

Naeher, Dominik

**Article — Published Version**

## Technology adoption under costly information processing

International Economic Review

**Provided in Cooperation with:**

John Wiley & Sons

*Suggested Citation:* Naeher, Dominik (2021) : Technology adoption under costly information processing, International Economic Review, ISSN 1468-2354, Wiley Periodicals, Inc., Hoboken, USA, Vol. 63, Iss. 2, pp. 699-753,  
<https://doi.org/10.1111/iere.12545>

This Version is available at:

<https://hdl.handle.net/10419/264470>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<http://creativecommons.org/licenses/by/4.0/>

## TECHNOLOGY ADOPTION UNDER COSTLY INFORMATION PROCESSING\*

BY DOMINIK NAEHER

*University of Goettingen, Germany*

A new explanation for imperfect technology adoption is proposed. In the model, agents allocate scarce attention rationally between actions and decide whether and how to adopt a technology. Introducing constraints to information processing gives rise to attentional barriers to technology adoption, which affect optimal adoption behavior. The model can explain why individuals (i) fail to adopt profitable technologies, (ii) fail to make the best use of adopted technology, (iii) treat complementary technologies as substitutes, and (iv) change their adoption decision back and forth over time. The model complements existing learning models and is consistent with empirical evidence.

### 1. INTRODUCTION

According to macroeconomic growth literature, much of the variation in per-capita income levels across countries can be attributed to differences in total factor productivity (Prescott, 1998; Hall and Jones, 1999; Easterly and Levine, 2001). An important determinant of the associated productivity gaps between countries is the range of technologies used (Aghion and Howitt, 1992; Grossman and Helpman, 1993; Keller, 2004). Adoption of technological advancements is therefore commonly seen as a driver of productivity growth and as a key pathway for less-developed countries to catch up with richer economies (Comin and Hobijn, 2004; Banerjee and Duflo, 2005; Hall, 2005).<sup>1</sup> Yet, a large body of evidence shows that many innovations are only very slowly adopted, and, conditional on being adopted, frequently used in sub-optimal ways.<sup>2</sup>

This article proposes a new explanation for why this may be the case. The explanation centers around the idea that adoption decisions are affected by attentional costs associated with adequately managing and operating technology in the presence of constraints to information processing. There are several reasons for why these costs may be of economic relevance.

[Correction added on January 19, 2022 after first Online publication: Footnote numbers and manuscript received and revised information updated]

Manuscript received May 2020; revised December 2020 and May 2021.

I thank Matthias Schündeln, Mirko Wiederholt, and anonymous referees for extensive comments and suggestions. I also thank Sandro Ambuehl, Luc Christiaensen, Ester Faia, Rema Hanna, Dany Jaimovich, Stefan Klonner, David McKenzie, Devesh Rustagi, Sara Savastano, and seminar participants at Goethe University Frankfurt, Heidelberg University, Peking University, and University College Dublin for helpful insights and discussions. Please address correspondence to:

Dominik Naeyer, University of Goettingen, Waldweg 26, 37073 Goettingen, Germany. E-mail: [dnaeher@uni-goettingen.de](mailto:dnaeher@uni-goettingen.de).

<sup>1</sup> This view is supported by a growing body of empirical evidence showing that returns to investments into modern technology are often extremely high for households in developing countries (Miguel and Kremer, 2004; Udry and Anagol, 2006; Duflo et al., 2008). Several authors argue that answering the question of why technological diffusion is often so slow is crucial for understanding the broader process of economic growth and development (Rosenberg, 1972; Geroski, 2000; Atkin et al., 2017).

<sup>2</sup> This includes innovations in farming technology (Udry and Anagol, 2006; Duflo et al., 2011), health and sanitation (Dupas, 2009; Ahuja et al., 2010; Ashraf et al., 2010), and management practices (Bloom et al., 2013; Atkin et al., 2017). Further examples can be found in the surveys by Rosenberg (1972) and Keller (2004).

First, many modern technologies require users to make complex choices about the parameters of usage, such as the type, timing, and composition of different inputs. Second, the returns associated with different parameter choices are not necessarily fixed, but are often sensitive to time- and location-specific conditions (specific examples are discussed below). To apply such technologies in optimal ways, users are required to adapt their practices to the present conditions anew in each period. This imposes a continuous need to be attentive to new information, because in deciding how to manage and operate adopted technology, users can rely only to a limited extent on knowledge obtained from past observations, for example, based on learning about optimal input targets. As a consequence, users may be required to process large amounts of new information in limited amounts of time to apply adopted technology optimally.

In standard models based on rational expectations, such an overload of information does not affect optimal technological adoption decisions since agents are assumed to process available information instantaneously and without additional cost, that is, there are no constraints to information processing. However, a growing body of empirical evidence shows that in many situations, people do not (or only imperfectly) absorb readily available information and translate it into appropriate choices.<sup>3</sup> This indicates that humans possess only a limited capacity for processing information, which may affect optimal decisions about uptake and usage of modern technologies.

Consider, for example, the input choice problem of a textile manufacturer. To stay on the efficient production frontier, the company's managers will have to be attentive not only to the operations within their own factory, but also to changes in customer preferences, the supply network, and the prevailing legal rules and regulations (among others). At least to some extent, fluctuations in any of these domains will also affect the company's optimal production decisions. Similarly, a farmer who chooses whether to cultivate an improved crop variety needs to consider factors such as present soil and weather conditions, prices of agricultural inputs, and demand for specific crops. In both cases, applying even relatively simple technologies in optimal ways creates a continuous need to attend to information.

The model I propose for studying technology diffusion incorporates this feature by analyzing optimal adoption decisions in a choice framework with rationally inattentive agents (Sims, 2003; Matějka and McKay, 2015). In the model, returns to using a particular technology are dependent on accompanying parameter choices. Which parameter values are optimal depends on the fundamental state of the economy, which is not directly observed by agents. By processing relevant information, agents can reduce uncertainty about optimal parameters and thereby increase the expected value of adoption. As attention is a scarce resource, it has to be allocated between the choice problem associated with technology adoption and other relevant decisions. This allows to incorporate aspects of technology use that are typically neglected in the existing literature on technology diffusion. The model I construct provides a structural framework for studying the roles of information- and attention-related features of technology in determining optimal adoption behavior, including the degree of a technology's complexity, the cost of paying attention to information about optimal usage parameters, the returns to technology use under mistakes, and the properties of other decision problems that are competing for users' attention.

The model generates testable predictions that are helpful in understanding empirical adoption decisions. Most importantly, the model gives rise to the existence of "attentional barriers" to technology adoption. These barriers affect optimal adoption decisions and can cause nonadoption of otherwise profitable and economically feasible technology to be a rational decision. In addition, the model offers an explanation for why individuals and firms fail to

<sup>3</sup> This includes evidence from laboratory studies (Gabaix et al., 2006; Goecke et al., 2013; Ambuehl et al., 2018), field experiments (Hossain and Morgan, 2006; Hanna et al., 2014; Bartoš et al., 2016), and results based on observational data (Hortaçsu and Syverson, 2004; Hirshleifer et al., 2009; Handel, 2013; Bronnenberg et al., 2015; Grubb and Osborne, 2015; Bhargava et al., 2017). Further examples can be found in the surveys by Falkinger (2008), DellaVigna (2009), Handel and Schwartzstein (2018), and Maćkowiak et al. (2021).

make the best use of technologies they decided to adopt, treat complementary technologies as substitutes, and change their adoption decisions back and forth over time. I show that the predictions of the model are consistent with empirically observed adoption behaviors, including in contexts where existing theories are unable to (fully) account for these behaviors.

The article links to three strands of literature. First, I add to the literature on technology diffusion. Within this literature, various channels have been proposed to account for imperfect uptake and usage of technology. These include market imperfections such as credit rationing, incomplete insurance, and weak property rights (Feder, 1980; Moser and Barrett, 2006; Caselli and Coleman, 2001; Dercon and Christiaensen, 2011; Karlan et al., 2014), individual and social learning (Parente, 1994; Foster and Rosenzweig, 1995; Jovanovic and Nyarko, 1996; Bandiera and Rasul, 2006; Conley and Udry, 2010; Hanna et al., 2014), regulatory and organizational issues (Parente and Prescott, 1994; Bloom et al., 2013; Atkin et al., 2017), and behavioral biases (Ashraf et al., 2006; Duflo et al., 2011). All of these channels have been found to be empirically relevant in a variety of contexts. Yet, they can only account for part of the observed patterns of technology diffusion, and various observations remain puzzling (see Sections 2 and 5). I propose that costly information processing may be part of the explanation for these findings and develop a model that complements the insights of existing theories.

Within the large body of the literature on technology diffusion, my model relates most closely to learning (or epidemic) models, since these models also emphasize the importance of uncertainty about optimal usage parameters in explaining imperfect adoption behavior. In these models, learning is typically based on the discovery of new information that allows agents to refine their knowledge and choose better practices. As pointed out by Hanna et al. (2014), the explanation for imperfect adoption decisions that learning models generate is essentially based on a *lack of data*. In contrast, the model I propose for studying technology adoption focuses on situations where information that is useful for deriving optimal practices is in principle readily available to users, but *limits to data processing* restrict agents in absorbing and translating the information into appropriate decisions. Importantly, classical learning models such as target input models have the feature that agents can only reduce uncertainty about those components of optimal actions that are persistent across time. In contrast, in the model proposed here, agents reduce uncertainty about idiosyncratic conditions that are independent across time. This generates very different implications than those generated by learning models. As demonstrated below, the insights obtained from my model should be viewed as complementing the insights obtained from classical learning models, as the two approaches apply to opposite types of technologies.

Second, I add to the growing body of literature that stresses the role of attentional constraints for economic decision-making. This literature has considered applications across various topics, including macroeconomic contexts such as monetary transmission (Mankiw and Reis, 2002; Maćkowiak and Wiederholt, 2009), consumption dynamics (Luo, 2008; Tutino, 2013), and business cycles (Maćkowiak and Wiederholt, 2015), as well as applications in finance (Van Nieuwerburgh and Veldkamp, 2010; Kacperczyk et al., 2016), industrial organization (Sallee, 2014; Martin, 2017), and labor (Bartoš et al., 2016; Acharya and Wee, 2020).<sup>4</sup> I add to this literature by extending the analysis based on limited attention to technological diffusion processes, including an application to agricultural development.

Finally, in incorporating limitations to information processing into the model, I build on previous work in the literature on rational inattention started by Sims (2003), particularly on the discrete choice framework under rational inattention studied by Matějka and McKay (2015). An important feature of my model that, to the best of my knowledge, is novel to the rational inattention literature, is that it generates a discontinuity in the amount of attention allocated across multiple decision problems (this discontinuity arises at the threshold value of attentional capacity that defines whether adoption or nonadoption of the technology is optimal).

<sup>4</sup> Summaries of these papers are provided in the survey papers listed in footnote 3.

It should be noted that most of the predictions of my model arise independently of the underlying structure of information and thus could also be generated by a model with costly information acquisition that is not based on entropy reduction. Nevertheless, using a rational inattention approach to describe technological adoption decisions under costly information processing offers several advantages. First, the concept of rational inattention introduced by Sims (2003) is based on information theory, which provides a wide range of readily-available mathematical tools that are well understood. Second, and partly due to the existence of these tools, rational inattention models tend to allow for fairly tractable solutions. Moreover, using an entropy-based functional form for attention costs allows to endogenize not only the amount of information to be processed (i.e., the precision of the underlying signals), but also the structure of information (i.e., the distribution from which signals are drawn).<sup>5</sup> It has therefore been argued that rational inattention can be viewed as ideal for studying information acquisition based on endogenous considerations about the benefits of information instead of based on exogenous assumptions about the cost of information (Kőszegi and Matějka, 2020).

On the other hand, authors have pointed out that rational inattention maintains the strong (neoclassical) assumption that people are able to form accurate beliefs based on the information that is worth processing (Handel and Schwartzstein, 2018; Kremer et al., 2019). The empirical support for this assumption is mixed (see the survey by Handel and Schwartzstein, 2018), with several studies supporting the usefulness of rational inattention in explaining the outcomes of human choice behavior across different contexts (Caplin and Dean, 2013; Goecke et al., 2013; Bartoš et al., 2016; Ambuehl et al., 2018).<sup>6</sup>

The rest of the article is organized as follows: The next section summarizes features of modern technologies that motivate the analysis. Section 3 presents the model and derives testable predictions. Section 4 discusses the robustness of the theoretical results to changes in the assumed features of the model, and clarifies the model's relationship to other theories of technology adoption such as learning models. Section 5 applies the model to explain empirical adoption decisions of technologies in the context of agricultural development. Section 6 concludes.

## 2. FEATURES OF TECHNOLOGY THAT MOTIVATE THE ANALYSIS: THE CASE OF MODERN FARM INPUTS

The model I propose for studying technology adoption is a simple choice framework under rational inattention. The model focuses on contexts in which adoption decisions can be characterized by three stylized properties. First, applying the technology requires users to decide on a number of parameters, for example, the types and quantities of different inputs. Second, users face uncertainty about optimal parameter choices because the optimal values are sensitive to the fundamental state of the economy which users observe only imperfectly. Third, users have access to information that is helpful in reducing uncertainty about optimal ways of applying the technology, but have to exert costly mental effort to process this information.

There are many types of technology that could be argued to feature these stylized properties. One important example, which the application in Section 5 focuses on, are modern

<sup>5</sup> Earlier approaches to modeling limited attention typically either abstract completely from information processing or impose an exogenous signal structure on decision makers. In contrast, the rational inattention literature uses an entropy-based approach to model information processing that allows agents to endogenously choose the distribution from which signals are drawn.

<sup>6</sup> An alternative approach to model limited attention are salience-based models (e.g., Bordalo et al., 2013; Kőszegi and Szeidl, 2013). In these models, attention allocation is determined by a stimulus-driven process instead of the true value information carries. In explaining imperfect technology adoption, these models would thus focus on factors that are exogenous to users instead of on users' own optimizing behavior. For a more detailed discussion of the different approaches to attention used in economics, including a comprehensive discussion of the assumptions underlying the rational inattention approach, see Handel and Schwartzstein (2018).

agricultural input technologies such as fertilizer, pesticides, and improved seed varieties.<sup>7</sup> This focus is well placed, as there is a large body of empirical literature concerned with explaining adoption decisions of these technologies, especially in the development context (see the survey by Foster and Rosenzweig, 2010).<sup>8</sup> In addition, it is often argued that these technologies are likely to become more important over time as the world's growing demand for food has to be met with ever scarcer land and water resources, possibly exacerbated in coming decades by the effects of climate change (IFAD, 2013; FAO, 2014).

In this context, the economic literature has identified four puzzling findings which I will use in Section 5 to demonstrate the model's ability to contribute to the understanding of empirical adoption decisions.

- **Observation 1.** Farmers in developing countries spread complementary inputs across different agricultural plots instead of combining them on the same plots, thereby forgoing potential gains from agronomic synergies (Sheahan and Barrett, 2017).
- **Observation 2.** Farmers abstain persistently from using modern farming technologies, even in situations where these technologies are readily available, affordable, and profitable (Duflo et al., 2011; Datta and Mullainathan, 2014).
- **Observation 3.** Farmers who have adopted modern farm inputs often use them in suboptimal ways, for example, by not adapting their practices to idiosyncratic conditions and shocks (Morris et al., 2007; Duflo et al., 2008; Gollin and Udry, 2019).
- **Observation 4.** Farmers frequently switch in and out of using previously adopted technologies, including across subsequent agricultural seasons (Dercon and Christiaensen, 2011; Duflo et al., 2011; Suri, 2011).

Observation 1 is a puzzling finding because agronomic synergies are a well-known feature of modern farm inputs and many existing studies highlight the importance of joint adoption of these inputs to raise agricultural productivity.<sup>9</sup> Since this observation applies to farmers who are already using modern inputs on their farm, it is particularly useful to distinguish the model developed below from other factors commonly studied in the literature to explain imperfect adoption decisions (see Section 5). Therefore, I design the model to be able to capture complementarity between different technologies and multiple production locations. In addition, I show that the behavior of a rationally inattentive agent can account for imperfect uptake and usage of profitable technology more generally, including Observations 2–4.<sup>10</sup>

To provide a bit more motivation for the assumptions underlying the model constructed below, the rest of this section summarizes other relevant empirical findings. First, a large body of evidence suggests that the returns to modern farm inputs are very sensitive to the chosen cultivation practices, and that choosing profitable practices is not always an easy task for

<sup>7</sup> Although the application discussed in this article focuses on a relatively narrow set of technologies, it should be noted that the model is more broadly applicable and the derived insights are useful in explaining technological adoption decisions beyond the specific application studied here.

<sup>8</sup> The diffusion of modern farm inputs is commonly seen as playing an important role in the growth process of developing countries (Irz et al., 2001; Evenson and Gollin, 2003; Gollin, 2010; Christiaensen et al., 2011). This applies especially to sub-Saharan Africa, where a large share of the labor force works in agriculture, productivity gaps tend to be particularly large in the agricultural sector, and adoption rates of modern farming technologies are often persistently low (Morris et al., 2007; World Bank, 2008; Sheahan and Barrett, 2017).

<sup>9</sup> The literature has identified many such agronomic synergies. For example, fertilizer tends to yield higher returns on irrigated plots as it requires a certain soil moisture to effectively enhance plant growth (Duflo and Pande, 2007; Morris et al., 2007). Since fertilizer facilitates the growth of undesired weeds, it should optimally be combined with herbicide (Beaman et al., 2013; Sheahan and Barrett, 2017). Integrated soil fertility management is based on synergies arising from combined use of organic and mineral fertilizers (Place et al., 2003). Finally, many new seed varieties are designed to be paired with agrochemicals and therefore feature strong complementarities with these inputs. These synergies are only fully realized if inputs are combined in adequate ways, for example, regarding the timing and use of specific types of inputs (Sheahan et al., 2013; Nyangena and Juma, 2014; Abay et al., 2016).

<sup>10</sup> Although these observations can also be accounted for by other factors, I propose that costly information processing may be part of the explanation for these observations and show that it complements the insights of existing theories (particular those of learning models; see Subsection 4.2).



farmers. For example, Cole and Fernando (2012) state that due to seasonal variation in the types and resistance of pests, continued adjustments in the types and quantities of used pesticides are required to achieve effective pest control. In a similar way, many authors argue that the profitability of fertilizer use depends on tailoring the dosage, composition, and timing of application to field-specific and seasonal conditions (Marenya and Barrett, 2009; Conley and Udry, 2010; Sheahan et al., 2013). Evidence from a randomized field experiment by Duflo et al. (2008) shows that farmers in Kenya face considerable uncertainty about profitable quantities of fertilizer use, with appropriate quantities being highly profitable (raising farmers' annual net incomes by 15% on average) and suboptimal quantities being associated with negative net returns.<sup>11</sup> The problem of selecting optimal cultivation practices is exacerbated by the fact that farmers tend to face distinct growing conditions on different fields, including for fields located in the same village.<sup>12</sup> Finally, the view that choosing optimal ways of applying agricultural inputs represents a challenging task for farmers is supported by the growing literature on precision agriculture, which argues that even conditions at the micro level, that is, within individual plots, can play an important role in determining optimal usage practices (Koch et al., 2004; Lambert et al., 2006).

The model also builds on the assumption that users have access to information that is helpful in choosing optimal ways of using the technology, but face limited mental resources in attending to this information. In the context of modern farm inputs, such information may include features of the agricultural production process itself that are relevant for choosing optimal cultivation practices, for example, information about the soil conditions, growing stages of planted crops, and early signs of pests and diseases. Typically, these data can be directly observed by farmers. In addition, farmers may have access to external sources of information, such as weather forecasts, user manuals, advice from neighboring farmers, and agricultural extension services (some of which may be associated with monetary costs). Deviating from standard models without constraints to information processing, I follow the literature on limited attention and incorporate the idea that users of technology have to exert costly mental effort to absorb information and incorporate it into their decision-making. Unless one is willing to assume that farmers' information processing differs systematically from that of other people, the empirical support for this feature in the context of agricultural input use is essentially the same as the evidence discussed in Section 1 (including the studies listed in footnote 3). In addition, there are several reasons to believe that constraints to information processing may be especially severe in the context of rural development. Most importantly, farmers in poor countries tend to have very limited access to information in preprocessed form, for example, due to the unavailability of tools such as online search engines. Moreover, poor farmers are often unable to benefit from distraction-saving goods and services, such as stable electricity and water supply, which has been argued to deteriorate available mental resources (Banerjee and Mullainathan, 2008). Finally, some studies suggest that cognitive functioning is adversely affected by poverty (Mani et al., 2013; Haushofer and Fehr, 2014).

### 3. MODEL

I start by illustrating the main idea of the article in a simple benchmark model of technology choice in which the agent allocates available attentional capacity between two decisions under uncertainty. Subsection 3.4 extends the model to incorporate interactions between different technologies across multiple production location.

<sup>11</sup> Duflo et al. (2008) summarize their findings with the words, "...while fertilizer can be very profitable when used correctly, one reason why farmers may not use fertilizer and hybrid seeds is that the official recommendations are not adapted to many farmers in the region. This also suggests that fertilizer is not necessarily easy to use correctly, which implies that it may not be profitable for many farmers who do not use the right quantity."

<sup>12</sup> Conley and Udry (2010) report that while growing conditions are typically highly correlated across adjacent plots, they can significantly vary over a village as a whole, for example, because soil types, topographical features, and rainfall realizations frequently differ on opposite sides of a single village.

3.1. *Basic Setup.* Consider the decision problem of a single agent who faces two decisions. The first decision corresponds to the adoption of a certain technology of interest and is denoted by  $\tau$ . If the technology is used, the agent has to select between two different ways of applying it, for example, low and high quantity, or early and late timing. Let these two options be denoted as *low* ( $l$ ) and *high* ( $h$ ).<sup>13</sup> In addition, let the outside option of not using the technology be denoted as 0.

The payoff associated with each way of applying the technology depends on the state of the world. Let  $v_{\tau,i}$  denote the payoff for using the technology under option  $i_{\tau}$ . The payoff for a given usage practice depends on the fundamental  $z_{\tau}$ . One may think of  $z_{\tau}$  as capturing all relevant conditions that determine the returns to different ways of using the technology. In the basic setup considered here,  $z_{\tau}$  takes two possible realizations,  $l$  and  $h$ , with equal probability, and the payoff for each way of using the technology is given by

$$(1) \quad v_{\tau,high} = \begin{cases} a, & z_{\tau} = h \\ -b, & z_{\tau} = l \end{cases} \quad \text{and} \quad v_{\tau,low} = \begin{cases} a, & z_{\tau} = l \\ -b, & z_{\tau} = h, \end{cases}$$

where  $b > a > 0$ .<sup>14</sup> For example, if the realized state is  $h$ , using the technology with option *high* yields a return equal to  $a$ , whereas using the technology with option *low* yields a return equal to  $-b$  (and vice versa for  $z_{\tau} = l$ ). This captures the idea that adopting the technology is profitable if it is used correctly according to the realized state, but can lead to negative (net) returns under inadequate usage (see, e.g., the example of modern farm inputs discussed in Section 2).

If the agent chooses not to adopt the technology, a safe return is obtained which is normalized to zero. In this case, the agent does not have to engage in a choice about how to apply the technology. For example, this may be thought of as the possibility to apply a less complex technology that does not require any parameter choices.<sup>15</sup>

The second decision, denoted as  $-\tau$ , corresponds to another decision problem that requires the agent's attention. One may think of  $-\tau$  as capturing distractions in the spirit of Banerjee and Mullainathan (2008) or as the aggregate of all decisions which the agent currently faces other than the one associated with  $\tau$ . The choice problem for  $-\tau$  features the same payoff structure as the one for  $\tau$  given by expression (1), except that there is no safe outside option available for  $-\tau$ , that is, the agent is required to choose one of the two options  $i_{-\tau} \in \{low, high\}$ .<sup>16</sup> Importantly, I focus on the case where the states  $z_{\tau}$  and  $z_{-\tau}$  are independent of each other (relaxing this assumption is discussed in Subsection 4.1).

3.2. *Information Processing.* Although the agent is aware of the underlying payoff structure, the realized state of the economy is initially unobserved such that the agent faces uncertainty about the payoffs associated with different actions. To reduce uncertainty, the agent can process information.<sup>17</sup> Let the agent's prior belief over the realized state  $z_k$  be given by the probability mass function  $G(z_k)$ . Processing information about  $z_k$  can be modeled as receiving a signal  $s_k \in \mathbb{N}$  on the realization of  $z_k$  to update the belief  $G(z_k)$ . The agent's actions

<sup>13</sup> The model can be generalized to allow for more than two options without qualitatively changing the results (see Appendix A.1.1). Throughout the article, I use the terms "option," "usage practice," and "way of applying the technology" interchangeably.

<sup>14</sup> Assuming  $b > a$  is necessary to ensure that all options represent relevant alternatives for the agent. As shown in Appendix A.1.1, this assumption can be relaxed in a more general choice framework with more than two options (see also footnote 29).

<sup>15</sup> The existence of a riskless outside option is a common assumption in models of technology adoption (e.g., Munshi, 2004; Bandiera and Rasul, 2006; Foster and Rosenzweig, 2010).

<sup>16</sup> Alternative payoff structures are considered in the model extensions in Subsection 3.4 and Appendix A.1.1.

<sup>17</sup> Relevant information may come from external sources, such as other people and media, or consist of internal data generated by the production process itself, for example, production status and condition of equipment (see Section 2 for more examples).



are based on the posterior belief resulting from processing information, captured by the conditional probability mass function  $\Gamma(z_k | s_k)$ . Following the literature on rational inattention (Sims, 2003; Maćkowiak and Wiederholt, 2009; Matějka and McKay, 2015), there are no restrictions imposed on the structure of the signal (other than that prior and posterior beliefs must be consistent with each other). The agent chooses  $s_k$  freely by selecting  $\Gamma(s_k, z_k) \in \Delta$ , where  $\Delta$  is the set of all joint probability mass functions over signals and states.<sup>18</sup>

However, in line with a growing body of evidence on the limits of human cognition (see the references listed in footnote 3), processing information is assumed to require attentional effort, and the agent faces a limited attentional capacity. Let  $\kappa_k$  denote the amount of attention that is allocated to processing information about  $z_k$ , and let  $\bar{\kappa} > 0$  denote the agent’s total endowed attentional capacity.<sup>19</sup> Limited attention is modeled as a constraint on uncertainty reduction. As is standard in the rational inattention literature, uncertainty is quantified by entropy, a measure of the unpredictability of a random variable’s realization. This is formally captured by the constraint

$$(2) \quad H(G(z_k)) - E_s[H(\Gamma(z_k | s_k))] \leq \kappa_k,$$

where  $H(G(z_k))$  denotes the prior entropy associated with the choice problem for  $k \in \{\tau, -\tau\}$  and  $E_s[H(\Gamma(z_k | s_k))]$  is the expected posterior entropy after the agent has processed information. In the case of a discrete choice problem considered here, the entropy of the prior belief is defined as

$$(3) \quad H(G(z_k)) = - \sum_{z_k} G(z_k) \log G(z_k),$$

and the posterior entropy given the received signal  $s_k$  is given by

$$(4) \quad H(\Gamma(z_k | s_k)) = - \sum_{z_k} \Gamma(z_k | s_k) \log \Gamma(z_k | s_k).$$

Condition (2) thus states that the more attention is devoted to problem  $k \in \{\tau, -\tau\}$  (i.e., the larger  $\kappa_k$  is), the larger is the expected reduction in uncertainty (measured in entropy) about the realization of  $z_k$ .

**3.3. Decision Problem and Optimal Adoption Behavior.** The timing of the model is as follows: First, the agent chooses the allocation of attention. This includes both the decision of how much attention to devote to each choice problem and the structure of the signal. For example, the agent may choose to use all attention to discriminate between the possible states of  $z_\tau$ , while paying no attention to  $z_{-\tau}$ . Second, the state of nature realizes (e.g., the realized state of  $z_\tau$  is  $l$  and the realized state of  $z_{-\tau}$  is  $h$ ). Third, the agent receives a signal about the state of the fundamentals (e.g., the signal might say that the state of  $z_\tau$  is  $l$  and the state of  $z_{-\tau}$  is also  $l$ ). Recall that the signal is more likely to be correct the more attention the agent devotes to this decision. Fourth, the agent chooses an option for each decision based on the

<sup>18</sup> Notice that this endogenizes the structure of information processing in the model, that is, the agent chooses not only the number or precision of signals based on some exogenously imposed signal structure, but all aspects of the underlying distribution from which signals are drawn.

<sup>19</sup> In the rational inattention literature, scarce attention is modeled either as a fixed capacity of endowed attention (as done here) or as a unit cost of attention. In the model in Subsection 3.4, working with a unit cost would lead to separate allocation problems across different production locations, that is, the optimal amount of attention paid to one location would be independent of the decisions for other locations. In explaining heterogeneity across locations, such an approach would thus have to rely on different parameter values across locations. In contrast, the capacity-based approach taken here is able to generate different optimal actions across locations even for the case where locations are *ex ante* identical.

posterior beliefs (this includes the decision whether to adopt the technology or not). Fifth, the agent receives the payoff for each decision, which is determined by the combination of state and selected option, as described by expression (1).

Formally, the agent’s decision problem is given as follows:

$$(5) \quad \max_{\{\kappa_k, \Gamma(s_k, z_k)\}_{k \in \{\tau, -\tau\}}} \sum_{k \in \{\tau, -\tau\}} \left( \sum_{z_k} \sum_{s_k} v_{k, i^*} \Gamma(s_k | z_k) G(z_k) \right),$$

$$(5.1) \quad \text{s.t. } i^* = \arg \max_i E[v_{k, i} | s_k]$$

$$(5.2) \quad H(G(z_k)) - E_s[H(\Gamma(z_k | s_k))] \leq \kappa_k \quad \forall k \in \{\tau, -\tau\},$$

$$(5.3) \quad \sum_{k \in \{\tau, -\tau\}} \mu \kappa_k \leq \bar{\kappa}.$$

The first condition (5.1) states that the agent chooses the option with the highest expected payoff given posterior beliefs. This corresponds to a standard choice problem under uncertainty.<sup>20</sup> The second condition (5.2) states that larger reductions in uncertainty (measured in entropy) require more attention. The third condition (5.3) is the budget constraint for the scarce resource attention, where the parameter  $\mu$  captures how demanding it is to pay attention to a decision.<sup>21</sup>

The objective of the agent is defined by expression (5). The term in parentheses is the expected payoff obtained from decision  $k$  for the chosen signal  $s_k$  and realized state  $z_k$ . For each realized state, the expected payoff equals the product of the probability of the state,  $G(z_k)$ , the conditional probability of the signal given the realized state,  $\Gamma(s_k | z_k)$ , and the expected value of option  $i^*$  that is chosen under the obtained signal as specified in condition (5.1). The agent chooses how much attention  $\kappa_k$  is allocated to each decision by deciding on the structure of the signal, which is captured by the joint probabilities  $\Gamma(s_k, z_k)$ . In doing so, the agent seeks to achieve higher probabilities of selecting options that yield larger returns given the realized state. Overall, the agent allocates the available attention  $\bar{\kappa}$  so as to maximize the sum of expected payoffs obtained from  $\tau$  and  $-\tau$ , subject to the given constraints.<sup>22</sup>

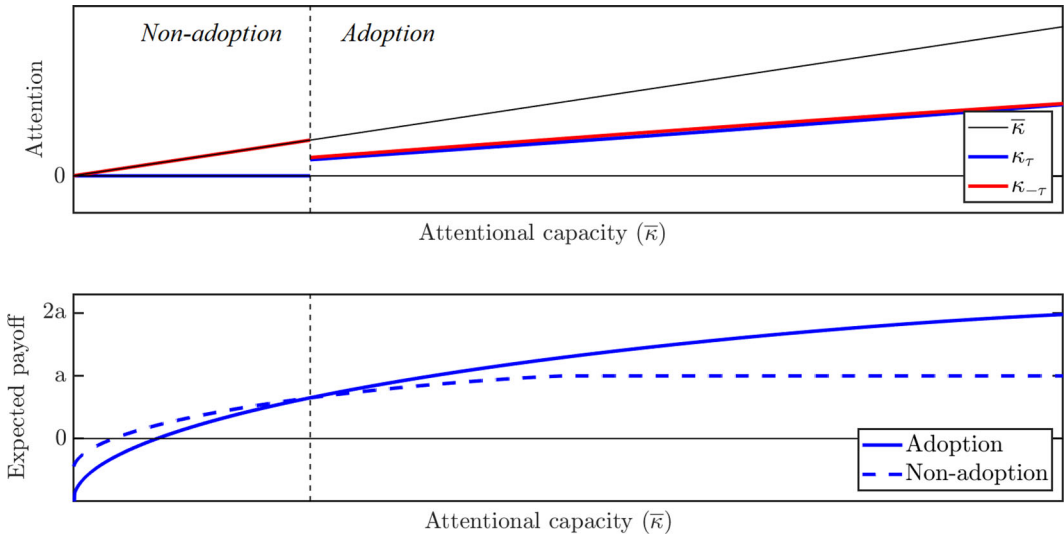
The solution to the optimization problem given by expressions (5) to (5.3) is derived in Appendix A.1.1. The solution shows that the optimal adoption behavior depends on the shadow price of attention, denoted as  $\lambda$ , which corresponds to the Lagrange multiplier of the budget constraint for attention (5.3).<sup>23</sup> Specifically, the optimal allocation of attention is given by the

<sup>20</sup> In related studies, this is sometimes referred to as the second-stage problem while choosing the allocation of attention is called the first-stage problem.

<sup>21</sup> The parameter  $\mu$  measures how much attention is needed to achieve a given reduction in the entropy associated with decision  $k$  (e.g., the agent may be more experienced in processing information about a certain type of decision or have access to information in preprocessed form, corresponding to smaller values of  $\mu$ ). Although in the framework studied here  $\mu$  is a constant, it would, in principle, also be possible to let  $\mu$  differ across decisions, meaning that paying attention to one decision would be more demanding than paying attention to another decision (see also the discussion around Corollary 1 in Subsection 3.4).

<sup>22</sup> Notice that the optimization problem given by expressions (5)–(5.3) is equivalent to a decision problem where the agent decides first whether to use the technology and then selects the allocation of attention and corresponding actions.

<sup>23</sup> When other parameters are fixed, the size of  $\lambda$  is fully determined by the available attentional capacity  $\bar{\kappa}$ . The relationship between  $\bar{\kappa}$  and  $\lambda$  is inverse, such that  $\lambda$  can be interpreted as a measure of the scarcity of attention in the



NOTE: Simulation for the case of two binary choice problems as specified in Subsection 3.1. Results are obtained by varying  $\bar{\kappa}$  on the interval between 0 and 2 for fixed parameter values  $(a, b, \mu) = (1, 2, 1)$ .

FIGURE 1

OPTIMAL ALLOCATION OF ATTENTION AND EXPECTED PAYOFF

following three cases:<sup>24</sup>

$$(6) \quad \kappa_{\tau}^* = \begin{cases} \log(2) & \text{if } \bar{\kappa} \geq \bar{m}_{\tau} \\ \log(2) - \frac{e^{\frac{a}{\lambda\mu}}}{e^{\frac{a}{\lambda\mu}} + e^{\frac{-b}{\lambda\mu}}} \log\left(\frac{e^{\frac{a}{\lambda\mu}} + e^{\frac{-b}{\lambda\mu}}}{e^{\frac{a}{\lambda\mu}}}\right) - \frac{e^{\frac{-b}{\lambda\mu}}}{e^{\frac{a}{\lambda\mu}} + e^{\frac{-b}{\lambda\mu}}} \log\left(\frac{e^{\frac{a}{\lambda\mu}} + e^{\frac{-b}{\lambda\mu}}}{e^{\frac{-b}{\lambda\mu}}}\right) & \text{if } \bar{\kappa} \in [m_{\tau}, \bar{m}_{\tau}), \\ 0 & \text{if } \bar{\kappa} < m_{\tau} \end{cases}$$

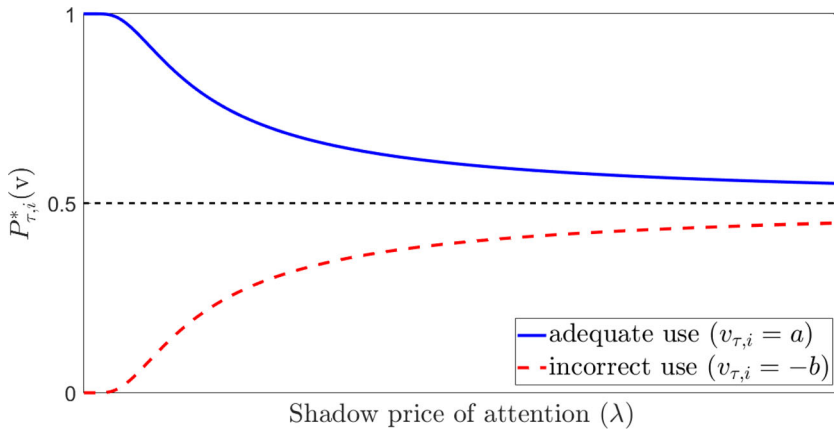
$$(7) \quad \kappa_{-\tau}^* = \frac{\bar{\kappa}}{\mu} - \kappa_{\tau}^*.$$

Figure 1 provides a graphical illustration of the agent’s optimal behavior implied by the results in Equations (6) and (7). The upper graph shows the optimal allocation of attention between  $\tau$  and  $-\tau$  for different values of  $\bar{\kappa}$ , keeping the other parameters fixed at  $(a, b, \mu) = (1, 2, 1)$ . The lower graph shows the associated expected payoff under adoption and nonadoption of the technology. If attention is abundantly available, that is,  $\bar{\kappa}$  is sufficiently large to yield a shadow price of attention  $\lambda$  equal to 0,<sup>25</sup> then the agent adopts the technology and uses it under optimal parameter choices. In this case (which corresponds to the first case for  $\kappa_{\tau}^*$  in Equation (6), that is, when  $\bar{\kappa} \geq \bar{m}_{\tau}$ ), the agent selects the optimal way of using the technology

agent’s overall decision problem. Although it is possible to derive an analytical expression that implicitly pins down  $\lambda$  as a function of the primitives of the model (i.e., the parameters  $a, b, \mu$ , and  $\bar{\kappa}$ ), it is in general not possible to derive a closed-form solution of the model. Most of the results are therefore presented as functions of  $\lambda$ . Appendix A.2 reports numerical results on the relationship between  $\lambda$  and individual parameters.

<sup>24</sup> Appendix A.1.1 derives the solution for a slightly more general case that allows for more than two options (e.g., more than two ways of applying the technology). Equations (6) and (7) represent a special case of that solution and can be obtained from the results stated in Proposition A.1 (Appendix A.1.1) by setting  $N_k = 2$ .

<sup>25</sup> Note that the total entropy in the agent’s decision problem is finite, such that  $\bar{\kappa}$  does not have to be infinite to lead to  $\lambda = 0$ . In the special case with two binary choice problems considered here, the entropy for each choice problem equals  $\log(2)$ , and the amount of attention that leads to  $\lambda = 0$  is given by  $\bar{m}_{\tau} = 2\mu \log(2)$ .



NOTE: Simulation for the case of two binary choice problems as specified in Subsection 3.1. Results are obtained by varying  $\lambda$  on the interval between 0 and 10 for fixed parameter values  $(a, b, \mu) = (1, 2, 1)$ .

FIGURE 2

PROBABILITIES OF ADEQUATE AND INCORRECT WAYS OF TECHNOLOGY USE

with probability one, so that the (expected) payoff from  $\tau$  equals  $a$  and the total payoff equals  $2a$  (see the right-hand edge of the lower graph in Figure 1).

As attention becomes scarcer, the agent uses the technology under imperfect attention, causing the agent to occasionally make mistakes (i.e., select incorrect parameters of usage associated with a payoff of  $-b$ ). In this case, corresponding to the second case in Equation (6) where  $m_\tau \leq \bar{\kappa} < \bar{m}_\tau$ , the expected payoff of using the technology is given by  $E[V_\tau(\kappa_\tau^*)] \in (0, a)$ . The smaller  $\bar{\kappa}$  (and thus the larger  $\lambda$ ), the less attention is available for processing information needed to operate the technology, and the more likely it is that mistakes occur. This is illustrated in Figure 2, which depicts the probability that the agent selects adequate ways (solid blue line) and incorrect ways (dashed red line) of using the technology for different values of  $\lambda$ . If the agent is perfectly attentive, that is,  $\lambda = 0$ , the probability of making a mistake equals zero. If attention is very scarce ( $\lambda$  is very large), the probability of making a mistake approaches the maximum value of one half. A larger shadow price of attention thus makes operating the technology in profitable ways more difficult, acting as a barrier to adoption.

As shown in the lower graph of Figure 1, the expected payoff decreases with smaller values of  $\bar{\kappa}$ . For a certain value of  $\bar{\kappa}$ , the optimal amount of attention allocated to each decision becomes too small to yield a positive expected return of using the technology. Since there is no outside option for  $-\tau$ , the optimal strategy is to switch to nonadoption of the technology. This allows the agent to obtain the secure return of zero for  $\tau$  and at the same time increase the expected payoff for  $-\tau$  as attention is shifted from  $\tau$  to  $-\tau$ . In Figure 1, the cutoff value  $m_\tau$  at which the agent switches to nonadoption corresponds to the value of  $\bar{\kappa}$  at the intersection of the two blue lines in the lower graph (as indicated by the vertical reference line).<sup>26</sup>

Importantly, notice that the agent’s optimal behavior features a discontinuity in attention allocation around the cutoff  $m_\tau$ . Intuitively, this discontinuity exists because it is never optimal to use the technology under very small amounts of attention  $\kappa_\tau$ , as the agent would be too much at risk of selecting inadequate usage practices.

<sup>26</sup> Note that, because any attention that is not used for  $\tau$  can be used to increase the expected payoff for  $-\tau$ , the switch from adoption to nonadoption occurs, in fact, before the expected return of adoption becomes negative. Also note that for sufficiently small values of  $\bar{\kappa}$ , even the expected total payoff under nonadoption is negative (see the left-hand edge of the lower graph in Figure 1). In the extreme case where no attention is available, the expected total payoff is given by  $E[V_{-\tau}] = 0.5(a - b) < 0$ , which equals the expected payoff in the absence of any useful information, that is, the simple expected value (this constitutes the best possible outcome for the agent in the absence of an outside option for  $-\tau$ ).

TABLE 1  
RETURN  $v_{j,i}$  GIVEN STATE  $z_j$  AND ACTION  $i$

	$i \in \mathcal{I}$	$z_j$				$E[v_{j,i} \mid \kappa_j = 0]^*$
		$(l, l)$	$(l, h)$	$(h, l)$	$(h, h)$	
Nonadoption	$(0,0)$	0	0	0	0	0
Single input	$(S_l, 0)$	$a$	$a$	$-b$	$-b$	$\frac{1}{2}(a - b)$
	$(S_h, 0)$	$-b$	$-b$	$a$	$a$	
	$(0, F_l)$	$a$	$-b$	$a$	$-b$	
	$(0, F_h)$	$-b$	$a$	$-b$	$a$	
Joint inputs	$(S_l, F_l)$	$2a(1 + s)$	$a - b$	$a - b$	$-2b$	$\frac{1}{2}a(2 + s) - b$
	$(S_l, F_h)$	$a - b$	$2a(1 + s)$	$-2b$	$a - b$	
	$(S_h, F_l)$	$a - b$	$-2b$	$2a(1 + s)$	$a - b$	
	$(S_h, F_h)$	$-2b$	$a - b$	$a - b$	$2a(1 + s)$	

NOTES: \*Expected return when no attention is devoted to the production location (i.e., simple expected value). Each state of the fundamental  $z_j$  is realized with equal probability 1/4.

Also note that, in the case of adoption, the agent allocates an equal amount of attention to  $\tau$  and to  $-\tau$ . This feature stems from the assumed symmetry of payoffs for  $\tau$  and  $-\tau$  as well as from the properties of the underlying information technology.<sup>27</sup> In particular, this highlights the model’s implicit assumption of preferences for generalized learning (as opposed to preferences for specialized learning; see Van Nieuwerburgh and Veldkamp, 2010). In the context of technology adoption, assuming preferences for generalized learning is plausible as long as it holds that users cannot arbitrarily change the relative roles of payoffs received from the actions corresponding to  $\tau$  and  $-\tau$ . In practice, this will be the case if the consequences of making suboptimal choices for the actions captured by  $-\tau$  cannot simply be avoided, for example, if  $-\tau$  represents distractions in the spirit of Banerjee and Mullainathan (2008) that must be dealt with (see Subsection 4.3 for a more detailed discussion).

3.4. *Complementary Inputs and Multiple Production Locations.* In the framework discussed so far, using technology is modeled as choosing a single input parameter (i.e., choosing among options  $i_\tau \in \{low, high\}$  in the baseline model described above or among options  $i_\tau \in \{1, \dots, N_\tau\}$  in the generalized version in Appendix A.1.1). Although this helps to simplify the analysis and is in line with other target input models commonly used to explain imperfect adoption behavior (see the references in Subsection 4.2), it is important to note that such an approach abstracts from possible dynamics between different aspects of technology use, including the existence of complementarity between technologies. In order to be able to study the role that these interactions play for optimal adoption decisions, I now present a special case of the model that incorporates the existence of complementary inputs as well as multiple production locations. As the application in Section 5 shows, the insights obtained from this extension are useful in explaining empirical patterns of technology use.

Consider the decision problem of an agent who operates identical production units in two different locations, for example, two factories or two agricultural plots, and has access to two complementary input technologies denoted  $S$  and  $F$ . For each location  $j \in \{1, 2\}$ , the agent chooses among  $i \in \mathcal{I}$  actions, which are listed in the first column of Table 1. In particular, the agent can choose not to adopt any of the two input technologies, to apply a single input, or to combine both inputs together. Whenever an input is used in a production location, it can be applied in two different ways. Let these two options be denoted as *low* ( $l$ ) and *high* ( $h$ ), respectively. In addition, let the outside option of not using a specific input be denoted as 0. In this stylized framework, the agent thus selects among nine possible actions for each

<sup>27</sup> The entropy-based information technology specified in Subsection 3.2 implies a convex cost function of attention that means that the expected marginal benefit of allocating more attention to  $-\tau$  decreases with higher levels of  $\kappa_{-\tau}$ , and analogously for  $\kappa_\tau$  in the case of adoption (see the second graph in Figure 1 as well as the discussion in Matějka and McKay, 2015, p. 279).

production location: nonadoption of any of the input technologies, four possible ways of using a single input in one unit, and four possible ways of using two complementary inputs jointly.<sup>28</sup>

The outcome of an action for production unit  $j$  depends on the state of the world at location  $j$ . Let  $v_{j,i}$  denote the return for action  $i$  in location  $j$ , and let  $\vec{v}_j$  denote the  $(9 \times 1)$  vector containing the returns for all possible actions in location  $j$ . The return for a given action depends on the state  $z_j$ . There are four possible state realizations, which are denoted as  $z_j \in \{(l, l), (l, h), (h, l), (h, h)\}$ . The notation is such that  $(l, h)$  indicates that in this state, the optimal way of applying the first input is *low* and the optimal way of applying the second input is *high*. If the state  $(l, h)$  realizes, then the vector of returns  $\vec{v}_j$  equals the corresponding column in Table 1. Nonadoption of any modern input yields a safe return that is normalized to zero. Adequate application of a single input yields a return equal to  $a$ . Therefore, the returns of actions  $(S_l, 0)$  and  $(0, F_h)$  in the  $(l, h)$  state equal  $a$ . Inadequate application of an input yields a return equal to  $-b$ . Hence, the returns of actions  $(S_h, 0)$  and  $(0, F_l)$  in the  $(l, l)$  state equal  $-b$ . Furthermore, adequate application of both inputs jointly creates synergy effects. For this reason, the return of the action  $(S_l, F_h)$  in the  $(l, h)$  state equals  $2a(1 + s)$  with  $s > 0$ . If one of the two inputs is applied inadequately, the synergy is lost and the joint return equals the sum of the individual returns.

In order to be able to capture the full spectrum of possible input choices, I focus on parameter value ranges that ensure that all actions represent relevant alternatives for the agent. In the stylized framework considered here, this requires that  $0 < a < b$  and  $s < \frac{b-a}{a}$  (see Appendix A.1.3).<sup>29</sup> In particular, this implies that using complementary inputs without devoting any attention to the respective production unit is dominated by the outside option (i.e., the expected values shown in the last column of Table 1 are smaller than the secure payoff obtained under nonadoption). As before, I focus on the case where each state is realized with equal probability and states are independent across the two locations (see Subsection 4.1 for a discussion on relaxing this assumption).

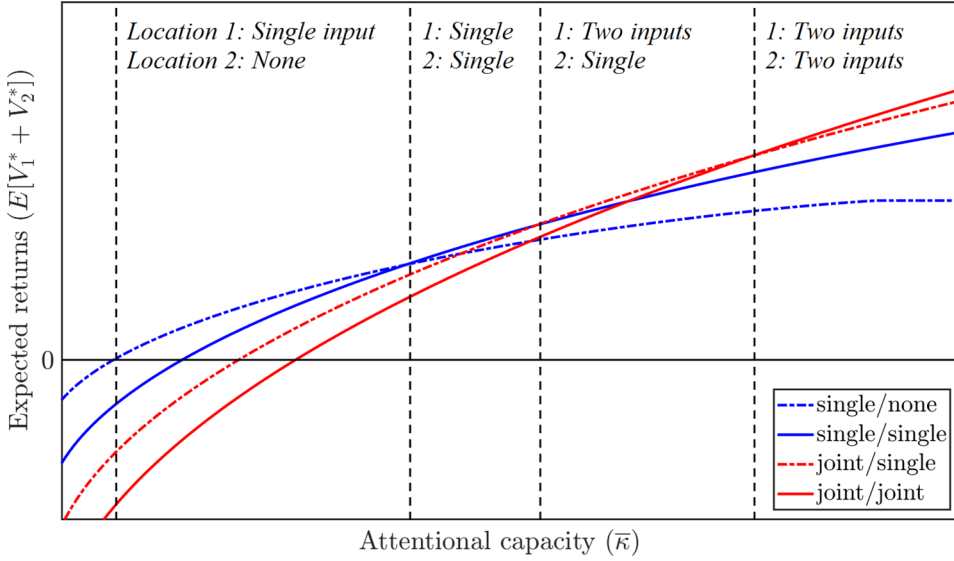
The distribution of payoffs specified in Table 1 captures two main ideas. First, the returns of combining complementary inputs depend on how well usage practices are adjusted to prevailing conditions. This introduces additional uncertainty into the agent's decision problem because parameters of usage for both inputs have to be chosen in conformity with the realized state to generate optimal outcomes, and optimal parameter choices may vary between different inputs and across locations. Second, joint adoption of complementary inputs is profitable, provided that the inputs are applied in adequate ways. At the same time, choosing suboptimal practices is associated with lower profitability and can, for some states of the world, lead to smaller net returns than those under nonadoption (e.g., this may be the case if the productivity gains associated with applying two complementary inputs in inadequate ways are not large enough to cover the costs incurred for purchasing these inputs). Importantly, it should be noted that the distribution of payoffs does not simply assume that there are states in which combining complementary inputs is not profitable, or less profitable than single input use. Rather, Table 1 specifies that joint input use always offers the largest potential payoff, but the realized payoff depends on the agent's ability to select the correct way of applying each input.

The timing of the model is analogous to the one in Subsection 3.3. First, the agent chooses how much attention to devote to the action in each of the two locations and sets the structure of the signal. Second, the state of nature realizes, for example, the realized state at location 1 may be  $(l, l)$  and the realized state at location 2 may be  $(h, h)$ . Third, the agent receives a

<sup>28</sup> Notice that the two locations may be thought of as two different decision problems similar to  $\tau$  and  $-\tau$  in the baseline model, where the nine actions in the first column of Table 1 correspond to the "options" in the baseline model (i.e., the model with complementary input technologies is a special case of the framework discussed in Appendix A.1.1 with  $N_k = 9$ ).

<sup>29</sup> These parameter restrictions can be relaxed by considering more than two possible options for each input. For example, if the agent had to choose the correct action for each input among  $N$  symmetric alternatives (as in the model in Appendix A.1.1), the expected value of using a single input without attention would be  $E[v_{j,i}] = \frac{1}{N}[a - b(N - 1)]$ , such that the restriction  $a < b$  could be relaxed to  $a < (N - 1)b$ .





NOTES: Simulations for different input combinations across two production locations as specified in Subsection 3.4. Results are obtained by varying  $\bar{\kappa}$  on the interval between 0 and 4 for fixed parameter values  $(a, b, s, \mu) = (1, 2, 0.05, 1)$ .

FIGURE 3

OPTIMAL INPUT CHOICES ACROSS TWO PRODUCTION LOCATIONS

signal about the state of the two locations, for example, the signal might say that the state of location 1 is  $(l, l)$  and the state of location 2 is  $(h, l)$ . Fourth, the agent takes an action in each location based on the posterior beliefs. Fifth, the agent receives the payoff for each production unit, which is determined by the combination of state and action, as described by Table 1.

The solution to the agent’s optimization problem with complementary input technologies is derived in Appendix A.1.3. Proposition 1 summarizes the agent’s optimal behavior:

**PROPOSITION 1.** *The optimal behavior of a rationally inattentive agent with access to two complementary input technologies and two ex ante identical production locations is characterized by a pair of cutoff values,  $m_1$  and  $m_2$ , specifying the minimum amount of attention  $\bar{\kappa}$  for which using a single input ( $m_1$ ) or using two inputs jointly ( $m_2$ ) at a location is profitable in expectation for given parameter values  $(a, b, s, \mu)$ .*

*If attention is sufficiently scarce, that is,  $\bar{\kappa} < m_1$ , the agent abstains from using any input technology. As  $\bar{\kappa}$  increases, actions corresponding to single or joint input use across the two locations become optimal, where the exact order in which these actions become optimal is sensitive to the parameter values. If parameters are such that all actions represent relevant alternatives, that is,  $0 < a < b$ ,  $s < \frac{b-a}{a}$ , and  $m_2 > m_1$ , then the order in which actions become optimal as  $\bar{\kappa}$  increases is as follows: For  $m_1 \leq \bar{\kappa} < 2m_1$ , the agent uses a single input in one location and no input in the other location. For  $2m_1 \leq \bar{\kappa} < m_1 + m_2$ , the agent uses a single input in both locations. For  $m_1 + m_2 \leq \bar{\kappa} < 2m_2$ , the agent uses two inputs jointly in one location and a single input in the other location. For  $\bar{\kappa} \geq 2m_2$ , the agent uses two inputs jointly in both locations.*

**PROOF.** See Appendix A.1.3. □

Figure 3 provides a graphical illustration of the agent’s optimal input choice behavior. It shows the expected returns of different input combinations across two locations for different values of  $\bar{\kappa}$ . As indicated by the intersections of the different graphs, there is no generally

dominant action, but the optimal behavior depends on the available attentional capacity. For very small values of  $\bar{\kappa}$ , that is, left of the first vertical reference line (corresponding to  $m_1$ ), the agent optimally chooses not to adopt any input. If the amount of available attention is somewhat larger, input choices that offer potentially higher returns—but also require more attention to achieve these higher returns in expectation—are optimal (where the order in which these input choices become optimal corresponds to the one described in Proposition 1).<sup>30</sup>

An important implication of the results in Proposition 1 is that choosing different combinations of inputs across production units can be an optimal outcome, even if the production units feature the same characteristics, that is, they are *ex ante* identical. For example, it may be optimal for the agent to combine both inputs in location 1 but use only a single input in location 2. This is the case, if the optimal amount of attention allocated to the two locations is insufficient to make combining inputs in both locations profitable in expectation. In addition, these results imply that it can be optimal for users to apply complementary inputs separately, for example, use one input in production location 1 and another input in location 2, despite the feature that combining both inputs at the same location offers potential synergies.

Moreover, the results give rise to the prediction that users may optimally decide to opt in and out of applying a particular technology over time, or vary the degree to which they combine complementary inputs in the same location. To see this, consider a simple dynamic extension of the model in which the choice problem of the agent in each period is given by the static optimization problem studied here. If the state of the world is uncorrelated over time and there is variation in  $\lambda$  across periods (e.g., in some periods, the agent faces more distractions or more demanding decisions than in other periods), then optimal behavior will change over time. The agent will thus adjust input practices across periods.<sup>31</sup> Notice that this offers a way to explain changes in input practices even in situations where fluctuations in external factors, for example, prices and weather conditions, which affect the profitability of these inputs, are not sufficient to (fully) account for observed changes in practices.

To complete the description of the solution of the model, I now explore how changes in the cost of processing information, captured by the parameter  $\mu$ , affect the agent's optimal adoption behavior. Recall that, for given attentional capacity  $\bar{\kappa}$ , different values of  $\mu$  may be thought of as capturing variation in factors that determine how demanding it is for users to absorb and evaluate information about the state of the world and associated optimal parameter choices.<sup>32</sup> In practice, there will be at least two possible sources of such variation. First, differences in  $\mu$  may reflect heterogeneity in user-specific characteristics that determine how costly it is for a firm or individual to process information. This may include differences in education, experience, and aspects of the environment in which the technology is operated. In addition, different values of  $\mu$  may capture the effect of external interventions that help users to reduce uncertainty about optimal usage practices. For example, this may include the provision of advisory services (Naeher and Schündeln, 2021), access to infrastructure, and tools such as a hotline (Cole and Fernando, 2012) or simply a user manual, which are helpful in reducing uncertainty about optimal usage practices. Irrespective of the underlying sources of variation in  $\mu$ , the effect on optimal input decisions in the model is summarized by the following corollary:

<sup>30</sup> Note that Figure 3 omits the possibility of combining two inputs in one location while using none of the inputs in the other location. This possibility is discussed in Appendix A.1.3.

<sup>31</sup> Consequently, the model also suggests that interventions aimed at promoting uptake and efficient usage of technology will have to provide continuous forms of support that are tailored to the current environment in which individual users operate (instead of, e.g., providing one-time support based on general “best practices”).

<sup>32</sup> Formally,  $\mu$  measures how much attention is needed to achieve a given reduction in the entropy associated with decision  $k$ . In the framework studied here,  $\mu$  is assumed to be constant across decisions, and is thus simply a scale factor of  $\bar{\kappa}$ . However, in principle it would also be possible to let  $\mu$  differ across decisions, meaning that paying attention to one decision would be more demanding than paying attention to another decision (see also footnote 21).

**COROLLARY 1.** *Users facing a lower cost  $\mu$  of processing information require a lower threshold in attentional capacity to adopt technology, and to combine complementary input technologies in the same production location.*

**PROOF.** See Appendix A.1.4. □

Intuitively, smaller values of  $\mu$  make it less costly for the agent to reduce uncertainty about how to optimally apply and combine complementary technologies, lowering the threshold in attention required to make adoption profitable in expectation. Consequently, the model suggests that a lower cost of processing information, stemming either from individual user characteristics or from external interventions, will tend to facilitate adoption and joint use of complementary input technologies.<sup>33</sup>

**3.5. Summary of Predictions.** Summarizing, the model gives rise to the following predictions about uptake and usage of technology in situations where profitable adoption requires users to process information.

**PREDICTION 1.** *If positive input synergies require usage practices to be tailored to idiosyncratic conditions, individuals with access to complementary input technologies but sufficiently scarce attention will choose to use these inputs separately instead of combining them in the same production location.*

This prediction corresponds to the agent's optimal behavior in the model with complementary input technologies (Subsection 3.4) when attention is sufficiently scarce, for example, when  $2m_1 \leq \bar{\kappa} < m_1 + m_2$  in Proposition 1 (left of the third reference line in Figure 3).

**PREDICTION 2.** *Individuals who have adopted modern technology will devote a significant amount of attention to managing and operating the technology. In doing so, users will direct their attention to a subset of the relevant information instead of absorbing all available information equally. In particular, users will devote more attention to technologies for which returns are more sensitive to the chosen parameters of usage, for example, when combining complementary inputs instead of using a single input.*

The first part of Prediction 2 corresponds to the implication of the model that it is never optimal to use the technology under very small amounts of attention, as the agent would be too much at risk of selecting inadequate usage practices (see the discontinuity in  $\kappa_\tau$  around the cutoff  $m_\tau$  in Figure 1). The second part of Prediction 2 corresponds to the results stated in Proposition 1, that is, that combining inputs in profitable ways requires more attention than using a single input profitably (in expectation).<sup>34</sup>

**PREDICTION 3.** *Individuals with sufficiently scarce attention will rationally choose not to adopt a technology, even if the technology is readily available and generally profitable when properly applied.*

Prediction 3 corresponds to the agent's optimal behavior in the third case specified in Equation (6), that is, when  $\bar{\kappa} < m_\tau$ , in the baseline model (and to the case when  $\bar{\kappa} < m_1$  in the model with complementary inputs; see Proposition 1).

<sup>33</sup> In general, the role of external interventions in facilitating technology adoption will also depend on whether users' learning behaviors are governed by strategic substitutes or strategic complements across different users (see the discussion in Subsection 4.4).

<sup>34</sup> Intuitively, this implication stems from the feature that, while joint input use offers potentially larger payoffs due to synergies, it also requires the agent to reduce uncertainty about both elements of the state (e.g.,  $l$  and  $h$  for the  $(l, h)$  state). In contrast, single input use requires attention only to one of the two elements of the state (only to the first element if  $S$  is used and only to the second element if  $F$  is used).

PREDICTION 4. *Individuals with a sufficiently large (but not perfect) attentional capacity will adopt the technology but fail to use it in optimal ways, for example, by adjusting usage parameters only imperfectly to idiosyncratic conditions. The probability of making these mistakes will tend to be higher, the scarcer users' attention is.*

Prediction 4 corresponds to the solution in the second case specified in Equation (6), that is, when  $m_\tau \leq \bar{\kappa} < \bar{m}_\tau$  in the baseline model. Similarly, in the model with complementary inputs, there is a strictly positive probability that the agent will fail to use a single input optimally, or to combine two inputs in optimal ways, if inputs are adopted while the constraint on information processing is binding. The probability of making these mistakes is higher, the larger the scarcity of attention (i.e., the larger the shadow price  $\lambda$ ) is in the agent's decision problem (recall Figure 2).<sup>35</sup> Importantly, even if these mistakes can sometimes cause the agent to incur negative payoffs, adoption may still represent optimal behavior (as long as the expected return of adoption exceeds the return under nonadoption).<sup>36</sup>

#### 4. DISCUSSION

This section discusses the roles of some of the (implicitly) assumed features of the model and the robustness of the derived results to changes in these assumptions. I also clarify the model's relationship to other theories of technology adoption, such as models based on learning or risk aversion.

4.1. *Independence of States across Decisions.* The model presented in Subsection 3.4 assumes that the realized states  $z_j$  are independent across locations  $j \in \{1, 2\}$ .<sup>37</sup> To provide some intuition about how relaxing this assumption will affect the agent's optimal adoption behavior, I now consider the possibility that states are correlated across decisions.<sup>38</sup> It should be noted that independent states are associated with the maximum possible uncertainty in the model. Specifically, the overall uncertainty that the agent faces about the state of the world can be quantified by the joint entropy of the two fundamentals. If  $z_1$  and  $z_2$  are independent, then the joint entropy is given by the sum of individual entropies, that is,

$$(8) \quad H(z_1, z_2) = H(z_1) + H(z_2).$$

If  $z_1$  and  $z_2$  are correlated, then the joint entropy is given by

$$(9) \quad H(z_1, z_2) = H(z_1) + H(z_2 | z_1),$$

which is strictly smaller than the joint entropy of independent variables specified in Equation (8) (see Cover and Thomas, 1991, p. 40). Intuitively, this is due to the fact that, when states are correlated, reducing uncertainty about the realized state in one location also helps the agent to reduce uncertainty about the realized state in the other location. In the case of full information, that is, if  $\bar{\kappa}$  is large enough to lead to  $\lambda = 0$  for given parameter values, it obviously does not matter whether states are correlated or not, since there is no uncertainty about the realized states. If attention is scarce (i.e.,  $\lambda > 0$ ), then the information technology specified in

<sup>35</sup> Formally, the probability of selecting the wrong option according to the realized state is given by the term  $e^{\frac{-b}{\lambda\mu}} / (e^{\frac{a}{\lambda\mu}} + e^{\frac{-b}{\lambda\mu}})$ , which depends positively on  $\lambda$  (see the respective expressions in Proposition A.1 in Appendix A.1.1 for the baseline model, and in Lemma A.1 in Appendix A.1.3 for the model with input complementarity).

<sup>36</sup> This applies to the case of risk-neutral agents considered here. The role of different utility functions and risk attitudes is discussed in Subsection 4.3.

<sup>37</sup> Analogously, the model in Subsection 3.1 and Appendix A.1.1 assumes that  $z_\tau$  and  $z_{-\tau}$  are independent of each other.

<sup>38</sup> Correlation here refers to dependence in an information-theoretic sense, that is, including all forms of stochastic dependence (not just linear dependence as captured, for example, by the Pearson correlation coefficient).

Subsection 3.2 implies that correlated states are associated with lower information processing costs to achieve the same reduction in uncertainty. In particular, if states are correlated, then for given parameter values  $(a, b, s, \mu)$  less attention  $\bar{\kappa}$  will be needed to make using complementary inputs together at the same location profitable in expectation compared to the case of independent states. Although a rigorous analysis of the model without the assumption of independent states would go beyond the scope of this article, these considerations suggest that the presence of correlation between different production locations will tend to facilitate adoption (and joint use) of technologies.

*4.2. Persistence of States and Relationship to Learning Models.* A large body of literature that seeks to explain empirical patterns of technology diffusion focuses on learning effects.<sup>39</sup> Similar to the proposed rational inattention model of technology adoption, the explanation for imperfect adoption behavior offered by the models developed in this literature is based on uncertainty about optimal parameter choices. However, there are several fundamental differences between the two approaches. Most importantly, learning models rely on the existence of persistence in optimal actions across time, since otherwise learning—in the sense captured by these models, that is, agents updating beliefs about optimal target values—would be impossible. On the other side, the proposed rational inattention approach abstracts from persistence in states and optimal actions, and instead focuses on the case where the agent’s choice of how to use technology is independent across time (with identical uniform priors over states in every period; see the discussion of a simple dynamic extension of the model in Subsection 3.4).

To illustrate the implications of this difference, consider the setup of a standard target input model with learning, for example, as in Foster and Rosenzweig (1995). In this setup, agents have to make a choice about how to adopt a new technology, and the associated returns are dependent on the proximity of the chosen parameters to predefined optimal values. Focusing on an agent with a single production unit, the optimal way of using the technology takes the form

$$(10) \quad a_{j,t}^* = \theta^* + u_{j,t},$$

where  $\theta^*$  is a fixed target value (referred to as “systematic component”) and  $u_{j,t}$  accounts for time-variant, individual-specific deviations from this target (“idiosyncratic component”).<sup>40</sup> Both components are initially unobserved, such that the agent faces uncertainty about the optimal action  $a_{j,t}^*$ . Based on this setup, learning models capture the idea that agents have to get access to more information to refine their knowledge about optimal actions. Importantly, learning in these models consists exclusively of updating beliefs about the systematic component  $\theta^*$ . Furthermore, learning in these models is only based on data that have been generated in the past.<sup>41</sup> In contrast, the model I propose for studying technology adoption formalizes the idea that users have to be attentive to “present conditions,” that is, to the idiosyncratic component  $u_{j,t}$ , to derive optimal parameter choices. In essence, the agents in learning models reduce uncertainty about  $\theta^*$ , whereas the agents in my model reduce uncertainty about  $u_{j,t}$ . In this particular sense, therefore, my model may be viewed as the other extreme to classical learning models. The insights obtained from my model thus complement the insights obtained from learning models.<sup>42</sup> In practice, most technologies will have features

<sup>39</sup> This includes Besley and Case (1993), Parente (1994), Foster and Rosenzweig (1995), Jovanovic and Nyarko (1996), Munshi (2004), Bandiera and Rasul (2006), Conley and Udry (2010), and Hanna et al. (2014).

<sup>40</sup> Although the systematic component does not necessarily have to be fixed to allow for learning, it needs to feature at least some persistence over time. In contrast, the idiosyncratic component may be completely independent across time.

<sup>41</sup> These data may have been generated by agents’ own experimentation (“learning by doing”) or obtained through knowledge spillovers from other users (“social learning”).

<sup>42</sup> Recent work on dynamic rational inattention (Steiner et al., 2017; Maćkowiak et al., 2018) may be viewed as combining these two approaches by studying uncertainty reduction for processes that involve both persistent and in-

that correspond to both approaches, that is, using technologies optimally will typically require users to learn about systematic components of the technology and to be attentive to idiosyncratic conditions. The relative degree to which each of these components matters likely differs across different types of technologies. For example, using a car requires knowledge that is highly persistent across time (knowing how to steer, change gears) as well as attention to present conditions (such as traffic, weather). A person who has learned how to drive is typically able to do so in various kinds of circumstances, so for the ability to adequately use a car the systematic component seems to play a relatively important role, that is, the magnitude of  $\theta^*$  is relatively large compared to  $u_{j,t}$ . In contrast, the fixed knowledge needed to use pesticide (e.g., how to spray pesticide on a plant) is relatively simple to obtain, but to apply pesticides adequately, including choosing the right type, quantity, and timing of treatment according to the prevailing conditions, often represents a challenging task for farmers (see the corresponding references in Section 2). Thus, for pesticides,  $u_{j,t}$  seems to be relatively important. Based on the discussion above, the ability of learning models to account for adoption decisions in each of these two cases differs fundamentally from the model I present. In particular, learning models are less applicable to technologies for which optimal parameter choices are very sensitive to idiosyncratic conditions (e.g., pesticides), because agents' choices in these models are exclusively based on knowledge about the systematic component. On the other hand, my model is less applicable to technologies for which optimal parameter choices are very persistent across time and states, and idiosyncratic conditions play only a minor role. This demonstrates that the insights obtained from my model and those obtained from learning models should be viewed as complementing each other (instead of contradicting each other), as they apply to opposite types of technologies with respect to the persistence of optimal parameters of usage.<sup>43</sup>

In addition, there are several other ways in which my model differs from learning models. First, notice that in learning models, updating beliefs is solely based on information that has been generated by previous outcomes. In contrast, the model I propose captures the idea that users also have to be attentive to changes in existing conditions to choose optimal practices, that is, it is not sufficient to rely on past experience.

Second, my model incorporates the idea that users are constrained in their ability to reduce uncertainty about optimal choices even if relevant information about  $\theta^*$  and  $u_{j,t}$  is readily available. In essence, the explanation for imperfect technology adoption provided by learning models is based on a *lack of data*, whereas the bottleneck in the suggested rational inattention approach consists of *limits to data processing*. Learning models are thus less suited to account for imperfect adoption in contexts where users face an overload of relevant information instead of a lack of information. The rational inattention approach, on the other hand, is in principle able to explain imperfect adoption decisions even in situations where all information needed to derive optimal parameter choices is readily available. One implication of this difference is that learning models are in general better suited for explaining uptake of recent innovations for which relatively little information is available to users. My model complements the insights of these models by offering a possibility to account for low adoption rates and suboptimal usage also in contexts where lack of access to information is not a binding constraint.

Third, learning models generate the prediction that usage practices tend to improve over time.<sup>44</sup> This stems from the feature that agents' knowledge about optimal actions monotonically increases in these models as more and more data become available (provided there is

dependent components across time. These studies show that agents with memory and limited attention will learn both about the current optimal action and the best predictors of future optimal actions.

<sup>43</sup> Note that this insight also applies to the basic setup discussed in Subsection 3.1, that is, the presence of complementarity between different inputs is not required to rationalize nonadoption of available and profitable technology for which the optimal parameters of usage are sensitive to idiosyncratic conditions.

<sup>44</sup> An exception is the model by Hanna et al. (2014) that combines learning with limited noticing. Still, their model assumes that updating beliefs is exclusively based on observations of past outcomes, and once an input dimension has been attended to, it is used optimally in all successive periods. Although this offers a way to explain why people may persistently fail to make optimal decisions about some input dimensions, their model cannot account for fluctuations



perfect recall). In contrast, the channel I propose can account for fluctuations in the quality of usage practices, such as the occurrence of occasional mistakes that have not been made before, since agents have to be attentive again in every new period to make optimal choices. Importantly, this mechanism does not imply that such mistakes must necessarily decrease over time. As long as the expected return of using the technology with imperfect attention is greater than the return under nonadoption, agents will apply the technology even if there is a strictly positive probability of making mistakes. This offers a possibility to account for persistently suboptimal usage of adopted technology in the long run.

Finally, there exists empirical evidence that people tend to switch back and forth between using and not using previously adopted technology, even long periods of time after the technology has been introduced (e.g., the evidence underlying Observation 4 in Section 2). Classical learning models are able to account for such switching behavior only in the short term, that is, right after a new technology has become available and users are still experimenting to find out whether (and how) the technology can be used in profitable ways (as in, e.g., Conley and Udry, 2010). The proposed rational inattention channel offers a possible way to rationalize postadoption switching behavior also in the long run, based on the recurring need to attend to idiosyncratic conditions (see the discussion of a simple dynamic extension of the model in Subsection 3.4), though a more rigorous dynamic analysis in this regard is left to future work.

*4.3. Different Utility Functions and Preferences for Specialized Learning.* One concern may be the robustness of the derived model predictions with respect to the underlying utility function. In the framework presented in Subsection 3.1 and Appendix A.1.1, the agent maximizes the sum of expected payoffs over two decision problems, with payoffs being symmetric across options. Importantly, this means that the implications derived from the model do not require agents to be risk averse (the same holds for the case with complementary input technologies studied in Subsection 3.4). If agents were assumed to be risk averse, for example, the utility associated with adopting the technology would depend not only on the expected return for  $\tau$  but also (negatively) on the variance of returns, then this would tend to amplify the models' ability to account for imperfect adoption decisions (Predictions 1 and 3) while leaving Predictions 2 and 4 qualitatively unchanged.

In addition, it should be noted that the restrictions on the parameters  $a$ ,  $b$  (and  $s$  in the model in Subsection 3.4) are made to ensure that all choices represent relevant alternatives for the agent. Relaxing or dropping these restrictions would lead to outcomes where, for some combinations of parameter values, the agent would either never adopt or always adopt the technology, irrespective of the value of  $\bar{\kappa}$ .<sup>45</sup> However, this would leave intact the model's general ability to generate Predictions 1–4 for particular parameter values.

This could potentially be different if not only the values of the payoffs were allowed to differ but also the assumed structure of payoffs. In particular, the solution of the model relies on the assumption that options are *a priori* homogeneous, that is, for each action, the available options seem equally attractive to the agent before processing information (generalizing this feature would make the model intractable analytically, though it might still be analyzed computationally). Together with the entropy-based information technology specified in Subsection 3.2, this feature implies preferences for generalized learning, that is, the agent tends to spread the available attentional capacity over all actions and all options (except for action  $\tau$  if the riskless outside option is taken). If instead, the agent had preferences for specialized learning, as in Van Nieuwerburgh and Veldkamp (2010), then the optimal allocation of

in the quality of input choices, for example, why farmers manage to apply fertilizer at the right time and in right quantities in some periods but fail to do so in other periods.

<sup>45</sup> Analogously, if the made parameter restrictions in the model in Subsection 3.4 would be dropped, then only a subset of the possible input combinations across two locations would be considered, for example, the agent would never use a single input if  $s$  is sufficiently large for given values of  $a$  and  $b$ .

attention would tend to be less diversified, possibly affecting some of the predictions of the model (as discussed below).

Whether assuming preferences for generalized learning is plausible depends on the nature of the technology adoption problem. For example, in the context of portfolio choice studied by Van Nieuwerburgh and Veldkamp (2010), investors can choose to invest more strongly in some assets and avoid the returns from other assets by abstaining to invest. This can lead to “feedback effects” and preferences for specialized learning where investors optimally pay attention only to a small subset of assets and trade heavily in these. The technology adoption problem I consider differs in several ways from the portfolio choice problem. In particular, financial investors can arbitrarily change the relative roles of payoffs received from different assets by adjusting their portfolio. In contrast, the scope for scaling up (or down) the use of technology is much more limited for most of the technologies studied in the economic literature (this is captured in the model by fixing the maximum payoff that the agent can receive from any domain). For example, many agricultural input decisions are made only once per growing season and the extent to which modern farm inputs can be used is constrained by the limited availability of land, labor, and credit.<sup>46</sup> The potential role of feedback effects leading to specialized learning is thus more limited in the context of technology adoption than, for example, in the portfolio choice problem. In addition, note that the choice problem with complementary input technologies studied in Subsection 3.4 somewhat relaxes the assumption of generalized learning by allowing the agent to choose between using a single input and combining two inputs at the same location. As summarized in Prediction 2, the optimal behavior in this case features a specialization in attention allocation where the agent allocates more attention to locations under joint input use than to locations operated with a single input.

If the assumption of preferences for generalized learning was to be further relaxed in the model, this would tend to cause the agent to focus the available attention more on a single domain. In principle, this could lead to a result where the agent specializes in  $\tau$  and adopts the technology even under small amounts of attentional capacity, potentially eliminating the non-adoption result in Prediction 3.<sup>47</sup> However, even with preferences for specialized learning the results summarized in the other three predictions would remain qualitatively the same under plausible conditions. Specifically, suppose that there was an outside option for  $-\tau$  and the payoffs were such that it would be optimal for the agent to allocate all available attention to  $\tau$ . As long as (i) payoffs are such that profitably combining complementary inputs requires more attention than profitably using a single input, (ii) both single and joint input use are relevant alternatives, and (iii) attention is a binding constraint, it will still hold that the agent will rather use a single input than combining complementary inputs if  $\bar{\tau}$  is sufficiently small (Prediction 1), devote more attention to locations under joint input use than to locations operated with a single input (Prediction 2), and occasionally use adopted technology in suboptimal ways (Prediction 4).

Finally, it should be noted that the entropy-based information technology specified in Subsection 3.2 has the built-in feature that it is independent of the labeling of states, that is, it is equally costly to distinguish nearby states and distant states (irrespective of the associated payoffs). One might thus be concerned whether the derived predictions would also hold under alternative functional forms of the cost of processing information that do not have this feature. In the baseline model in Subsection 3.1, relaxing the assumption that all states are equally costly to distinguish would not change the main results (i.e., those summarized in Predictions 1–4). To see this, suppose that the information technology was such that it was easier for the agent to reduce uncertainty about  $z_k$  if the states  $l$  and  $h$  are further apart (according to some metric). Although this would affect the quantitative results of the model (e.g., the absolute value of the cutoff  $m_\tau$  would depend on the distance between the states  $l$  and  $h$ ), apart

<sup>46</sup> This applies especially to peasant farmers (e.g., those studied in Section 5) and only to a lesser extent to commercial farms (or, in a more general context, large firms, or governments) that are less capital-constrained.

<sup>47</sup> This effect would be further exacerbated if there was persistence in states  $z_\tau$  over time and the agent expects to use the technology also in future periods (I thank an anonymous referee for pointing this out to me).

from that the solution for the agent's optimal behavior would remain qualitatively the same.<sup>48</sup> Similarly, in the choice problem with two complementary input technologies in Subsection 3.4, the main results would not change qualitatively if it was relatively easier for the agent to reduce uncertainty about the correct option of using one input than the other. Again, the absolute values of the cutoffs  $m_1$  and  $m_2$  would be sensitive to such changes (and there may be a clear preference introduced as to which of the two inputs gets adopted first), but other than that it would still hold that profitably using both inputs together requires more attention than using only a single input, and thus using synergistic inputs separately can be rational (Prediction 1).<sup>49</sup>

*4.4. Multiple Agents and Social Learning.* The model in Section 3 studies the adoption decision of a single agent and abstracts from the possibility of learning spillovers and strategic interactions between multiple agents. As discussed in Subsection 4.2, learning, in the sense of updating beliefs about optimal target values, requires that optimal actions are correlated over time. Similarly, learning spillovers are only possible if optimal actions are correlated across agents. In existing learning models (e.g., those listed in footnote 39), this is typically the case because optimal actions are partly determined by a systematic component that is persistent (or even constant) across both time and states (recall the parameter  $\theta^*$  in Equation (10)). In contrast, the rational inattention channel studied here focuses on the role of idiosyncratic conditions and individual uncertainty reduction in determining adoption decisions.<sup>50</sup>

Once the focus on idiosyncratic conditions is relaxed, however, social learning would in principle also be possible for the rational inattention channel. To see this, consider a simple dynamic multiagent extension of the model, in which the choice problem of each agent in each period is given by the static optimization problem in Subsection 3.3. If the state  $z_\tau$  is uncorrelated over time but correlated across agents within each period, then the uncertainty reduction about the realization of  $z_\tau$  achieved by one agent may potentially also benefit other agents.<sup>51</sup> Together with the existence of a mechanism through which agents can observe each other or exchange information, such a setting would give rise to similar strategic interactions as studied in many other contexts in the economic literature.

For example, with the presence of spillovers in uncertainty reduction, farmers may freeride on neighboring farmers' information, thereby facilitating widespread adoption (and joint use) of modern technologies. Given that such spillovers make it easier for individuals to choose profitable ways of applying technology, this will tend to weaken the model's ability to account for imperfect adoption. However, as long as agents' learning behaviors feature strategic complements, this will mainly affect the outcomes of the model quantitatively (i.e., the derived cutoff values characterizing optimal adoption behavior will change), whereas most of the qualitative predictions of the model will remain the same. Specifically, as long as costly information processing remains a binding constraint even with such spillovers, it will still hold that agents will prefer using a single input to combining complementary inputs if attentional capacity is sufficiently scarce (Prediction 1), devote more attention to complex usage decisions

<sup>48</sup> In particular, such a relaxation would lead to the same solution as multiplying the right-hand side of Equation (5.2) with an appropriate number for given parameter values.

<sup>49</sup> Although the feature that all states are equally costly to distinguish does not play a crucial role for the results in the single-agent model studied here, this might be different when moving to a framework with multiple agents and strategic interactions (see Morris and Yang, 2019, and the discussion in Subsection 4.4).

<sup>50</sup> Empirical evidence in the literature on technology adoption suggests that the roles of individual and social learning differ across contexts and technologies. Although in some contexts, social learning seems to be an important factor in the diffusion process (e.g., Foster and Rosenzweig, 1995; Conley and Udry, 2010), other studies find that for technologies whose performance is more sensitive to idiosyncratic conditions people tend to rely predominantly on individual learning (Munshi, 2004; Dufo et al., 2010). For related theoretical insights, see Smith and Sørensen (2000) and Ali (2018).

<sup>51</sup> Note that this only applies to each period separately. The associated spillovers in uncertainty reduction thus represent a static form of social learning that is different from the dynamic concept of social learning captured in classical learning models.

(Prediction 2), and occasionally use adopted technology in suboptimal ways (Prediction 4). In this case, decentralized markets may not yield the socially optimal level of attention, thereby providing support for additional coordination or government intervention.

If agents' learning behaviors are instead governed by strategic substitutes, such that it is optimal to specialize in acquiring different types of information (see Hellwig and Veldkamp, 2009; Veldkamp, 2011), then some of the model's predictions may change also qualitatively (e.g., the result in Prediction 2 that individuals will devote more attention to complex usage decisions might be offset by additional effects stemming from strategic considerations in a multiagent model). In this case, there will be less of a need for government intervention to facilitate information exchange across agents. Future work concerned with the difference between the optimal social level of information processing and the optimal individual level of attention might find it interesting to embed the presented single-agent model of technology adoption into a larger social learning framework.

## 5. APPLICATION

"The small minority of households that are using multiple modern inputs tend to spread them across plots rather than concentrating them on single plots. This behavior has gone largely unstudied to date and raises important questions about prospective untapped productivity gains from coordinated modern inputs use."

Sheahan and Barrett (2017, p. 18)

As argued in Section 2, adoption decisions of modern farm inputs in developing countries provide a suitable context for applying the model, because farmers often face considerable uncertainty about optimal ways of applying these inputs and because constraints to information processing are often severe in the context of rural development. In this context, Sheahan and Barrett (2017) find that there is surprisingly low correlation between the use of commonly paired modern inputs at the household and plot level in nationally representative samples of farmers in sub-Saharan Africa.<sup>52</sup> Using the same data source to quantify joint adoption rates of specific sets of modern inputs, I find that only 18% of households use improved seeds and inorganic fertilizer together on their farms, 10% of households apply fertilizer jointly with other agrochemicals (herbicide or pesticide), and less than 4% of households have adopted sets of at least three of these modern farm inputs.<sup>53</sup>

It is important to note that, in principle, the low correlation between the use of these inputs at the household level could be due to any of the explanations for nonadoption of (individual) farm technologies discussed in the economic literature, such as risk aversion or lack of awareness about the profitability of modern farm inputs. A more differentiated picture arises when looking at adoption decisions at the plot level (farmers typically cultivate multiple plots, with a mean of two plots in my sample). Table 2 shows joint adoption rates of modern inputs at the plot level when focusing exclusively on those households who are using at least one modern input on their farm. This suggests that the included households generally have access to modern farm inputs and are aware of their benefits. Column (1) of Table 2 shows that households who already use modern inputs on the farm combine complementary inputs only on 25% of their plots. At the same time, 46% of plots receive a single modern input and 29% of plots are cultivated without any modern input. As shown in columns (2)–(5), similar results are obtained when conditioning on household-level use of specific inputs (instead of any of

<sup>52</sup> This is a puzzling finding since a well-known feature of these modern farm inputs is that they are most productive when applied in combination with each other (see the discussion in Section 2, including the references listed in footnote 9).

<sup>53</sup> Figures are based on nationally representative survey data from Malawi, Nigeria, Tanzania, and Uganda, collected by the World Bank's Living Standard Measurement Study - Integrated Surveys on Agriculture (see Appendix A.3 for more details).

TABLE 2  
PLOT-LEVEL USE OF MODERN FARM INPUTS CONDITIONAL ON ADOPTION ON THE FARM

Number of Modern Inputs Used on Plot	Modern Inputs Adopted on the Farm				
	Any Input (1)	Inorganic Fertilizer (2)	Improved Seeds (3)	Other Agrochemicals (4)	Irrigation (5)
None	28.5	16.7	20.6	18.7	18.7
One	46.1	44.4	43.1	39.7	31.2
Two	22.4	33.9	30.9	32.9	25.6
Three or more	3.0	5.0	5.5	8.6	24.5
Observations	14904	8918	7187	4893	507

NOTES: Numbers are percentages of plots cultivated with a certain number of modern inputs, conditional on households using at least one modern input on their farm as specified in each column (i.e., excluding plots that are cultivated by farmers who use no modern inputs in column (1), and farmers who do not use the input specified in columns (2) to (5), respectively). Plots that are rented out or left fallow are excluded. *Source:* Author's calculations based on nationally representative survey data from Malawi, Nigeria, Tanzania, and Uganda in the World Bank's Living Standards Measurement Study - Integrated Surveys on Agriculture (see Appendix A.3 for more details).

the observed modern inputs as done in column 1).<sup>54</sup> The results in Table 2 suggest that even among the subset of households who are generally able and willing to use modern farm inputs, a majority of households are forgoing potential gains from agronomic synergies by failing to combine complementary inputs on the same plots. This largely rules out explanations for low joint usage rates that are based on factors at the household level, such as limited access, lack of awareness, and behavioral biases. In particular, if any of these factors were a binding constraint for the included households, we would not observe these households use modern farm inputs on other plots.

The model presented in Subsection 3.4 offers a possible way to rationalize these findings. For example, the model implies that farmers may optimally decide to use only a single input on a plot, even if complementary inputs are readily available and joint usage under adequate practices leads to positive synergies (Prediction 1). Furthermore, the model predicts that this behavior will be more likely for farmers who face higher attentional costs associated with reducing uncertainty about optimal input practices. A direct test of this prediction would require data on farmers' allocation of attention across different plots and the underlying attentional costs. To the best of my knowledge, such data are so far not available. However, other studies have argued that individuals' costs of processing information can, at least to some extent, be inferred from characteristics such as age, education, and complexity of the studied task (Greenwood and Parasuraman, 1991; Verhaeghen et al., 2003; Ambuehl et al., 2018). In the context of smallholder farmers, it also seems plausible to assume that farmers face larger costs of being attentive to the conditions on their plots, the more plots they cultivate, and the further away these plots are located from their home. Based on these considerations, there should be an empirical link between these factors and farmers' input choices, even when holding other variables that are correlated with these factors fixed.

In addition, reducing uncertainty about optimal input choices is likely more demanding in situations where farmers face growing conditions that differ very strongly from the conditions encountered in previous seasons, for example, because farmers can rely less on past experience. According to Corollary 1, farmers should then be less likely to use modern farm inputs, and less likely to combine complementary inputs on the same plots, when facing irregular growing conditions.

To test the empirical plausibility of these predictions, I estimate a set of linear probability models of the form

$$(11) \quad y_{cij} = \beta_0 + \beta_1 \Delta Rainfall_{ci} + \beta_2 Rainfall_{ci} + \beta_3 X_{cij}^1 + \beta_4 X_{cij}^2 + \lambda_c + \epsilon_{cij},$$

where the dependent variable is a dummy capturing different numbers of modern farm inputs used on plot  $j$  of household  $i$  in country  $c$ . The explanatory variable of interest is  $\Delta Rainfall_{ci}$ , which is a vector of two variables measuring deviations in the amount and timing of rainfall from past average conditions for each household. The first of these two variables is constructed by calculating the absolute deviation of 12-month total rainfall for each household from the household's long-term average value (covering the past 10–15 years, depending on data availability), expressed in units of standard deviation.<sup>55</sup> The second variable is constructed in the same way based on the start of the wettest quarter (measured in dekads), which usually corresponds to the beginning of the main growing season. These two variables thus capture the degree by which current growing conditions that differ from the

<sup>54</sup> For example, column (2) of Table 2 indicates that 39% of plots cultivated by farmers who have adopted inorganic fertilizer receive combinations of at least two modern inputs, whereas 61% of plots receive either no modern input or only a single input. When I restrict the analysis to plots cultivated by households who use at least two modern inputs on the farm, still more than half of the plots (54%) receive less than two inputs, and only 7% of plots receive sets of three inputs. Similar results are also obtained when I focus on plots planted with maize, which is known to feature relatively large input complementarities, particularly for combinations of improved seeds, fertilizer, and pesticide (Sheahan et al., 2013; MacRobert et al., 2014; Nyangena and Juma, 2014).

<sup>55</sup> If rainfall is serially uncorrelated within small areas across years, that is, local rainfall follows a white-noise process, then the average will also correspond to the expected value (see Paxson, 1992).



conditions farmers are used to based on previous rainfall patterns. According to the considerations above, these two variables should be negatively associated with modern input use. To control for the effects of rainfall on input choices through other channels than increased uncertainty about optimal usage practices, the actual realizations of the quantity and timing of rainfall, denoted by the vector  $Rainfall_{ci}$ , are themselves also included.  $X_{cij}^1$  and  $X_{ci}^2$  are vectors of plot and household-level variables,  $\lambda_c$  denotes country fixed effects, and  $\beta_0$  is a constant. The error term  $\epsilon_{cij}$  is permitted to be heteroskedastic and spatially correlated as a function of physical distance (see Conley, 1999; Conley and Udry, 2010).<sup>56</sup> A more detailed description of the construction of variables and underlying data sources is provided in Appendix A.3, along with basic summary statistics.

Table 3 reports estimates of the regression model specified in Equation (11) for different numbers of farm inputs as dependent variable (the included inputs are improved seed varieties, organic and inorganic fertilizer, agrochemicals, and irrigation). In the first three columns, the dependent variable is a dummy that equals 1 if the plot was cultivated using at least one of these inputs. The dependent variable in columns (4)–(6) equals 1 if the plot received at least two inputs, and the dependent variable in columns (7)–(9) equals 1 if the plot received at least three inputs. The results in Table 3 show that, irrespective of whether additional plot and household controls are included or not, deviations in the timing and amount of rainfall from past realizations are negatively correlated with modern input use. Importantly, this holds despite the fact that the actual rainfall shocks, that is, the realizations of the quantity and timing of rainfall, are also included (see columns (3), (6), and (9)). For deviation in the timing of rainfall, the estimated coefficients are highly statistically significant in all specifications. For deviation in the amount of rainfall, the estimated coefficients are significant at the 10-% significance level when the dependent variable is an indicator for joint use of at least two inputs or of at least three inputs and controls are included (columns (6) and (9)). When the dependent variable captures use of at least one modern input (column (3)), the point estimate for deviation in the amount of rainfall is negative, as anticipated, but cannot be statistically distinguished from zero at conventional levels of significance. As discussed in Appendix A.3.2, most of these results are robust to different ways of clustering standard errors and to alternative definitions of the considered rainfall variables. I thus interpret the existence and direction of the correlations in Table 3 as being in line with the predictions of the model, but I also stress that the observational nature of the data limits my ability to make causal claims. The results suggest that the magnitude of the association between rainfall irregularities and farmers' input choices is of moderate economic relevance. For example, the estimated coefficient of deviation in rainfall timing in column (3) of Table 3 indicates that a one-standard-deviation increase in the irregularity of rainfall timing (corresponding to a shift in the start of the wettest quarter of about 30 days) is associated with a decrease in the probability of modern input use of 2.2 percentage points.

Furthermore, the results in columns (3), (6), and (9) of Table 3 show that use of modern farm inputs is negatively correlated with the distance of the plot to the farm house, number of plots cultivated by the household, and age of the households head. The dummy indicating whether any household member has attained secondary education (or more) is positively correlated with the dependent variable in columns (3) and (6). Of course, these findings cannot be interpreted as causal, but the directions of these correlations seem to be consistent with the model. From the perspective of the model, these findings could be attributed to the idea that it is more demanding for farmers to be attentive to the conditions on their plots, the more different plots they cultivate, and the further away these plots are located from their home. Also, according to the studies cited above, limited attention may be more likely to represent a binding constraint for farmers who are older and less educated.

<sup>56</sup> The use of spatial standard errors is motivated by the fact that shocks to growing conditions are likely correlated across space instead of based on political boundaries such as districts. All regressions use a conical (Bartlett) spatial weighting kernel that decays linearly with distance. The regressions in Table 3 are based on a cutoff distance of 50 km. Robustness checks are conducted with a cutoff of 100 km and for clustered standard errors at the district level (see Appendix A.3.2).

TABLE 3  
RAINFALL IRREGULARITIES AND MODERN FARM INPUT USE

Number of Modern Inputs Used on Plot:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	One or More Inputs			Two or more inputs			Three or more inputs		
$\Delta$ Rainfall timing (absolute deviation)	-0.038*** (0.009)	-0.021V*** (0.010)	-0.022** (0.009)	-0.041*** (0.007)	-0.024*** (0.007)	-0.025*** (0.006)	-0.020*** (0.003)	-0.011*** (0.003)	-0.011*** (0.003)
$\Delta$ Rainfall amount (absolute deviation)	-0.023** (0.011)	-0.009 (0.011)	-0.013 (0.010)	-0.025** (0.010)	-0.011 (0.010)	-0.015* (0.009)	-0.013*** (0.005)	-0.008* (0.004)	-0.008* (0.004)
Rainfall timing: start of wettest quarter		0.003 (0.002)	0.002 (0.002)	0.003 (0.001)	0.003** (0.001)	0.003*** (0.001)	0.001** (0.001)	0.001** (0.001)	0.001** (0.000)
Rainfall amount: 12-month total rainfall		-0.019*** (0.004)	-0.020*** (0.003)	-0.019*** (0.003)	-0.019*** (0.003)	-0.020*** (0.003)	-0.010*** (0.002)	-0.010*** (0.002)	-0.010*** (0.002)
Distance of plot to farm house > 1 km			-0.061*** (0.010)			-0.033*** (0.009)			-0.010** (0.004)
Number of cultivated plots			-0.026*** (0.005)			-0.015*** (0.004)			-0.007*** (0.002)
Secondary education or above			0.055*** (0.011)			0.023** (0.010)			-0.001 (0.005)
Age of household head			-0.001*** (0.000)			-0.001*** (0.000)			-0.001*** (0.000)
Country fixed effects	yes	yes	yes	yes	yes	yes	yes	yes	yes
Additional control variables	21584	21584	21584	21584	21584	21584	21584	21584	21584
Observations	0.179	0.190	0.225	0.104	0.120	0.154	0.038	0.051	0.070
R <sup>2</sup>	0.532	0.532	0.532	0.221	0.221	0.221	0.055	0.055	0.055
Mean dependent variable									

NOTES: \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Conley spatial standard errors reported in parentheses are based on a Bartlett weighting kernel with 50 km bandwidth. Inputs included in the dependent variable are improved seed varieties, organic and inorganic fertilizer, agrochemicals, and irrigation. Additional control variables are as follows: At the plot level, plot size and dummies for good soil quality, sloped, erosion, rented, and land title are included. At the household level, farm size (cultivated area), number of household members, distance to road, distance to market, food expenditure, nonfood expenditure, and dummies for rural household and male household head are included. All regressions include an additional constant which is not reported. *Source:* Author's calculations based on nationally representative survey data from Malawi, Nigeria, Tanzania, and Uganda in the World Bank's Living Standards Measurement Study - Integrated Surveys on Agriculture (see Appendix A.3.1 for more details).

The model is also able to account for other features of empirical adoption decisions of modern farm inputs. Summarizing the existing evidence on low uptake of fertilizer in developing countries, Datta and Mullainathan (2014, p. 13) state that, “fertilizer is available, affordable, effective, and appreciated. But it is still not used by farmers.”<sup>57</sup> Importantly, theories based on limited access to fertilizer or low returns are unable to account for the low adoption rate of fertilizer in this context. In addition, Duflo et al. (2011) argue that neither learning models nor theories based on credit constraints and risk aversion can account for the observed behavior.<sup>58</sup> Instead, these authors propose a behavioral explanation centered around the idea that farmers are stochastically present biased and not fully sophisticated about this bias. The rational inattention model I propose for studying technology adoption provides an alternative channel to explain the observed behavior. In particular, Prediction 3 of my model implies that farmers may optimally decide to abstain from using fertilizer even if it is readily available and generally profitable when properly applied, because choosing profitable ways of applying fertilizer requires costly attention. This suggests that low uptake of fertilizer, including in the context studied by Duflo et al. (2011), may also be the result of fully rational behavior in the presence of constraints to information processing.

Another stylized observation is suboptimal technology usage given adoption. This includes farmers not adapting their practices perfectly to idiosyncratic conditions and shocks (Morris et al., 2007; Duflo et al., 2008), and failing to allocate available resources efficiently (Gollin and Udry, 2019). These findings are difficult to explain based on standard models in which farmers are perceived as rational profit maximizers with full information (e.g., Schultz, 1964; Bardhan and Udry, 1999). Moreover, classical learning models are in general only able to account for such imperfect usage in the short run, that is, for newly available technologies (recall the discussion in Subsection 4.2). Green Revolution technologies such as inorganic fertilizer, pesticides, and improved seeds, however, have been available in Africa already for several decades. Indeed, many authors argue that learning effects are therefore unlikely to be an important factor behind imperfect usage of these technologies observed in Africa today (Duflo et al., 2010; Foster and Rosenzweig, 2010; Suri, 2011).<sup>59</sup> The channel I propose for explaining imperfect technology adoption offers a possible way to rationalize suboptimal usage of modern farm inputs based on scarce attention and a (recurring) need to attend to idiosyncratic conditions that determine optimal usage (Prediction 4). Importantly, my model does not imply that suboptimal usage practices will necessarily improve over time (see the respective discussion in Subsection 4.2). Unlike classical learning models, therefore, the channel I study offers a possibility to account for suboptimal usage of modern farm inputs also in the long run.

Finally, another observation is that farmers frequently switch back and forth between using and not using adopted technologies (Dercon and Christiaensen, 2011; Duflo et al., 2011; Suri, 2011).<sup>60</sup> Suri (2011) argues that this switching behavior may be driven by changes in external factors (such as prices and availability) affecting the profitability of modern farm inputs. My model complements this view by showing that such switching behavior may also be the result

<sup>57</sup> This conclusion is mainly based on a series of field experiments conducted by Duflo et al. (2008, 2011), who show that fertilizer use is generally very profitable in the studied area, and that farmers are very aware of the benefits of using fertilizer. Specifically, Duflo et al. (2011) report that when farmers are asked directly, less than 2% of them respond that they believe fertilizer to be unprofitable. Nevertheless, many farmers do not use fertilizer, including when it is readily available for purchase.

<sup>58</sup> In particular, these authors stress that fertilizer has been available in the studied area for many decades, is a highly divisible technology, and features large returns at small levels of investment. According to additional results presented in Duflo et al. (2008), risk aversion may be able to explain underutilization of fertilizer, but it cannot account for nonadoption.

<sup>59</sup> Duflo et al. (2010) also point out the long history of large-scale extension work in Africa aimed at teaching farmers how to use modern farm inputs. When estimating learning effects in fertilizer use directly, these authors find that individual learning is very small and there is no evidence of diffusion to friends and neighbors. Conley and Udry (2010) find significant learning effects for a newly introduced technology in Ghana.

<sup>60</sup> Duflo et al. (2011) report that the  $R^2$  of a regression of fertilizer use on use during the previous year as well as a full vector of controls is only 0.25. Suri (2011) documents significant switching behavior in a nationally representative sample of maize farmers in Kenya.

of fluctuations in household-internal factors that affect farmers' shadow price of attention (see the discussion of a simple dynamic extension of the model in Subsection 3.4). This gives rise to the possibility to rationalize switching behavior also in situations where fluctuations in external factors on their own are not sufficient to explain observed changes in input practices.

Overall, the predictions of the model appear to be consistent with empirical adoption decisions of agricultural technologies in the development context, and the assumptions underlying the construction of the model are supported by empirical features of these technologies (see Section 2). However, I stress that many of the considered observations may also be related to other channels that affect farmers' input choices, and the direct evidence on the empirical role of limited attention in determining agricultural input choices is very limited.<sup>61</sup> I thus clearly do not wish to argue that the channel proposed in this article constitutes the primary force behind farmers' input choices. Rather, I suggest that costly information processing may be one of several pieces of the puzzle to explain imperfect technology adoption, which can complement the insights of existing theories in this context.

## 6. CONCLUSION

Understanding differences in technology adoption is key to understanding differences in living standards (Parente and Prescott, 1994; Keller, 2004). This article proposes a new explanation for low adoption rates and suboptimal usage of profitable technologies. The explanation centers around the idea that adoption decisions are affected by the anticipated (opportunity) cost users face when processing information required to choose adequate parameters of usage. I show that these "attentional barriers" to technology adoption constitute a new channel to account for empirical patterns in technology diffusion, complementing the insights of existing theories. The proposed explanation has three features: (i) the returns to new technology depend on accompanying parameter choices users are required to make; (ii) applying adopted technology in profitable ways requires users to be repeatedly attentive to idiosyncratic conditions that determine optimal parameter choices; (iii) due to limited information processing capacity, users are unable to attend to all relevant conditions and are therefore at risk of selecting suboptimal practices. The article identifies circumstances under which individuals or firms will optimally choose to abstain from using profitable and economically feasible technologies.

I study a rational inattention model in which agents decide how to allocate their scarce mental resources between different actions and whether they should adopt a particular technology. I find that nonadoption of the technology can be an optimal outcome if agents' opportunity cost of attention is too large to make using the technology profitable in expectation. In addition, the model can explain why users apply complementary technologies separately across different production locations, and occasionally use adopted technology in suboptimal ways. The model also generates a discontinuity in the amount of attention allocated among multiple decision problems, which, to the best of my knowledge, is a novel feature in the literature on rational inattention.

The model applies primarily to technologies for which optimal usage practices are largely determined by idiosyncratic conditions (instead of fixed target values). The model therefore complements existing learning models commonly used to explain imperfect technology diffusion, because these models focus on knowledge about fixed target values and abstract from the possibility of adjusting usage practices to conditions that are uncorrelated over time. In

<sup>61</sup> In particular, I am unable to assess the empirical plausibility of Prediction 2 as this would require data on farmers' allocation of attention across different plots that are, as far as I am aware, not available. At most, one might argue that the rational inattention channel is indirectly supported by field experiments that show that providing farmers with relevant information in preprocessed form (e.g., based on personalized summary statistics and reminders) improves input practices significantly more than only providing mere access to the same information (Cole and Fernando, 2012; Casaburi et al., 2014; Hanna et al., 2014). In particular, such interventions should only have an impact on input choices if farmers are indeed devoting attention to such information and are reactive to changes in the attentional cost associated with processing information.

some contexts, authors have recently argued that explaining imperfect adoption decisions requires the presence of behavioral biases, such as procrastination and limited self-control (DellaVigna, 2009; Duflo et al., 2011; Bernheim et al., 2015). The implications of my model suggest that some of the observations discussed in these studies may also be the result of fully rational behavior in the presence of constraints to information processing.

The model gives rise to several implications. Overall, the model suggests a more nuanced approach to policies aimed at promoting uptake and efficient use of modern technology. For example, in the context of modern farming technology, many governments focus on price subsidies to increase the adoption of fertilizer and pesticide (see Duflo et al., 2011). Such programs are often seen as increasing the risk of overuse of inputs, which can lead to negative environmental consequences such as soil degradation (FAO, 2019). In contrast, programs that can help farmers achieve higher profits by reducing uncertainty about optimal ways of using these inputs may be able to increase adoption rates without increasing the risk of overuse. Different from existing theories in this context, for example, those based on classical learning models, the limited attention channel implies that it may not be sufficient to teach people once how to optimally operate a new technology. Rather, a continuous form of support that helps users to adjust their practices to changes in the idiosyncratic conditions in which they operate may be needed to facilitate diffusion and efficient use of technology.

This article focuses on a stylized decision problem of a single agent with symmetric actions and independent states. Future work concerned with developing a more general theory of technology adoption under costly information processing might consider embedding some of the presented ideas into a dynamic rational inattention framework, and allowing for interactions of multiple agents (including the possibility of social learning). Although I do not undertake a structural test of the model, many of the model's implications are consistent with a broad set of empirical evidence, particularly on the diffusion of modern agricultural technology in developing economies. However, as the empirical evidence on attention constraints to technology diffusion is weak, any statement about the quantitative role of attentional barriers in determining imperfect adoption behavior would be speculative. I thus do not wish to argue that the mechanism proposed in this article constitutes the primary force behind (imperfect) empirical adoption decisions. Rather, I suggest that it can complement the insights of existing theories and should be considered as an additional channel to contribute to explanations of observed patterns in technology diffusion.

#### ACKNOWLEDGEMENTS

Open access funding enabled and organized by Projekt DEAL.

#### APPENDIX

**A.1 Solving the Model.** This appendix is organized as follows: Subsection A.1.1 derives the solution to the baseline model presented in Subsections 3.1–3.3. Subsection A.1.2 presents additional derivations for the basic setup. Subsection A.1.3 derives the solution to the decision problem with complementary input technologies presented in Subsection 3.4. Appendix A.1.4 presents the proof of Corollary 1.

**A.1.1 Solution to the baseline model.** This section derives the solution for a slightly more general case than the basic setup presented in Subsection 3.1, where each of the two choice problems  $k \in \{\tau, -\tau\}$  is allowed to feature  $i \in \{1, \dots, N_k\}$  options (instead of just two options). The solution to the basic setup with  $i \in \{low, high\}$  is included as a special case with  $N_k = 2$ .

With  $i \in \{1, \dots, N_k\}$  options, let  $v_{k,i}$  denote the payoff associated with option  $i$  for problem  $k \in \{\tau, -\tau\}$ . Furthermore, let  $\mathbf{v}_k = (v_{k,1}, \dots, v_{k,N_k})$  denote the  $(1 \times N_k)$  vector of possible payoffs for  $k$ . For each  $k$ ,  $\mathbf{v}_k$  depends on the realization of the state  $z_k$ , which is initially unobserved by the agent. I focus on the case where each state is realized with equal probability and states are independent of each other. To ensure that both adoption and non-adoption are

relevant alternatives for the agent, the payoff structure is such that (i)  $v_\tau$  contains at least one element that is larger than the safe return of the outside option (which is normalized to zero), and (ii) randomly choosing one of the  $N_\tau$  options is associated with an expected payoff below the return of the outside option. For  $k = -\tau$ , options take different values with positive probability as otherwise the behavior of the agent for these decisions would be irrelevant, and there is no outside option available for  $-\tau$ . A simple payoff structure that features these properties is given by the following specification:

$$(A.1) \quad v_{k,i} = \begin{cases} a_k, & z_k = i \\ -b_k, & z_k \in \{1, \dots, N_k\} \setminus \{i\}, \end{cases}$$

where each state  $z_k \in \{1, \dots, N_k\}$  is realized with equal probability. In this symmetric case, the expected payoff of randomly selecting an option for decision  $k$  is given by  $E[v_{k,i}] = \frac{1}{N_k}[a_k - b_k(N_k - 1)]$ , such that conditions (i) and (ii) are fulfilled as long as

$$(A.2) \quad -b_{-\tau} < a_{-\tau}$$

and

$$(A.3) \quad -b_\tau < 0 < a_\tau < (N_\tau - 1)b_\tau.$$

This specification offers a simple interpretation. In each possible state, there exists exactly one adequate way of using the technology. Choosing the correct option leads to a payoff of  $a_\tau$  that is larger than the secure payoff under nonadoption. Choosing a suboptimal option yields a payoff of  $-b_\tau < 0$ . This corresponds to the idea that using a more complex technology (captured by a higher number  $N_\tau$  of different parameter choices) can be profitable, but at the same time requires more precise choices about the parameters of usage to lead to the desired outcome.

**Reformulated problem.** The agent’s optimization problem with  $i \in \{1, \dots, N_k\}$  options is given analogously to expressions (5)–(5.3) in Subsection 3.3, with the payoff structure specified by expressions (A.1)–(A.3). Given that the agent can choose among all possible joint distributions  $\Gamma(s_k, z_k)$  of signals and states, deriving the optimal signal structure is in general very difficult. In solving the model, I therefore follow the approach taken by Matějka and McKay (2015) and calculate the agent’s optimal allocation of attention as the solution to a reformulated problem that is based on state-contingent choice probabilities instead of signals.<sup>62</sup> Specifically, these authors show that in the case of a rationally inattentive decision maker facing a discrete choice problem, every signal structure along with a prior belief induces a joint distribution between states and chosen options. Let

$$(A.4) \quad \mathcal{P}_{k,i}(v) = \Pr(i \mid \vec{v}_k)$$

denote the induced probability of selecting option  $i$  in decision problem  $k$  conditional on realized values  $\vec{v}_k$ , that is, after receiving the signal. Furthermore, let

$$(A.5) \quad \mathcal{P}_{k,i}^0 = \int_V \mathcal{P}_{k,i}(v)G(dV)$$

<sup>62</sup> Matějka and McKay (2015) show that the reformulated decision problem leads to the same optimal allocation of attention and taken actions as the original problem. This applies both to models where the cost of processing information is proportional to Shannon’s mutual information between the signal and the state (as in the model of Matějka and McKay, 2015) and to models where the agent is constrained to using signals whose mutual information is below some fixed capacity, as in Sims (2003) and in the model studied here (the capacity-based approach represents the dual problem of the cost-based version when adjusted by an appropriate number multiplying mutual information; see Matějka and McKay, 2015, p. 282).



denote the *ex ante* probability of selecting option  $i$  in decision problem  $k$  (before any information is processed). Using this approach, the agent's optimization problem can be reformulated as:

$$(A.6) \quad \max_{\{\mathcal{P}_{k,i}(\mathbf{v})\}_{i=1}^{N_k}} \sum_{k \in \{\tau, -\tau\}} \left( \sum_{i=1}^{N_k} \int_V v_{k,i} \mathcal{P}_{k,i}(\mathbf{v}) G(dV) \right),$$

$$(A.7) \quad \text{s.t. } H(G(\mathbf{v}_k)) - E[H(\Gamma(\mathbf{v}_k))] \leq \kappa_k \quad \forall k \in \{\tau, -\tau\},$$

$$(A.8) \quad \sum_{k \in \{\tau, -\tau\}} \mu \kappa_k \leq \bar{\kappa},$$

$$(A.9) \quad \sum_{i=1}^{N_k} \mathcal{P}_{k,i}(\mathbf{v}) = 1 \quad \forall k \in \{\tau, -\tau\}.$$

In the transformed problem, the agent directly sets the choice probabilities  $\mathcal{P}_{k,i}(\mathbf{v})$  for each decision  $k \in \{\tau, -\tau\}$  so that it is not necessary anymore to explicitly model signals. Conditions (A.7) and (A.8) are equivalent to conditions (5.2) and (5.3) in the original problem. In addition, condition (A.9) states that the conditional probabilities of all options must sum up to one for each choice problem. The solution to the transformed problem is the information strategy, that is, the set of probabilities  $\{\mathcal{P}_{k,i}(\mathbf{v})\}_{i=1}^{N_k}$  with  $\mathcal{P}_{k,i}^0 = \int_V \mathcal{P}_{k,i}(\mathbf{v}) G(dV)$  for each  $k \in \{\tau, -\tau\}$ , which maximizes the sum of expected payoffs subject to the given constraints.

Note that, as long as attention is scarce, that is, condition (A.8) is binding, the restrictions (A.7) and (A.8) must hold with equality, since otherwise attention would be wasted. In this case, using the mathematical definition of entropy in condition (A.7), the amount of attention allocated to the choice problem  $k \in \{\tau, -\tau\}$  can be written as

$$(A.10) \quad \kappa_k = - \sum_{i=1}^{N_k} \mathcal{P}_{k,i}^0 \log \mathcal{P}_{k,i}^0 + \int_V \sum_{i=1}^{N_k} \mathcal{P}_{k,i}(\mathbf{v}) \log \mathcal{P}_{k,i}(\mathbf{v}) G(dV).$$

Plugging this expression into condition (A.8) yields

$$(A.11) \quad \bar{\kappa} = - \sum_{k \in \{\tau, -\tau\}} \mu \left( \sum_{i=1}^{N_k} \mathcal{P}_{k,i}^0 \log \mathcal{P}_{k,i}^0 - \int_V \sum_{i=1}^{N_k} \mathcal{P}_{k,i}(\mathbf{v}) \log \mathcal{P}_{k,i}(\mathbf{v}) G(dV) \right).$$

Equation (A.11) can be used to replace conditions (A.7) and (A.8) in the reformulated decision problem. The Lagrangian of the resulting optimization problem is then given by

$$\begin{aligned} \max_{\{\mathcal{P}_{k,i}(\mathbf{v})\}_{i=1}^{N_k}} \mathcal{L} = & \sum_{k \in \{\tau, -\tau\}} \sum_{i=1}^{N_k} \int_V v_{k,i} \mathcal{P}_{k,i}(\mathbf{v}) G(dV) \\ & + \lambda \left[ \bar{\kappa} + \sum_{k \in \{\tau, -\tau\}} \mu \left( \sum_{i=1}^{N_k} \mathcal{P}_{k,i}^0 \log \mathcal{P}_{k,i}^0 - \int_V \sum_{i=1}^{N_k} \mathcal{P}_{k,i}(\mathbf{v}) \log \mathcal{P}_{k,i}(\mathbf{v}) G(dV) \right) \right] \end{aligned}$$

$$(A.12) \quad - \sum_{k \in \{\tau, -\tau\}} \int_V \phi_k(v) \left( \sum_{i=1}^{N_k} \mathcal{P}_{k,i}(v) - 1 \right) G(dV),$$

where  $\lambda$  and  $\phi_k(v)$  denote Lagrange multipliers.

**Optimal choice probabilities.** Taking the derivative of the Lagrangian (A.12) with respect to  $\mathcal{P}_{k,i}(v)$  gives a first-order condition that can be further transformed along the same steps as shown in Matějka and McKay (2015).<sup>63</sup> This leads to the optimality condition for the conditional choice probabilities

$$(A.13) \quad \mathcal{P}_{k,i}(v) = \frac{\mathcal{P}_{k,i}^0 e^{v_{k,i}/\lambda\mu}}{\sum_{j=1}^{N_k} \mathcal{P}_{k,j}^0 e^{v_{k,j}/\lambda\mu}}.$$

Equation (A.13) implies that for each  $k \in \{\tau, -\tau\}$ , the optimal probability of selecting option  $i$  conditional on realized values  $v_k$  depends positively on the prior belief about option  $i$ 's payoff relative to the payoff of other options. As the cost of processing information increases, less attention is allocated to the choice problem and the weight of the prior rises. This is the case if  $\mu$  is larger, that is, it is more demanding to process information, or if  $\lambda$  increases due to a smaller attentional capacity  $\bar{\kappa}$  (holding other parameters fixed).

Note that the optimality condition (A.13) applies both under adoption and under nonadoption of the technology. If nonadoption is chosen, the agent obtains a secure payoff of zero for  $k = \tau$  and uses all attentional capacity for  $-\tau$ . The optimization problem in this case takes the same form as the Lagrangian (A.12), except that the terms associated with  $k = \tau$  do not appear.

Equation (A.13) still depends on the unconditional probabilities  $\mathcal{P}_{k,i}^0$ . Before showing how the optimal values for these probabilities can be derived, note that if attention is not scarce, that is, constraint (A.8) is not binding, then the shadow price of attention captured by the multiplier  $\lambda$  in the Lagrangian (A.12) equals 0 and the term stemming from Equation (A.11) does not appear in the optimization problem. In this special case (which corresponds to the standard case of full information), the agent is perfectly attentive to all available information and able to reduce uncertainty about the optimal option for each decision to 0. The agent is thus able to select the optimal option with probability 1 for each decision. Given the payoff structure specified by expressions (A.1)–(A.3), it follows that the agent will adopt the technology and use it under optimal parameter choices.

If attention is scarce, that is, constraint (A.8) is binding, then the agent decides how to allocate the available capacity  $\bar{\kappa}$  between the different decisions. In this case, the agent's optimal behavior features the unconditional probabilities

$$(A.14) \quad \left( \mathcal{P}_{\tau,0}^{0,*}, \mathcal{P}_{\tau,1}^{0,*}, \dots, \mathcal{P}_{\tau,N_\tau}^{0,*} \right) \in \left\{ \left( 0, \frac{1}{N_\tau}, \dots, \frac{1}{N_\tau} \right), (1, 0, \dots, 0) \right\}$$

$$(A.15) \quad \mathcal{P}_{-\tau,i}^{0,*} = \frac{1}{N_{-\tau}} \quad \forall i \in \{1, \dots, N_{-\tau}\},$$

where  $\mathcal{P}_{\tau,0}^{0,*}$  denotes the probability of choosing the outside option of non-adoption of the technology. The results in Equations (A.14) and (A.15) follow directly from the results derived in Matějka and McKay (2015, Appendix B). To see this, first, notice that the assumed payoff structure specified in Equation (A.1) implies *a priori* homogeneous options, because

<sup>63</sup> Notice that in Matějka and McKay (2015), the parameter  $\lambda$  denotes a fixed unit cost of attention, whereas in the capacity-based approach taken, here  $\lambda$  is a Lagrange parameter that depends on the values of the other parameters of the model.

all options seem initially identical to the agent before processing any information (i.e.,  $G(v_k)$  is invariant to permutations of the entries of  $v_k$ ). As shown by Matějka and McKay (2015), this implies that the optimal unconditional probabilities are uniform. It thus follows that  $\mathcal{P}_{k,i}^{0,*}$  must be equal to  $\frac{1}{N_k}$  for each relevant option. For the choice problem associated with technology use, this result applies if the agent decides to adopt the technology. In case of non-adoption, the agent selects the outside option directly by setting  $\mathcal{P}_{k,0}^{0,*}$  equal to 1, and thus, all other unconditional probabilities equal to 0. Appendix A.1.2 shows how the results in Equations (A.14) and (A.15) can be explicitly derived for the case with two binary choice problems ( $N_k = 2$ ) considered in the basic setup in Subsection 3.1.

Equations (A.14) and (A.15) imply that there is a simple solution for the agent's optimal behavior, characterized by two possible cases. The first set of probabilities specified in expression (A.14) corresponds to the case where the agent chooses to adopt the technology. In this case, each possible way of using the technology features the same *ex ante* probability of being selected. This result is intuitive, as each option is initially equally likely to lead to the desired maximum payoff. In the second case, the agent chooses nonadoption and assigns all probability to the outside option.

Several important implications follow from these results. First, the agent will never adopt the technology and use it only under a subset of the possible options. Instead, if the agent decides to use the technology, then all options that might potentially be adequate feature a positive *ex ante* probability of being selected.

Second, it is never optimal for the agent to pay attention to the technology but then decide not to adopt it. This follows from the fact that there is no case in expression (A.14) in which the agent assigns positive probabilities to both the outside option and any of the other  $N_\tau$  options. The intuition behind this result is that using the technology in adequate ways does always offer larger returns than the outside option, and the agent knows in advance how much attention is needed to reduce uncertainty about optimal usage practices by enough to make adoption profitable in expectation.<sup>64</sup>

**Optimal allocation of attention and expected payoff.** Based on the optimal probabilities specified in Equations (A.14) and (A.15), analytical expressions for the agent's optimal allocation of attention and the associated expected payoff of each choice problem can be derived. The results are summarized in the following proposition:

**PROPOSITION A.1.** *If the agent uses the technology, the optimal amount of attention allocated to each choice problem  $k \in \{\tau, -\tau\}$  is given by*

$$(A.16) \quad \kappa_k^*(\lambda) = \log(N_k) + \frac{e^{\frac{a}{\lambda\mu}}}{\Theta} \log\left(\frac{e^{\frac{a}{\lambda\mu}}}{\Theta}\right) + (N_k - 1) \frac{e^{\frac{-b}{\lambda\mu}}}{\Theta} \log\left(\frac{e^{\frac{-b}{\lambda\mu}}}{\Theta}\right),$$

where  $\Theta = e^{\frac{a}{\lambda\mu}} + (N_k - 1)e^{\frac{-b}{\lambda\mu}}$  and  $\lambda > 0$  is the agent's shadow price of attention. The expected payoff for each choice problem is given by

$$(A.17) \quad E[V_k(\kappa_k^*)] = \frac{a e^{\frac{a}{\lambda\mu}} - (N_k - 1) b e^{\frac{-b}{\lambda\mu}}}{e^{\frac{a}{\lambda\mu}} + (N_k - 1) e^{\frac{-b}{\lambda\mu}}} \quad \forall k \in \{\tau, -\tau\}.$$

*If the technology is not adopted, the agent sets  $\kappa_\tau^*$  equal to 0 and allocates all available attention to  $-\tau$ . In this case, the expected payoff from  $-\tau$  features the same functional form as in Equation (A.17) but with a different value of  $\lambda$ .*

<sup>64</sup> This result differs from the solution to the problem with asymmetric prior considered by Matějka and McKay (2015), where for most safe returns of the outside option, the agent processes information and *ex ante* mixes between the uncertain option and the outside option.

PROOF. Starting with Equation (A.10), that is,

$$(A.18) \quad \kappa_k = - \sum_{i=1}^{N_k} \mathcal{P}_{k,i}^0 \log \mathcal{P}_{k,i}^0 + \int_V \sum_{i=1}^{N_k} \mathcal{P}_{k,i}(v) \log \mathcal{P}_{k,i}(v) G(dV),$$

plugging in the optimal values of  $\mathcal{P}_{k,i}^{0,*}$  from Equations (A.14) and (A.15) in the case of adoption shows that the first summand in Equation (A.18) can be reduced to

$$(A.19) \quad - \sum_{i=1}^{N_k} \frac{1}{N_k} \log \left( \frac{1}{N_k} \right) = \log(N_k).$$

The second summand in Equation (A.18) can be simplified as follows: First, plugging the values of  $\mathcal{P}_{k,i}^{0,*}$  from Equations (A.14) and (A.15) into the optimality condition (A.13) shows that the optimal conditional probabilities can be written as

$$(A.20) \quad \mathcal{P}_{k,i}^*(v) = \frac{e^{v_{k,i}/\lambda\mu}}{\sum_{j=1}^{N_k} e^{v_{k,j}/\lambda\mu}}.$$

Since each state is associated with exactly one option for each decision that yields the maximum return, the sum in the denominator leads to the same expression for every realization of the fundamental. Therefore, using the distribution of payoffs specified in Equation (A.1), the optimal conditional probabilities are given by

$$(A.21) \quad \mathcal{P}_{k,i}^*(v) = \begin{cases} \frac{e^{\frac{a}{\lambda\mu}}}{e^{\frac{a}{\lambda\mu}} + (N_k - 1) e^{\frac{-b}{\lambda\mu}}} & \text{if } z_k = i \\ \frac{e^{\frac{-b}{\lambda\mu}}}{e^{\frac{a}{\lambda\mu}} + (N_k - 1) e^{\frac{-b}{\lambda\mu}}} & \text{if } z_k \in \{1, \dots, N_k\} \setminus \{i\}. \end{cases}$$

To lighten notation, let  $\Theta = e^{\frac{a}{\lambda\mu}} + (N_k - 1) e^{\frac{-b}{\lambda\mu}}$ . Since options are symmetric, it holds that for each possible realization of  $v_k$ , the second sum in Equation (A.18) amounts to

$$(A.22) \quad \sum_{i=1}^{N_k} \mathcal{P}_{k,i}(v) \log \mathcal{P}_{k,i}(v) = \frac{e^{\frac{a}{\lambda\mu}}}{\Theta} \log \left( \frac{e^{\frac{a}{\lambda\mu}}}{\Theta} \right) + (N_k - 1) \frac{e^{\frac{-b}{\lambda\mu}}}{\Theta} \log \left( \frac{e^{\frac{-b}{\lambda\mu}}}{\Theta} \right).$$

Together, using Equations (A.19) and (A.22) to simplify Equation (A.18) shows that  $\kappa_k^*$  is given by the expression stated in Proposition A.1.

The associated expected payoff can be derived as follows: Plugging the optimal conditional probabilities from Equation (A.20) into the corresponding part of the agent's objective function (A.6) leads to

$$(A.23) \quad E[V_k(\kappa_k^*)] = \sum_{i=1}^{N_k} \int_V v_{k,i} \frac{e^{\frac{v_{k,i}}{\lambda\mu}}}{e^{\frac{a}{\lambda\mu}} + (N_k - 1) e^{\frac{-b}{\lambda\mu}}} G(dV).$$

Since each option yields the desired payoff  $a$  for exactly one realization of the fundamental state and otherwise yields the payoff  $-b$ , integrating over all states leads to the same expression for every option, such that

$$(A.24) \quad E[V_k(\kappa_k^*)] = \sum_{i=1}^{N_k} \frac{1}{N_k} \frac{a e^{\frac{a}{\lambda\mu}}}{e^{\frac{a}{\lambda\mu}} + (N_k - 1) e^{\frac{-b}{\lambda\mu}}} + \frac{N_k - 1}{N_k} \frac{-b e^{\frac{-b}{\lambda\mu}}}{e^{\frac{a}{\lambda\mu}} + (N_k - 1) e^{\frac{-b}{\lambda\mu}}}.$$

Using again the symmetry of options, it follows that the expected payoff of each decision under optimal allocation of attention is given by the expression stated in Proposition A.1, that is,

$$(A.25) \quad E[V_k(\kappa_k^*)] = \frac{a e^{\frac{a}{\lambda\mu}} - (N_k - 1) b e^{\frac{-b}{\lambda\mu}}}{e^{\frac{a}{\lambda\mu}} + (N_k - 1) e^{\frac{-b}{\lambda\mu}}}.$$

In the case of nonadoption, the agent sets  $\kappa_\tau^* = 0$  and allocates all available attention to  $-\tau$ . In this case,  $E[V_{-\tau}(\kappa_{-\tau}^*)]$  can be obtained analogously to the case of adoption, resulting in an expression of the same functional form as in Equation (A.25) but with a different value of  $\lambda$ . This completes the proof of Proposition A.1.  $\square$

**Cutoff value for adoption ( $m_\tau$ ).** Equation (A.25) can be used to show that the agent’s optimal behavior is characterized by a cutoff value specifying the minimum amount of endowed attention  $\bar{\kappa}$  for which adopting the technology is optimal (corresponding to  $m_\tau$  in Equation (6)). Notice that the optimal allocation of attention and associated expected payoffs specified in Proposition A.1 depend on the opportunity cost of attention, captured by the Lagrange multiplier  $\lambda$ . When other parameters are fixed,  $\lambda$  is fully determined by the value of  $\bar{\kappa}$ . To see this, one can insert the expression for  $\kappa_k^*(\lambda)$  from Equation (A.16) into the budget constraint of attention (A.8), which yields an equation that contains only  $\lambda$ ,  $\bar{\kappa}$ , and the other primitives of the model (i.e.,  $a, b, \mu, N_k$ ). The existence of a unique threshold in attentional capacity that determines whether adoption or nonadoption is optimal follows from three properties. First, as shown below, the expected payoff of using the technology is monotonically increasing in  $\bar{\kappa}$  (decreasing in  $\lambda$ ). Second, if  $\bar{\kappa} = 0$ , the expected payoff of using the technology is smaller than the secure payoff of nonadoption. This follows directly from the payoffs specified in Equation (A.1) under the made parameter restrictions (A.3). Third, for sufficiently large values of  $\bar{\kappa}$ , the expected payoff under adoption is larger than the expected payoff of nonadoption. In particular, under perfect attention (i.e., when  $\bar{\kappa}$  is large enough to lead to  $\lambda = 0$ ), the agent can select the optimal option with certainty, such that the realized payoff under adoption equals  $a > 0$  (in the basic setup with two options, this corresponds to the first case in Equation (6), that is, when  $\bar{\kappa} \geq \bar{m}_\tau$ ). Together, these three properties imply that, for given parameter values ( $a, b, \mu, N_k$ ), there is exactly one value for  $\bar{\kappa}$  at which the expected payoffs under adoption and under nonadoption cross. This is the cutoff value denoted by  $m_\tau$  in Equation (6) (the other cutoff in Equation (6),  $\bar{m}_\tau$ , is defined in footnote 25).

The monotonicity of expected payoffs with respect to scarcity of attention can be derived as follows: Starting with Equation (A.25), dividing both numerator and denominator by  $e^{\frac{a}{\lambda\mu}}$  leads to

$$(A.26) \quad E[V_k(\kappa_k^*)] = \frac{a - (N_k - 1)b e^{\frac{-(a+b)}{\lambda\mu}}}{1 + (N_k - 1)e^{\frac{-(a+b)}{\lambda\mu}}}.$$

Taking the first derivative with respect to  $\lambda$  and simplifying the resulting expression yields

$$(A.27) \quad \frac{\partial E[V_k(\kappa_k^*)]}{\partial \lambda} = -\frac{(N_k - 1)(a + b)^2 \log(2)}{\lambda^2 \mu} e^{\frac{-(a+b)}{\lambda\mu}} \left(1 + e^{\frac{-(a+b)}{\lambda\mu}}\right)^{-2}.$$

As the derivative is negative (recall that  $N_k \geq 2$  and  $a, b, \mu > 0$ ), the expected payoff of using the technology is monotonically decreasing in  $\lambda$ .

*A.1.2 Additional derivations for the basic setup.* This appendix shows how the optimal unconditional probabilities  $\mathcal{P}_{k,i}^0$  stated in Equations (A.14) and (A.15) in Appendix A.1.1 can be explicitly derived for the case with two binary choice problems ( $N_k = 2$ ) considered in the

basic setup in Subsection 3.1. Some of the results derived here will also be used in the solution to the model with complementary input technologies discussed in Appendix A.1.3.

As stated in Appendix A.1.1, the unconditional probabilities  $\mathcal{P}_{k,i}^0$  must be internally consistent in that they fulfill  $\mathcal{P}_{k,i}^0 = \int_V \mathcal{P}_{k,i}(v)G(dV)$ . Substituting the choice probabilities  $\mathcal{P}_{k,i}(v)$  with the optimal expression from Equation (A.13) and dividing both sides by  $\mathcal{P}_{k,i}^0$  leads to

$$(A.28) \quad 1 = \int_V \frac{e^{v_{k,i}/\lambda\mu}}{\sum_{j=1}^2 \mathcal{P}_{k,j}^0 e^{v_{k,j}/\lambda\mu}} G(dV) \quad \text{if } \mathcal{P}_{k,i}^0 > 0.$$

For  $k = -\tau$ , plugging the values of  $v_{-\tau,i}$  specified in expression (1) into Equation (A.28) leads to the following two conditions:

$$(A.29) \quad 1 = 0.5 \frac{e^{\frac{a}{\lambda\mu}}}{\mathcal{P}_{-\tau,l}^0 e^{\frac{a}{\lambda\mu}} + \mathcal{P}_{-\tau,h}^0 e^{\frac{-b}{\lambda\mu}}} + 0.5 \frac{e^{\frac{-b}{\lambda\mu}}}{\mathcal{P}_{-\tau,l}^0 e^{\frac{-b}{\lambda\mu}} + \mathcal{P}_{-\tau,h}^0 e^{\frac{a}{\lambda\mu}}} \quad \text{if } \mathcal{P}_{-\tau,l}^0 > 0,$$

$$(A.30) \quad 1 = 0.5 \frac{e^{\frac{-b}{\lambda\mu}}}{\mathcal{P}_{-\tau,l}^0 e^{\frac{a}{\lambda\mu}} + \mathcal{P}_{-\tau,h}^0 e^{\frac{-b}{\lambda\mu}}} + 0.5 \frac{e^{\frac{a}{\lambda\mu}}}{\mathcal{P}_{-\tau,l}^0 e^{\frac{-b}{\lambda\mu}} + \mathcal{P}_{-\tau,h}^0 e^{\frac{a}{\lambda\mu}}} \quad \text{if } \mathcal{P}_{-\tau,h}^0 > 0.$$

Since  $\mathcal{P}_{-\tau,l}^0$  and  $\mathcal{P}_{-\tau,h}^0$  must sum up to 1, there are three possible solutions to this pair of equations. One where both unconditional probabilities are greater than 0, and the two cases where either  $\mathcal{P}_{-\tau,l}^0$  or  $\mathcal{P}_{-\tau,h}^0$  is equal to 0. In the first case, combining Equations (A.29) and (A.30) leads to

$$(A.31) \quad 0 = \frac{e^{\frac{a}{\lambda\mu}} - e^{\frac{-b}{\lambda\mu}}}{\mathcal{P}_{-\tau,l}^0 e^{\frac{a}{\lambda\mu}} + \mathcal{P}_{-\tau,h}^0 e^{\frac{-b}{\lambda\mu}}} + \frac{e^{\frac{-b}{\lambda\mu}} - e^{\frac{a}{\lambda\mu}}}{\mathcal{P}_{-\tau,l}^0 e^{\frac{-b}{\lambda\mu}} + \mathcal{P}_{-\tau,h}^0 e^{\frac{a}{\lambda\mu}}}.$$

Expanding both fractions to the common denominator and simplifying the resulting expression yields

$$(A.32) \quad 0 = \left( e^{\frac{a}{\lambda\mu}} - e^{\frac{-b}{\lambda\mu}} \right)^2 (\mathcal{P}_{-\tau,h}^0 - \mathcal{P}_{-\tau,l}^0).$$

Since  $b > a$ , the equation only holds for  $\mathcal{P}_{-\tau,l}^0 = \mathcal{P}_{-\tau,h}^0$ . This result is intuitive, as the two options are symmetric. As the sum of both unconditional probabilities must be equal to 1, it follows that  $(\mathcal{P}_{-\tau,l}^{0,*}, \mathcal{P}_{-\tau,h}^{0,*}) = (\frac{1}{2}, \frac{1}{2})$  whenever  $\mathcal{P}_{-\tau,l}^0, \mathcal{P}_{-\tau,h}^0 > 0$ .

The two remaining cases, where either  $\mathcal{P}_{-\tau,l}^0$  or  $\mathcal{P}_{-\tau,h}^0$  equals 0, would correspond to situations in which the agent chooses one of the two options directly, that is, without processing any information for  $-\tau$ . For the setup in Subsection 3.1, this is never an optimal strategy. To see this, recall that the two choice problems  $\tau$  and  $-\tau$  feature symmetric payoff structures, except that for  $\tau$  there is a safe outside option available. Therefore, whenever the agent chooses to allocate attention only to one of the two problems, then this will be  $-\tau$  as there is no outside option for this problem. This means that, as long as  $\bar{\kappa} > 0$  (and no attention is wasted), the agent will allocate a positive amount of attention to  $-\tau$  by setting  $\mathcal{P}_{-\tau,l}^0, \mathcal{P}_{-\tau,h}^0 > 0$ . Thus,  $(\mathcal{P}_{-\tau,l}^{0,*}, \mathcal{P}_{-\tau,h}^{0,*}) = (\frac{1}{2}, \frac{1}{2})$  is the unique solution for  $k = -\tau$ .



For  $k = \tau$ , plugging the corresponding values of  $v_{\tau,i}$  into Equation (A.28) leads to the following three conditions:

$$(A.33) \quad 1 = 0.5 \frac{e^{\frac{a}{\lambda\mu}}}{\mathcal{P}_{\tau,l}^0 e^{\frac{a}{\lambda\mu}} + \mathcal{P}_{\tau,h}^0 e^{\frac{-b}{\lambda\mu}} + \mathcal{P}_{\tau,0}^0} + 0.5 \frac{e^{\frac{-b}{\lambda\mu}}}{\mathcal{P}_{\tau,l}^0 e^{\frac{-b}{\lambda\mu}} + \mathcal{P}_{\tau,h}^0 e^{\frac{a}{\lambda\mu}} + \mathcal{P}_{\tau,0}^0} \quad \text{if } \mathcal{P}_{\tau,l}^0 > 0,$$

$$(A.34) \quad 1 = 0.5 \frac{e^{\frac{-b}{\lambda\mu}}}{\mathcal{P}_{\tau,l}^0 e^{\frac{a}{\lambda\mu}} + \mathcal{P}_{\tau,h}^0 e^{\frac{-b}{\lambda\mu}} + \mathcal{P}_{\tau,0}^0} + 0.5 \frac{e^{\frac{a}{\lambda\mu}}}{\mathcal{P}_{\tau,l}^0 e^{\frac{-b}{\lambda\mu}} + \mathcal{P}_{\tau,h}^0 e^{\frac{a}{\lambda\mu}} + \mathcal{P}_{\tau,0}^0} \quad \text{if } \mathcal{P}_{\tau,h}^0 > 0,$$

$$(A.35) \quad 1 = 0.5 \frac{1}{\mathcal{P}_{\tau,l}^0 e^{\frac{a}{\lambda\mu}} + \mathcal{P}_{\tau,h}^0 e^{\frac{-b}{\lambda\mu}} + \mathcal{P}_{\tau,0}^0} + 0.5 \frac{1}{\mathcal{P}_{\tau,l}^0 e^{\frac{-b}{\lambda\mu}} + \mathcal{P}_{\tau,h}^0 e^{\frac{a}{\lambda\mu}} + \mathcal{P}_{\tau,0}^0} \quad \text{if } \mathcal{P}_{\tau,0}^0 > 0.$$

There are two solutions to this system of equations. To see this, first note that a solution will feature either  $\mathcal{P}_{\tau,0}^0 = 0$ ,  $\mathcal{P}_{\tau,0}^0 = 1$  or  $\mathcal{P}_{\tau,0}^0 \in (0, 1)$ . If  $\mathcal{P}_{\tau,0}^0 = 0$ , that is, the outside option is never chosen, then the problem reduces to a problem that is equivalent to the one for  $k = -\tau$ , and thus, the solution is given by  $(\mathcal{P}_{\tau,0}^{0,*}, \mathcal{P}_{\tau,l}^{0,*}, \mathcal{P}_{\tau,h}^{0,*}) = (0, \frac{1}{2}, \frac{1}{2})$ . In this case, the agent adopts the technology and devotes a positive amount of attention to  $\tau$  to distinguish between *low* and *high* as the adequate way of using the technology. If  $\mathcal{P}_{\tau,0}^0 = 1$ , then the solution is given by  $(\mathcal{P}_{\tau,0}^{0,*}, \mathcal{P}_{\tau,l}^{0,*}, \mathcal{P}_{\tau,h}^{0,*}) = (1, 0, 0)$  as the sum of the three probabilities must be equal to 1. In this case, the agent chooses not to use the technology, and therefore, no attention is devoted to  $\tau$ .

Finally, suppose that there was a solution that featured  $\mathcal{P}_{\tau,0}^0 \in (0, 1)$ . Since the sum of all three probabilities must be equal to 1, at least one of the other two probabilities,  $\mathcal{P}_{\tau,l}^0$  and  $\mathcal{P}_{\tau,h}^0$ , must be strictly larger than 0. In addition, the symmetry of payoffs for  $\tau$  implies that if the agent believes that option *low* is inadequate given the realized state, then automatically *high* is believed to be optimal (and vice versa). This is due to the fact that, in each possible state, either option *low* or option *high* leads to a payoff which is larger than the safe return under nonadoption, that is, there is no state in which the technology cannot be profitably used. Together, these features imply that it is never optimal to adopt the technology and restrict usage only to a single option. Consequently, if the technology is used, then both ways of doing so will have a positive *a priori* probability of being selected, such that  $\mathcal{P}_{\tau,l}^0, \mathcal{P}_{\tau,h}^0 > 0$ . If  $\mathcal{P}_{\tau,0}^0 \in (0, 1)$ , then a solution would thus have to fulfill all three Equations (A.33)–(A.35). Combining Equations (A.33) and (A.34) leads to an equation that shows that  $\mathcal{P}_{\tau,l}^0$  and  $\mathcal{P}_{\tau,h}^0$  must be equal. Using this result to replace  $\mathcal{P}_{\tau,l}^0$  in Equation (A.35) leads to

$$(A.36) \quad \mathcal{P}_{\tau,0}^0 = 1 - \mathcal{P}_{\tau,h}^0 \left( e^{\frac{a}{\lambda\mu}} - e^{\frac{-b}{\lambda\mu}} \right).$$

In addition, probabilities must sum up to 1 such that  $2\mathcal{P}_{\tau,h}^0 + \mathcal{P}_{\tau,0}^0 = 1$ . Using this to replace  $\mathcal{P}_{\tau,h}^0$  in Equation (A.36) gives

$$(A.37) \quad \mathcal{P}_{\tau,0}^0 = 1 - \frac{1 - \mathcal{P}_{\tau,0}^0}{2} \left( e^{\frac{a}{\lambda\mu}} - e^{\frac{-b}{\lambda\mu}} \right).$$

Solving Equation (A.37) gives  $\mathcal{P}_{\tau,0}^0 = 1$ , which is contradictory to the original presumption that  $\mathcal{P}_{\tau,0}^0 \in (0, 1)$ . Thus, there is no solution with  $\mathcal{P}_{\tau,0}^0 \in (0, 1)$ .

Summarizing, the optimal behavior for  $k = \tau$  is characterized by two possible sets of unconditional probabilities, namely,  $(\mathcal{P}_{\tau,0}^{0,*}, \mathcal{P}_{\tau,l}^{0,*}, \mathcal{P}_{\tau,h}^{0,*}) = (0, \frac{1}{2}, \frac{1}{2})$  if the agent adopts the technology and  $(\mathcal{P}_{\tau,0}^{0,*}, \mathcal{P}_{\tau,l}^{0,*}, \mathcal{P}_{\tau,h}^{0,*}) = (1, 0, 0)$  in the case of nonadoption. The solution for  $k = -\tau$  always

features  $(\mathcal{P}_{-\tau,l}^{0,*}, \mathcal{P}_{-\tau,h}^{0,*}) = (\frac{1}{2}, \frac{1}{2})$ . Note that these results coincide with the results for the more general case with  $i \in \{1, \dots, N_k\}$  options, as stated in Equations (A.14) and (A.15).

**A.1.3 Solution with Complementary Input Technologies.** This section derives the agent’s optimal behavior for the model with complementary input technologies presented in Subsection 3.4 (including the proof of Proposition 1). Notice first that the agent’s problem in Subsection 3.4 represents a special case of the problem solved in Appendix A.1.1 where each of the two choice problems (actions) features  $N_k = 9$  possible options, and the payoffs are specified in Table 1. The Langrangian of the reformulated decision problem with complementary inputs thus features the same functional form as the one in Appendix A.1.1 (albeit with a different payoff structure). In particular, it holds that the amount of attention allocated to location  $j$  is analogous to the expression stated in Equation (A.10) and can be written as

$$(A.38) \quad \kappa_j = - \sum_{i \in \mathcal{I}} \mathcal{P}_{j,i}^0 \log \mathcal{P}_{j,i}^0 + \int_V \sum_{i \in \mathcal{I}} \mathcal{P}_{j,i}(v) \log \mathcal{P}_{j,i}(v) G(dV).$$

Furthermore, Equation (A.13) implies that the optimal choice probabilities are given by

$$(A.39) \quad \mathcal{P}_{j,i}^*(v) = \frac{\mathcal{P}_{j,i}^0 e^{v_{j,i}/\lambda\mu}}{\sum_{l \in \mathcal{K}} \mathcal{P}_{j,l}^0 e^{v_{j,l}/\lambda\mu}},$$

where  $\lambda$  denotes the Lagrange multiplier of the budget constraint for the scarce resource attention. As above, using the fact that the unconditional probabilities  $\mathcal{P}_{k,i}^0$  must be internally consistent, it follows that the agent’s optimal behavior must fulfill the condition

$$(A.40) \quad 1 = \int_V \frac{e^{v_{j,i}/\lambda\mu}}{\sum_{l \in \mathcal{K}} \mathcal{P}_{j,l}^0 e^{v_{j,l}/\lambda\mu}} G(dV) \quad \text{if } \mathcal{P}_{j,i}^0 > 0.$$

In the following, I first use the payoffs specified in Table 1 to derive the agent’s optimal behavior separately for each of the three types of actions (i.e., nonadoption, single input use, and joint input use) for a single production location. In a second step, these results are then aggregated to obtain the agent’s optimal behavior across two locations.

**Case 1: Nonadoption.** If the agent chooses not to adopt any complementary input in location  $j$ , the outside option  $i = (0, 0)$  is selected by setting  $\mathcal{P}_{j,(0,0)}(v)$  equal to 1 for all states and the probabilities of all other actions equal to 0. In this case, the conditional and unconditional probabilities are identical, and the secure return of 0 is realized with certainty. Plugging these probabilities into Equation (A.38) shows that the amount of attention allocated to the plot equals 0.

**Case 2: Single input use.** If a single input is used in a location, there are two possible ways of applying the input, given by  $i \in \{(S_l, 0), (S_h, 0)\}$  or  $i \in \{(0, F_l), (0, F_h)\}$ , depending on which input is chosen (recall that inputs are assumed to be homogeneous). After processing information, the agent seeks to select the option that best conforms to the realized state of the location. The optimal behavior at a given location under single input use is summarized by Lemma A.1.

LEMMA A.1. *The optimal behavior of a rationally inattentive agent who adopts a single input technology ( $S$ ) in location  $j$  is characterized by the choice probabilities*

$$\left( \mathcal{P}_{j,(S_l,0)}^*(\mathbf{v}), \mathcal{P}_{j,(S_h,0)}^*(\mathbf{v}) \right) = \begin{cases} \left( \frac{e^{\frac{a}{\lambda\mu}}}{e^{\frac{a}{\lambda\mu}} + e^{\frac{-b}{\lambda\mu}}}, \frac{e^{\frac{-b}{\lambda\mu}}}{e^{\frac{a}{\lambda\mu}} + e^{\frac{-b}{\lambda\mu}}} \right) & \text{if } z_j \in \{(l, l), (l, h)\} \\ \left( \frac{e^{\frac{-b}{\lambda\mu}}}{e^{\frac{-b}{\lambda\mu}} + e^{\frac{a}{\lambda\mu}}}, \frac{e^{\frac{a}{\lambda\mu}}}{e^{\frac{-b}{\lambda\mu}} + e^{\frac{a}{\lambda\mu}}} \right) & \text{if } z_j \in \{(h, l), (h, h)\}, \end{cases}$$

$$\left( \mathcal{P}_{j,(S_l,0)}^{0,*}, \mathcal{P}_{j,(S_h,0)}^{0,*} \right) = \left( \frac{1}{2}, \frac{1}{2} \right),$$

where  $\lambda$  denotes the agent's opportunity cost of attention (if the agent uses  $F$  as a single input, the corresponding probabilities follow analogously). The optimal amount of attention allocated to a location with single input use and the associated expected return are

$$\kappa_j^*(\lambda) = \log(2) + \frac{e^{\frac{a}{\lambda\mu}}}{e^{\frac{a}{\lambda\mu}} + e^{\frac{-b}{\lambda\mu}}} \log \left( \frac{e^{\frac{a}{\lambda\mu}}}{e^{\frac{a}{\lambda\mu}} + e^{\frac{-b}{\lambda\mu}}} \right) + \frac{e^{\frac{-b}{\lambda\mu}}}{e^{\frac{-b}{\lambda\mu}} + e^{\frac{a}{\lambda\mu}}} \log \left( \frac{e^{\frac{-b}{\lambda\mu}}}{e^{\frac{-b}{\lambda\mu}} + e^{\frac{a}{\lambda\mu}}} \right),$$

$$E[V_j(\kappa_j^*)] = \frac{a e^{\frac{a}{\lambda\mu}} - b e^{\frac{-b}{\lambda\mu}}}{e^{\frac{a}{\lambda\mu}} + e^{\frac{-b}{\lambda\mu}}}.$$

PROOF. Given that the two inputs are homogeneous, it is sufficient to derive the optimal probabilities for one of the two inputs. Without loss of generality, let  $S$  be the applied input whenever the agent uses a single input in production location  $j$ . Plugging the associated pay-offs from Table 1 into Equation (A.40) yields the following two conditions:

$$1 = 0.5 \frac{e^{\frac{a}{\lambda\mu}}}{\mathcal{P}_{j,(S_l,0)}^0 e^{\frac{a}{\lambda\mu}} + \mathcal{P}_{j,(S_h,0)}^0 e^{\frac{-b}{\lambda\mu}}} + 0.5 \frac{e^{\frac{-b}{\lambda\mu}}}{\mathcal{P}_{j,(S_l,0)}^0 e^{\frac{-b}{\lambda\mu}} + \mathcal{P}_{j,(S_h,0)}^0 e^{\frac{a}{\lambda\mu}}} \quad \text{if } \mathcal{P}_{j,(S_l,0)}^0 > 0,$$

$$1 = 0.5 \frac{e^{\frac{-b}{\lambda\mu}}}{\mathcal{P}_{j,(S_l,0)}^0 e^{\frac{a}{\lambda\mu}} + \mathcal{P}_{j,(S_h,0)}^0 e^{\frac{-b}{\lambda\mu}}} + 0.5 \frac{e^{\frac{a}{\lambda\mu}}}{\mathcal{P}_{j,(S_l,0)}^0 e^{\frac{-b}{\lambda\mu}} + \mathcal{P}_{j,(S_h,0)}^0 e^{\frac{a}{\lambda\mu}}} \quad \text{if } \mathcal{P}_{j,(S_h,0)}^0 > 0.$$

Notice that these two conditions are equivalent to Equations (A.29) and (A.30) obtained for the baseline model. The reason for this is that the choice problem associated with using a single input in one production location is equivalent to the choice problem for  $k = -\tau$  in the baseline model, that is, the agent essentially faces a binary choice between *low* and *high*. Following the same steps as in Appendix A.1.2 shows that the optimal unconditional probabilities in the case of single input use are given by  $(\mathcal{P}_{j,(S_l,0)}^{0,*}, \mathcal{P}_{j,(S_h,0)}^{0,*}) = (\frac{1}{2}, \frac{1}{2})$ . Using this result, the optimal choice probabilities  $\mathcal{P}_{j,i}^*(\mathbf{v})$  can be obtained by plugging the respective values of  $\mathcal{P}_{j,i}^{0,*}$  into Equation (A.39). This directly yields the expression stated in Lemma A.1.

The solution for  $\kappa_j^*$  can be found by plugging the optimal probabilities  $\mathcal{P}_{j,i}^*(\mathbf{v})$  and  $\mathcal{P}_{j,i}^{0,*}$  into Equation (A.38). This leads to

$$\begin{aligned} \kappa_j^* &= -0.5 \log(0.5) - 0.5 \log(0.5) \\ &+ 0.5 \left[ \frac{e^{\frac{a}{\lambda\mu}}}{e^{\frac{a}{\lambda\mu}} + e^{\frac{-b}{\lambda\mu}}} \log \left( \frac{e^{\frac{a}{\lambda\mu}}}{e^{\frac{a}{\lambda\mu}} + e^{\frac{-b}{\lambda\mu}}} \right) + \frac{e^{\frac{-b}{\lambda\mu}}}{e^{\frac{-b}{\lambda\mu}} + e^{\frac{a}{\lambda\mu}}} \log \left( \frac{e^{\frac{-b}{\lambda\mu}}}{e^{\frac{-b}{\lambda\mu}} + e^{\frac{a}{\lambda\mu}}} \right) \right] \\ &+ 0.5 \left[ \frac{e^{\frac{-b}{\lambda\mu}}}{e^{\frac{-b}{\lambda\mu}} + e^{\frac{a}{\lambda\mu}}} \log \left( \frac{e^{\frac{-b}{\lambda\mu}}}{e^{\frac{-b}{\lambda\mu}} + e^{\frac{a}{\lambda\mu}}} \right) + \frac{e^{\frac{a}{\lambda\mu}}}{e^{\frac{-b}{\lambda\mu}} + e^{\frac{a}{\lambda\mu}}} \log \left( \frac{e^{\frac{a}{\lambda\mu}}}{e^{\frac{-b}{\lambda\mu}} + e^{\frac{a}{\lambda\mu}}} \right) \right], \end{aligned}$$

which can be simplified to the expression stated in Lemma A.1.

In a similar way, the optimal probabilities can be used to derive an expression for the expected payoff at a location with single input use, corresponding to the part of the agent’s objective given by

$$(A.41) \quad E[V_j(\kappa_j)] = \sum_{i \in \mathcal{I}} \int_V v_{j,i} \mathcal{P}_{j,i}(v) G(dV).$$

Plugging  $\mathcal{P}_{j,i}^*(v)$  and  $\mathcal{P}_{j,i}^{0,*}$  into Equation (A.41) leads to

$$E[V_j(\kappa_j^*)] = 0.5 a \frac{e^{\frac{a}{\lambda\mu}}}{e^{\frac{a}{\lambda\mu}} + e^{\frac{-b}{\lambda\mu}}} - 0.5 b \frac{e^{\frac{-b}{\lambda\mu}}}{e^{\frac{-b}{\lambda\mu}} + e^{\frac{a}{\lambda\mu}}} - 0.5 b \frac{e^{\frac{-b}{\lambda\mu}}}{e^{\frac{a}{\lambda\mu}} + e^{\frac{-b}{\lambda\mu}}} + 0.5 a \frac{e^{\frac{a}{\lambda\mu}}}{e^{\frac{-b}{\lambda\mu}} + e^{\frac{a}{\lambda\mu}}},$$

which can be further simplified to the expression stated in Lemma A.1. This concludes the proof of Lemma A.1.  $\square$

**Case 3: Joint input use.** In case, the agent uses both inputs together in a location, the four possible ways of combining inputs with each other are given by  $i \in \{(S_l, F_l), (S_l, F_h), (S_h, F_l), (S_h, F_h)\}$ . The agent seeks to select the combination that conforms best to the realized state and permits the agent to take advantage of the complementarity between the two inputs. The optimal behavior at a given location under joint input use is summarized by Lemma A.2.

LEMMA A.2. *The optimal behavior of a rationally inattentive agent using both complementary inputs together in location  $j$  is characterized by the set of choice probabilities*

$$\left( \mathcal{P}_{j,(S_l,F_l)}^*(v), \mathcal{P}_{j,(S_l,F_h)}^*(v), \mathcal{P}_{j,(S_h,F_l)}^*(v), \mathcal{P}_{j,(S_h,F_h)}^*(v) \right) = \begin{cases} \left( \frac{\Psi^+}{\Theta}, \frac{\Psi}{\Theta}, \frac{\Psi}{\Theta}, \frac{\Psi^-}{\Theta} \right) & \text{if } z_j = (l, l) \\ \left( \frac{\Psi}{\Theta}, \frac{\Psi^+}{\Theta}, \frac{\Psi^-}{\Theta}, \frac{\Psi}{\Theta} \right) & \text{if } z_j = (l, h) \\ \left( \frac{\Psi}{\Theta}, \frac{\Psi^-}{\Theta}, \frac{\Psi^+}{\Theta}, \frac{\Psi}{\Theta} \right) & \text{if } z_j = (h, l) \\ \left( \frac{\Psi^-}{\Theta}, \frac{\Psi}{\Theta}, \frac{\Psi}{\Theta}, \frac{\Psi^+}{\Theta} \right) & \text{if } z_j = (h, h) \end{cases}$$

$$\left( \mathcal{P}_{j,(S_l,F_l)}^{0,*}, \mathcal{P}_{j,(S_l,F_h)}^{0,*}, \mathcal{P}_{j,(S_h,F_l)}^{0,*}, \mathcal{P}_{j,(S_h,F_h)}^{0,*} \right) = \left( \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4} \right),$$

where  $\Psi^+ = e^{\frac{2a(1+s)}{\lambda\mu}}$ ,  $\Psi = e^{\frac{a-b}{\lambda\mu}}$ ,  $\Psi^- = e^{\frac{-2b}{\lambda\mu}}$ , and  $\Theta = \Psi^+ + 2\Psi + \Psi^-$ . The optimal amount of attention devoted to a production location with joint input use and the associated expected return are given as

$$\kappa_j^*(\lambda) = \log(4) + \frac{\Psi^+}{\Theta} \log\left(\frac{\Psi^+}{\Theta}\right) + 2\frac{\Psi}{\Theta} \log\left(\frac{\Psi}{\Theta}\right) + \frac{\Psi^-}{\Theta} \log\left(\frac{\Psi^-}{\Theta}\right)$$

$$E[V_j(\kappa_j^*)] = \Theta^{-1} \left[ 2a(1+s) e^{\frac{2a(1+s)}{\lambda\mu}} + 2(a-b) e^{\frac{a-b}{\lambda\mu}} - 2b e^{\frac{-2b}{\lambda\mu}} \right].$$

PROOF. The optimal probabilities for joint input use in a single production unit can be derived along the same lines as in the case of single input use. Plugging the corresponding payoffs from Table 1 into the optimality condition (A.40) yields four conditions, one for each of the four possible ways of applying both inputs together. For the first action,  $i = (S_l, F_l)$ , the

condition reads

$$\begin{aligned}
 \text{(A.42)} \quad 1 &= 0.25 \frac{\Psi^+}{\mathcal{P}_{j,(S_l,F_l)}^0 \Psi^+ + \mathcal{P}_{j,(S_l,F_h)}^0 \Psi^+ + \mathcal{P}_{j,(S_h,F_l)}^0 \Psi^- + \mathcal{P}_{j,(S_h,F_h)}^0 \Psi^-} \\
 &+ 0.25 \frac{\Psi}{\mathcal{P}_{j,(S_l,F_l)}^0 \Psi + \mathcal{P}_{j,(S_l,F_h)}^0 \Psi^+ + \mathcal{P}_{j,(S_h,F_l)}^0 \Psi^- + \mathcal{P}_{j,(S_h,F_h)}^0 \Psi} \\
 &+ 0.25 \frac{\Psi}{\mathcal{P}_{j,(S_l,F_l)}^0 \Psi + \mathcal{P}_{j,(S_l,F_h)}^0 \Psi^- + \mathcal{P}_{j,(S_h,F_l)}^0 \Psi^+ + \mathcal{P}_{j,(S_h,F_h)}^0 \Psi} \\
 &+ 0.25 \frac{\Psi^-}{\mathcal{P}_{j,(S_l,F_l)}^0 \Psi^- + \mathcal{P}_{j,(S_l,F_h)}^0 \Psi + \mathcal{P}_{j,(S_h,F_l)}^0 \Psi + \mathcal{P}_{j,(S_h,F_h)}^0 \Psi^+} \quad \text{if } \mathcal{P}_{j,(S_l,F_l)}^0 > 0,
 \end{aligned}$$

where  $\Psi^+ = e^{\frac{2a(1+s)}{\lambda\mu}}$ ,  $\Psi = e^{\frac{a-b}{\lambda\mu}}$ , and  $\Psi^- = e^{\frac{-2b}{\lambda\mu}}$ . The conditions for the other three possible actions are given by analogous terms. The solution to the system of equations given by these four conditions is obtained by assigning equal probabilities  $\mathcal{P}_{j,i}^0$  to each of the four actions. This can be shown by verifying the guess

$$\text{(A.43)} \quad \left( \mathcal{P}_{j,(S_l,F_l)}^{0,*}, \mathcal{P}_{j,(S_l,F_h)}^{0,*}, \mathcal{P}_{j,(S_h,F_l)}^{0,*}, \mathcal{P}_{j,(S_h,F_h)}^{0,*} \right) = \left( \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4} \right),$$

as outlined in the following. Plugging the presumed values into Equation (A.42) and simplifying the obtained expression leads to

$$\begin{aligned}
 1 &= 0.25 \frac{\Psi^+}{\frac{1}{4}\Psi^+ + \frac{1}{4}\Psi + \frac{1}{4}\Psi + \frac{1}{4}\Psi^-} + 0.25 \frac{\Psi}{\frac{1}{4}\Psi + \frac{1}{4}\Psi^+ + \frac{1}{4}\Psi^- + \frac{1}{4}\Psi} \\
 &+ 0.25 \frac{\Psi}{\frac{1}{4}\Psi + \frac{1}{4}\Psi^- + \frac{1}{4}\Psi^+ + \frac{1}{4}\Psi} + 0.25 \frac{\Psi^-}{\frac{1}{4}\Psi^- + \frac{1}{4}\Psi + \frac{1}{4}\Psi + \frac{1}{4}\Psi^+}.
 \end{aligned}$$

Further simplifying the equation shows that this is equivalent to

$$\text{(A.44)} \quad 1 = \frac{\Psi^+ + 2\Psi + \Psi^-}{\Psi^+ + 2\Psi + \Psi^-}.$$

In the same way, the guess can be verified for the other three conditions. Therefore, if the agent decides to use both inputs together in the same production unit, each possible way of combining the two inputs features the same *ex ante* probability of being selected. Based on this result,  $\mathcal{P}_{j,i}^*(v)$  can be obtained by plugging the solution for  $\mathcal{P}_{j,i}^{0,*}$  into Equation (A.39). This directly gives the expression stated in Lemma A.2.

The solution for  $\kappa_j^*$  in case of joint input use can be obtained by plugging the respective optimal probabilities  $\mathcal{P}_{j,i}^*(v)$  and  $\mathcal{P}_{j,i}^{0,*}$  into Equation (A.38). This leads to

$$\begin{aligned}
 \kappa_j^* &= -4 \left[ \frac{1}{4} \log \left( \frac{1}{4} \right) \right] + \frac{\Psi^+}{\Psi^+ + 2\Psi + \Psi^-} \log \left( \frac{\Psi^+}{\Psi^+ + 2\Psi + \Psi^-} \right) \\
 &+ 2 \frac{\Psi}{\Psi^+ + 2\Psi + \Psi^-} \log \left( \frac{\Psi}{\Psi^+ + 2\Psi + \Psi^-} \right) + \frac{\Psi^-}{\Psi^+ + 2\Psi + \Psi^-} \log \left( \frac{\Psi^-}{\Psi^+ + 2\Psi + \Psi^-} \right),
 \end{aligned}$$

which can be simplified to the expression stated in Lemma A.2.

Furthermore, using the optimal probabilities  $\mathcal{P}_{j,i}^*(v)$  and  $\mathcal{P}_{j,i}^{0,*}$  in the corresponding part of the agent’s objective function directly shows that the expected payoff for joint input use is given by the expression in Lemma A.2. This concludes the proof of Lemma A.2.  $\square$

**PROOF OF PROPOSITION 1.** The results in Lemmas A.1 and A.2 can be used to show that the agent’s optimal behavior is characterized by a pair of cutoff values in attentional capacity  $\bar{\kappa}$  that determine whether nonadoption, adoption of a single input, or joint adoption of both inputs is optimal for given parameter values. I first derive these cutoff values for a single production location. Given that the two locations are *ex ante* identical, these results can then be aggregated to obtain the corresponding cutoff values for two production location as stated in Proposition 1.

**Step 1: Single production location.** Consider first the choice between nonadoption and single input use at one location. Given the payoffs specified in Table 1 (including the parameter restriction  $0 < a < b$ ), this is equivalent to the choice of adopting the technology  $\tau$  in the baseline model in Subsection 3.1. The existence of a unique cutoff value that determines whether single input use is preferable to nonadoption thus follows analogously to the proof of Proposition A.1 in Appendix A.1.1 (note that the expected payoff for single input use specified in Lemma A.1 is equivalent to the expected payoff of adopting  $\tau$  if one sets  $N_k = 2$  and  $k = j$  in Equation (A.25)). Let this threshold be denoted as  $m_1$  here. If  $\bar{\kappa}$  is smaller than  $m_1$ , no input is used in this production location. If  $\bar{\kappa}$  is larger than  $m_1$ , the agent uses (at least) a single input.

Next, consider the choice between single and joint input use. The existence of a cutoff value for  $\bar{\kappa}$  that determines which of the two actions is optimal again follows from three properties. First, given the made parameter restriction  $s < \frac{b-a}{a}$ , the expected return of combining inputs without paying any attention to the production location is smaller than the expected return of using a single input (see the last column in Table 1). Second, for sufficiently large values of  $\bar{\kappa}$ , the expected payoff for joint input use is higher than for single input use. This follows under perfect attention, where the realized payoff under joint input use equals  $2a(1 + s) > a$ . Third, it can be shown that the expected payoff of joint input use is monotonically increasing in  $\bar{\kappa}$  (see below). Together, these three properties imply that there is a unique value for  $\bar{\kappa}$  at which the expected payoff of joint input use and single input use cross. Let this threshold be denoted as  $m_2$ . It captures the minimum amount of endowed attention for which combining two inputs in a location yields a larger expected payoff than single input use.

It should be noted that, in principle, it is possible that the second cutoff lies below the first one, so that the agent will switch from nonadoption directly to joint input use as  $\bar{\kappa}$  increases (and never use only a single input). This case may arise if joint input use is relatively attractive compared to single input use, because the synergy  $s$  is very large. In order to be able to capture the full spectrum of possible outcomes, I focus on the case where parameter values are such that  $m_2 > m_1$ .<sup>65</sup>

Before deriving the optimal behavior across two production locations, the monotonicity of expected payoffs with respect to scarcity of attention can be derived as follows: The expected payoff of joint input use is given in Lemma A.2 as

$$(A.45) \quad E[V_j(\kappa_j^*)] = \underbrace{\frac{2a(1+s)}{\Theta} e^{\frac{2a(1+s)}{\lambda\mu}}}_{(A)} + 2 \underbrace{\frac{(a-b)}{\Theta} e^{\frac{a-b}{\lambda\mu}}}_{(B)} - \underbrace{\frac{2b}{\Theta} e^{\frac{-2b}{\lambda\mu}}}_{(C)},$$

<sup>65</sup> A sufficient condition for  $m_2 > m_1$  to hold is that  $s > 0$  is infinitely small. Deriving a necessary condition for  $s$  seems only possible numerically (for given values of  $a$  and  $b$ ).



where  $\Theta = e^{\frac{2a(1+s)}{\lambda\mu}} + 2 \cdot e^{\frac{a-b}{\lambda\mu}} + e^{\frac{-2b}{\lambda\mu}}$ . The derivative with respect to  $\lambda$  can be computed as the sum of derivatives of the three summands. These are given by

$$(A.46) \quad \frac{\partial(A)}{\partial\lambda} = \left[ -\frac{4a^2(1+s)^2 \log(2)}{\lambda^2\mu} e^{\frac{2a(1+s)}{\lambda\mu}} \Theta - 2a(1+s) e^{\frac{2a(1+s)}{\lambda\mu}} \Theta' \right] \Theta^{-2},$$

$$(A.47) \quad \frac{\partial(B)}{\partial\lambda} = \left[ -\frac{2(a-b)^2 \log(2)}{\lambda^2\mu} e^{\frac{a-b}{\lambda\mu}} \Theta - 2(a-b) e^{\frac{a-b}{\lambda\mu}} \Theta' \right] \Theta^{-2},$$

$$(A.48) \quad \frac{\partial(C)}{\partial\lambda} = \left[ -\frac{4b^2 \log(2)}{\lambda^2\mu} e^{\frac{-2b}{\lambda\mu}} \Theta + 2b e^{\frac{-2b}{\lambda\mu}} \Theta' \right] \Theta^{-2},$$

where  $\Theta'$  denotes the derivative of  $\Theta$  with respect to  $\lambda$ , that is,

$$(A.49) \quad \frac{\partial\Theta}{\partial\lambda} = -\frac{2a(1+s) \log(2)}{\lambda^2\mu} e^{\frac{2a(1+s)}{\lambda\mu}} - \frac{2(a-b) \log(2)}{\lambda^2\mu} e^{\frac{a-b}{\lambda\mu}} + \frac{2b \log(2)}{\lambda^2\mu} e^{\frac{-2b}{\lambda\mu}}.$$

For the derivative of the first term (A), it can easily be verified that the term in squared brackets is negative (simplifying leads to a sum in which each term is negative). Since the remaining factor  $\Theta^{-2}$  is positive, it follows that  $\frac{\partial(A)}{\partial\lambda} < 0$ .

For the derivative of the second term (B), inserting  $\Theta$  and  $\Theta'$  in the term in squared brackets, and simplifying the resulting expression leads to

$$(A.50) \quad \frac{\partial(B)}{\partial\lambda} = \left[ \frac{2(a-b)(b+a+2as) \log(2)}{\lambda^2\mu} e^{\frac{(a-b)+2a(1+s)}{\lambda\mu}} + \frac{2(b^2 - a^2) \log(2)}{\lambda^2\mu} e^{\frac{a-3b}{\lambda\mu}} \right] \Theta^{-2}.$$

In a similar way, the derivative of the third term (C) can be written as

$$(A.51) \quad \frac{\partial(C)}{\partial\lambda} = \left[ -\frac{2b[2b+2a(1+s)] \log(2)}{\lambda^2\mu} e^{\frac{-2b+2a(1+s)}{\lambda\mu}} - \frac{4b(a+b) \log(2)}{\lambda^2\mu} e^{\frac{a-3b}{\lambda\mu}} \right] \Theta^{-2}.$$

Since  $b > a$ , the first fraction in Equation (A.50) is negative, whereas the second fraction is positive. In Equation (A.51), both fractions are negative such that  $\frac{\partial(C)}{\partial\lambda} < 0$ . To show that the overall derivative of Equation (A.45) is negative, one can take the sum of the second (only positive) summand in Equation (A.50) and the second summand in Equation (A.51). As both are multiplied with the term  $e^{\frac{a-3b}{\lambda\mu}}$ , the resulting sum is

$$(A.52) \quad \left[ \frac{2(b^2 - a^2) \log(2)}{\lambda^2\mu} - \frac{4b(a+b) \log(2)}{\lambda^2\mu} \right] e^{\frac{a-3b}{\lambda\mu}}.$$

Simplifying leads to a negative expression

$$(A.53) \quad -\frac{2(a+b)^2 \log(2)}{\lambda^2\mu} e^{\frac{a-3b}{\lambda\mu}} < 0.$$

Therefore, the derivative of Equation (A.45) with respect to  $\lambda$  is negative. It thus holds that the expected payoff of joint input use is monotonically decreasing in  $\lambda$  (and increasing in  $\bar{\kappa}$ ).

**Step 2: Two production locations.** For *ex ante* identical production units, the same amount of attention is needed in each location to make single input use more attractive than nonadoption, and joint input use more attractive than single input use, respectively. Hence,  $\bar{\kappa}$  has to equal at least  $2m_1$  to make using a single input in both locations desirable. For values smaller than  $2m_1$ , the agent either adopts the input only in one location or uses no modern input at all. In the same way,  $\bar{\kappa}$  has to equal at least  $2m_2$  to make joint input use desirable in both locations.

Therefore, as long as all actions represent relevant alternatives (i.e.,  $m_2 > m_1$ ), this implies the following order of optimal actions. For  $\bar{\kappa} = 0$ , the agent adopts no modern input. As  $\bar{\kappa}$  increases, it first becomes optimal to use a single input in one location (i.e., when  $\bar{\kappa} = m_1$ ), and then in both locations ( $\bar{\kappa} = 2m_1$ ). For larger values of  $\bar{\kappa}$ , it eventually becomes optimal to apply two inputs jointly in one location and use a single input in the other location ( $\bar{\kappa} = m_2 + m_1$ ), and finally to combine inputs in both locations ( $\bar{\kappa} = 2m_2$ ). Whether it first becomes optimal to apply a single input in both locations or combine inputs in one location while using no input in the other location depends on the relation of the two cutoff values, which is determined by the deep parameters of the model. If the parameter values are such that  $2m_1 < m_2$ , then the agent first uses a single input in both locations. If  $2m_1 > m_2$ , then the agent first combines inputs in one location while using no input in the other location. This concludes the proof of Proposition 1.  $\square$

A.1.4 *Proof of Corollary 1.* The budget constraint for attention is given by

$$(A.54) \quad \mu(\kappa_1 + \kappa_2) \leq \bar{\kappa}.$$

For a fixed capacity of attention  $\bar{\kappa}$ , a reduction in  $\mu$  thus increases the amount of attention ( $\kappa_1 + \kappa_2$ ) that is available to reduce uncertainty about the state of nature according to condition (A.38). Analogous to an increase in  $\bar{\kappa}$  itself, this makes attention in the agent's decision problem less scarce and thus lowers the associated shadow price  $\lambda$ . Using the results derived in Appendix A.1.3, it follows that the expected payoffs for single and joint input use are increasing with smaller values of  $\mu$ . Holding the payoff under nonadoption fixed, this implies that the threshold  $m_1$  of attentional capacity that is needed to make using a modern input profitable in expectation is lowered when  $\mu$  decreases. In the same way, sufficiently large reductions in  $\mu$  also lead to a lower cutoff  $m_2$  at which combining inputs becomes more attractive than using a single input. This follows from the two features that (i) for  $\bar{\kappa}$  fixed, values of  $\mu$  sufficiently close to zero allow for arbitrarily large reductions in uncertainty, as the sum of  $\kappa_1$  and  $\kappa_2$  in condition (A.54) increases, and (ii) under perfect attention combining complementary inputs is more profitable than using a single input. Given that production units are homogeneous, these results carry over to the case of two locations by following the same procedure as described in Appendix A.1.3.

A.2 *Simulations.* Deriving analytical comparative statics for the model would be very cumbersome, if at all possible. Instead, this appendix provides some insights on the roles of individual parameters using numerical simulations. Recall that the optimal behavior of the agent in the baseline model is characterized by the existence of a cutoff value,  $m_\tau$ , which captures the minimum amount of endowed attention for which adopting the technology is profitable in expectation. In studying the roles of individual parameters in determining optimal adoption behavior, I use the analytical expressions from Equation (6) and Proposition A.1 to simulate the cutoff associated with different parameter value combinations. A lower value of  $m_\tau$  indicates that less attention is required to make using the technology profitable in expectation, thus facilitating adoption.

A.2.1 *Simulations for (a, b).* Table A.1 shows the cutoffs  $m_\tau$  associated with different combinations of return parameters, holding other parameters fixed. In Panel A, the returns

TABLE A.1  
SIMULATIONS FOR  $(a, b)$

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<i>Panel A: <math>(a_{-\tau}, b_{-\tau}) = (1, 2)</math></i>									
$a_{-\tau}$	<b>1.0</b>	<b>1.2</b>	<b>1.4</b>	<b>0.8</b>	<b>0.6</b>	1.0	1.0	1.0	1.0
$-b_{-\tau}$	<b>-2.0</b>	-2.0	-2.0	-2.0	-2.0	<b>-1.8</b>	<b>-1.6</b>	<b>-2.2</b>	<b>-2.4</b>
$m_{\tau}$	0.43	0.23	0.11	0.71	1.01	0.35	0.26	0.50	0.56
<i>Panel B: <math>(a_{\tau}, b_{\tau}) = (1, 2)</math></i>									
$a_{-\tau}$	<b>1.0</b>	<b>1.2</b>	<b>1.4</b>	<b>0.8</b>	<b>0.6</b>	1.0	1.0	1.0	1.0
$-b_{-\tau}$	<b>-2.0</b>	-2.0	-2.0	-2.0	-2.0	<b>-1.8</b>	<b>-1.6</b>	<b>-2.2</b>	<b>-2.4</b>
$m_{\tau}$	0.43	0.46	0.49	0.40	0.37	0.40	0.37	0.46	0.49

NOTES: Values are based on simulations for the case of two binary choice problems specified in Section 3.1 and  $\mu = 1$ . The cutoff value  $m_{\tau}$  corresponds to the minimum amount of endowed attention  $\bar{\kappa}$  for which adopting the technology is profitable in expectation.

of technology use,  $a_{\tau}$  and  $b_{\tau}$ , are varied, whereas the returns of the other decision problem,  $a_{-\tau}$  and  $b_{-\tau}$ , are kept constant. Based on the numbers in the first five columns in Panel A, the minimum attentional capacity that makes adoption desirable decreases with larger returns to optimal usage (columns (2) and (3)), and increases when  $a_{\tau}$  becomes smaller (columns (4) and (5)). Similarly, the last four columns in Panel A show that  $m_{\tau}$  increases with larger possible losses for inadequate usage (columns (8) and (9)), and decreases when  $b_{\tau}$  becomes smaller (columns (6) and (7)). These results are intuitive as they imply that making the technology more attractive (either by rising the potential payoff  $a_{\tau}$  or by reducing possible losses  $b_{\tau}$ ) facilitates adoption.

Panel B in Table A.1 shows the resulting values of  $m_{\tau}$  when the returns of technology use remain fixed and, instead,  $a_{-\tau}$  and  $b_{-\tau}$  are varied. According to the first five columns, the attentional threshold for adoption increases when the maximum payoff obtained from competing decision problems becomes larger (columns (2) and (3)), and decreases when  $a_{-\tau}$  becomes smaller (columns (4) and (5)). This is intuitive, as it means that, everything else being equal, uptake of a new technology is more likely in situations where it is less attractive for users to devote attention to other decisions.

With respect to potential losses for other decision, Panel B in Table A.1 indicates that the attentional threshold for adoption decreases with smaller absolute values of  $b_{-\tau}$  (columns (6) and (7)), and increases when the absolute value of  $b_{-\tau}$  becomes larger (columns (8) and (9)). The intuition for these results stems from the fact that there is no outside option available for  $-\tau$ . Thus, the more costly mistakes are for  $-\tau$ , the higher are the incentives for the agent to devote attention to avoiding these mistakes.

**A.2.2 Simulations for  $N_{\tau}$  and  $\bar{\kappa}$ .** Figure A.1 shows the expected payoffs associated with different values of  $N_{\tau}$  when the agent faces two decision problems, that is,  $\tau$  and  $-\tau$  (recall that  $N_{\tau}$  captures the number of different parameter combinations from which users have to choose when applying the technology). Three sets of simulations are depicted in Figure A.1, namely, for  $\bar{\kappa}$  equal to 1, 1.5, and 2. Note that in the case of nonadoption, the value of  $N_{\tau}$  plays no role and any value  $\bar{\kappa} \geq 1$  allows the agent to reduce the uncertainty associated with  $-\tau$  to 0. Therefore, the payoff of nonadoption equals  $a$  for all three depicted sets of simulations (shown by the red circle markers). In the case of adoption, Figure A.1 shows that, for each depicted value of  $\bar{\kappa}$ , the expected payoff decreases with larger values of  $N_{\tau}$ . This is intuitive, as applying the technology in profitable ways is more difficult if the parameters of usage have to be selected from a wider range of possible options. The results in Figure A.1 thus imply that if individuals face a choice between different technologies, everything else being equal, then naturally, the technology associated with the least complex decision about optimal usage will be preferred.

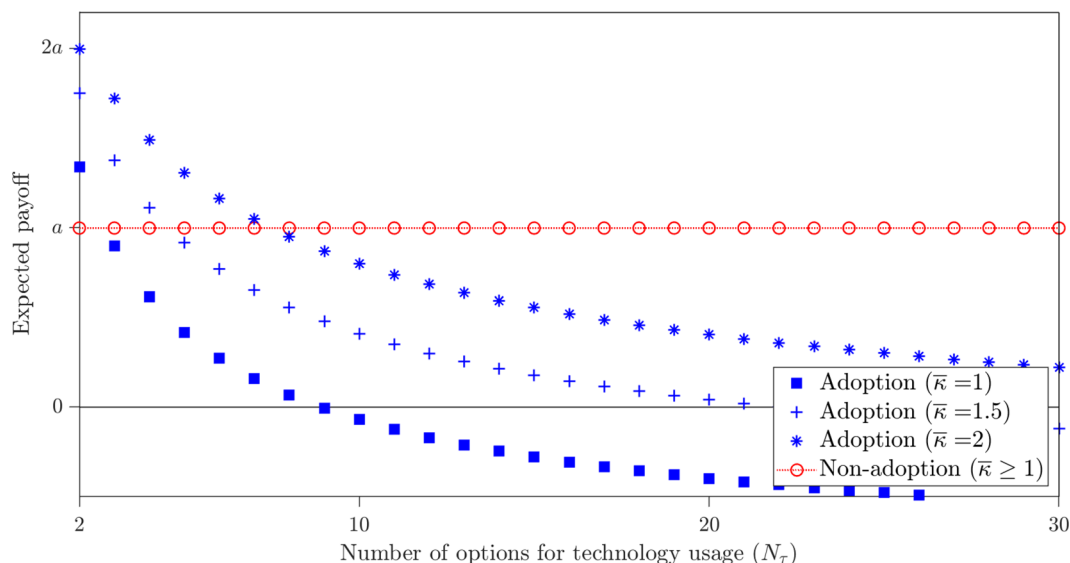


FIGURE A.1

SIMULATIONS FOR  $N_\tau$  AND  $\bar{\kappa}$

NOTES: SIMULATIONS ARE BASED ON  $K = 2$  CHOICE PROBLEMS WITH  $N_{-\tau} = 2$  AND FIXED PARAMETER VALUES  $(a, b, \mu) = (1, 2, 1)$ .

In addition, Figure A.1 shows that, consistent with the analytical results in Equation (6), adoption of profitable but attentionally demanding technology is facilitated by larger attentional capacity  $\bar{\kappa}$ . In particular, for each value of  $N_\tau$  in Figure A.1, the expected payoff of adoption is higher for larger values of  $\bar{\kappa}$ .

### A.3 Application to Modern Farming Technology.

**A.3.1 Data and variables.** The empirical results presented in Section 5 are based on data from nationally representative household surveys in Malawi, Nigeria, Tanzania, and Uganda. These were collected between 2011 and 2016 as part of the World Bank’s Living Standard Measurement Study - Integrated Surveys on Agriculture (LSMS-ISA). For the analysis, I pool the information from four survey rounds (Malawi 2013, Nigeria 2015/16, Tanzania 2012/13, Uganda 2011/12), so that my sample includes 11,057 observations at the household (farm) level and 22,160 observations at the plot level. This includes all households in the original LSMS-ISA data sets that cultivated at least one agricultural plot during the main growing season (plots that are rented out or left fallow are excluded). Although the LSMS-ISA project seeks to collect data in a comparable way across countries, some differences in survey designs remain. As far as possible, I restrict my analysis to variables for which the available information is comparable across countries. A list of the included variables, together with basic summary statistics, is provided in Table A.2. The following provides additional information on the construction of variables and handling of the unique survey design in each country. For more details on the data collection process and utilized survey tools, I refer to the resources available at World Bank (2018).

**Modern input use.** Observed inputs include improved seeds, inorganic fertilizer, other agrochemicals (pesticides and herbicides), organic fertilizer, and irrigation systems. All of these variables are based on information about binary input use decisions at the plot level, that is, whether farmers used the respective input (in any quantity) on individual plots or parcels. Based on this plot-level information, the corresponding household-level indicators are set

TABLE A.2  
LIST OF VARIABLES

Variable	Full Sample <sup>a</sup>	Malawi 2013	Nigeria 2015–16	Tanzania 2012–13	Uganda 2011–12	Description <sup>b</sup>
<i>Household level:</i>						
Farm size	1.61	0.79	1.88	2.19	1.59	Cultivated area, farmer-assessed (ha)
Number of plots	2.00	2.05	2.00	2.03	1.91	Number of cultivated plots
Household size	5.78	5.28	6.26	5.67	6.00	Number of HH members
Distance to road	11.8	9.1	7.1	21.7	8.1	Distance to road (km)
Distance to market	53.7	23.5	71.2	82.3	32.7	Distance to market (km)
Rural	0.87	0.87	0.89	0.84	0.88	D: HH lives in rural area
Food expenditure	16.5	26.1	16.0	13.0	8.6	Food expenditure, last 7 days (USD)
Non-food expenditure	23.0	33.6	16.5	25.4	13.4	Non-food exp., last 1 month (USD)
Secondary education	0.43	0.40	0.61	0.32	0.41	D: Secondary education or more in HH
Age head	48.5	43.9	53.2	48.9	48.1	Age of HH head (years)
Male head	0.77	0.76	0.84	0.77	0.69	D: HH head is male
<i>Input use</i>						
Improved seeds	0.31	0.57	0.17	0.24	0.21	D: HH uses improved seeds
Inorganic fertilizer	0.38	0.79	0.45	0.15	0.04	D: HH uses inorganic fertilizer
Agrochemicals	0.20	0.06	0.46	0.15	0.11	D: HH uses agrochemicals
Organic fertilizer	0.21	0.21	0.28	0.19	0.13	D: HH uses organic fertilizer
Irrigation	0.02	0.01	0.02	0.03	0.01	D: HH uses irrigation systems
<i>Rainfall</i>						
Amount	10.25	8.31	11.88	8.82	12.89	12-month total rainfall (dm)
Amount (average)	9.99	8.45	12.56	8.23	11.24	Average 12-month rainfall (dm)
Amount (abs. dev.)	1.10	0.67	1.00	1.22	1.65	Distance to avg. 12-month rainfall (dm)
Timing	16.13	16.39	17.56	17.36	12.11	Start of wettest quarter (dekads)
Timing (average)	15.95	16.58	17.78	18.77	8.60	Avg. start of wettest quarter (dekads)
Timing (abs. dev.)	2.53	0.39	3.00	3.50	3.59	Distance to average start of wettest quarter (dekads)
<b>Observations (max)</b>	<b>11057</b>	<b>3038</b>	<b>2834</b>	<b>3023</b>	<b>2162</b>	
<i>Plot level:</i>						
Plot size	0.80	0.39	0.94	1.08	0.83	Size of plot, farmer-assessed (ha)
Distance > 1 km	0.39	0.32	0.33	0.50	0.44	D: Distance to farm house > 1 km
Good soil quality	0.58	0.47	0.81	0.45	0.64	D: Good soil quality, farmer-assessed
Sloped	0.35	0.41	0.25	0.30	0.47	D: Plot is sloped
Erosion	0.18	0.36	0.06	0.10	0.19	D: Plot has problems with erosion
Rented	0.19	0.22	0.16	0.17	0.22	D: Plot is rented in
Land title	0.09	0.03	0.06	0.13	0.14	D: Household has land title for plot
Maize	0.38	0.66	0.16	0.41	0.20	D: Main crop on plot is maize
Improved seeds	0.23	0.42	0.14	0.19	0.14	D: Improved seeds used on plot
Inorganic fertilizer	0.30	0.57	0.39	0.11	0.03	D: Inorganic fertilizer used on plot
Agrochemicals	0.16	0.04	0.42	0.10	0.07	D: Agrochemicals used on plot
Organic fertilizer	0.14	0.13	0.23	0.12	0.08	D: Organic fertilizer used on plot
Irrigation	0.01	0.01	0.01	0.02	0.01	D: Plot is irrigated
<b>Observations (max)</b>	<b>22160</b>	<b>6229</b>	<b>5672</b>	<b>6124</b>	<b>4136</b>	

NOTES: Numbers are mean values. Only agricultural plots which are cultivated (i.e., not rented out or left fallow) are included. <sup>a</sup>The full sample is obtained by pooling data across countries. <sup>b</sup>“D” indicates dummy variables. *Source:* Author’s computation based on survey data from the Living Standards Measurement Study - Integrated Surveys on Agriculture (World Bank, 2018). Rainfall estimates come from the National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center.

equal to 1 if farmers report having used the respective input on at least one plot during a given season. Information about seed types is as follows: For Uganda, the indicator for improved seeds is based on a distinction between improved and traditional seed varieties. For Nigeria, the indicator lumps together “improved” and “hybrid” seeds. In Tanzania, the questionnaire contains a third category for “improved recycled” seeds, which I do not include in the indicator for improved seeds. In Malawi, the distinction between improved and traditional seeds is only possible for maize, tobacco, groundnuts, and rice, which comprise about 80%

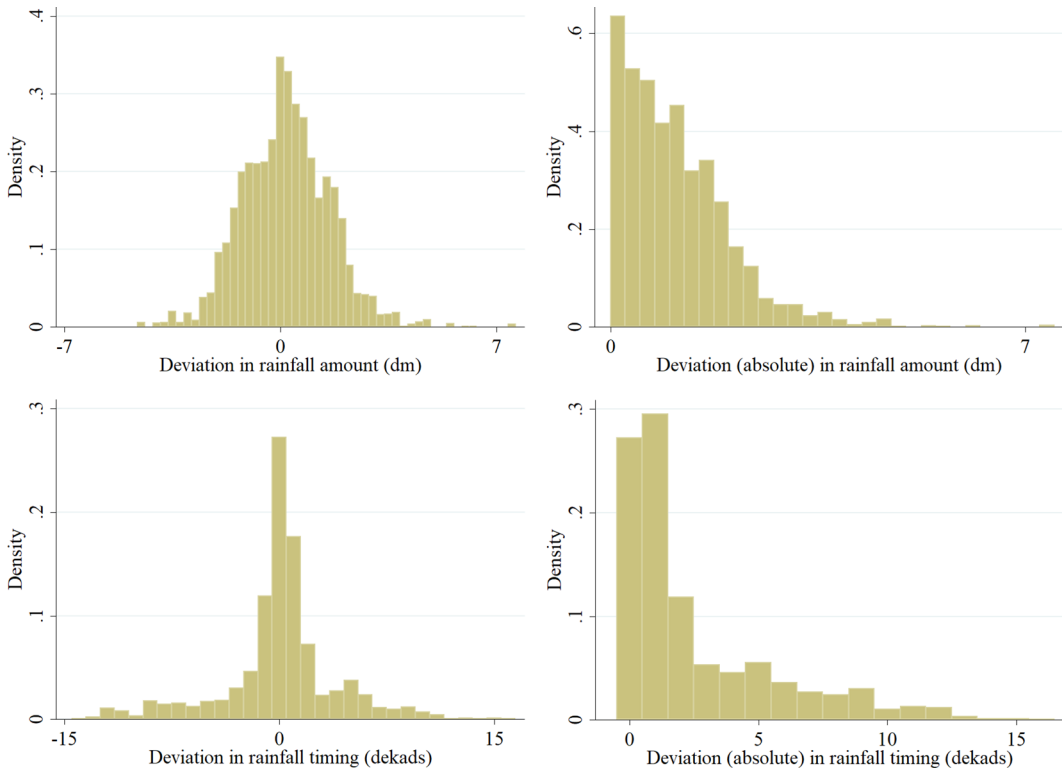
of plots (the corresponding adoption rate for Malawi should therefore be seen as a lower bound). The indicator for inorganic fertilizer is based on a binary response whether any inorganic or chemical fertilizer was used on a plot or parcel. The indicator for agrochemicals is based on a binary response about the use of pesticides or herbicides (in Nigeria, the indicator is based on two separate questions for pesticides and herbicides). The indicator for organic fertilizer is based on a binary response about organic fertilizer use (including manure). The indicator for irrigation is based on a binary response whether a plot or parcel was irrigated.

**Rainfall irregularities.** Households in the LSMS-ISA data sets are georeferenced, so that their locations can be linked to satellite data on climate and weather conditions. The rainfall variables used in the analysis are based on rainfall estimates from the National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center, which combine rain gauge data and multiple types of satellite data (with a final resolution of 0.1 decimal degrees). I use two variables provided by NOAA: 12-month total rainfall, measured in decimeters (dm), and start of the wettest quarter, expressed in dekads (each month is divided into three dekads, so that a dekad corresponds to a period of 8–11 days). To test the implications of the model, I construct two measures of irregularities in growing conditions, corresponding to deviations in the amount and timing of rainfall from previous conditions for each household. The first of these variables is calculated as the absolute deviation of 12-month total rainfall from the household's long-term average value of 12-month total rainfall estimates. Depending on data availability, the average value covers the past 10–15 years. The second variable is constructed in the same way using start of the wettest quarter (which typically corresponds to the beginning of the main growing season). Figure A.2 shows the resulting distributions of irregularities in the amount and timing of rainfall. For the regression analysis, both variables are rescaled so that values are expressed in units of standard deviation.

**Plot characteristics.** Observed plot characteristics include size, distance to farm house, ownership status, and a number of properties such as soil quality, slope, and erosion (a full list of the included plot-level variables is provided in Table A.2). All of these variables are based on farmer-reported values (for plot areas, GPS estimates by enumerators are frequently missing in the data sets, for example, due to flooding, security concerns, or because plots are located too far away from the household). In Malawi, Nigeria, and Tanzania, the data are collected at the plot level. In Uganda, the corresponding unit of observation is called parcel. I follow the approach of other authors and aggregate information across plots belonging to the same parcel (see Sheahan and Barrett, 2017). The indicator for the distance of plots to the farm house is based on the following information. In Malawi, Nigeria, and Tanzania, farmers are asked to state the distance (in km) from each plot to the farm house. In Uganda, the corresponding question refers to the time it takes to walk to the plot. To transform this into a comparable measure, I assume an average walking speed of 4 km/h. For the other included plot-level indicators, the underlying information is largely comparable across countries.

**Household characteristics.** Observed household characteristics include farm and household size, number of cultivated plots, distance to the nearest road and market, educational background, food and nonfood expenditures, and age and gender of the household head (see also Table A.2). Farm size includes only plots and parcels that are cultivated by the household in a given season (i.e., excluding fallow, pasture, and forest land, as well as plots that are given or rented out by the household). Information on household expenditures is based on a list of specific food and nonfood consumption items, which differ across countries. The two expenditure variables included in the analysis are constructed as the total value of all mentioned purchases of food and nonfood items, respectively, in each country. Local currencies are transformed into U.S. dollars using yearly average exchange rates from the World Development Indicator database. Information on quantities and area that was collected based on local units of measurement is transformed into standard units by using (as far as available) the conversion factors provided in the LSMS-ISA database.





NOTES: Irregularities in the amount of rainfall are calculated for each household as the (absolute) deviation of 12-months total rainfall (measured in decimeter) from the household's long-term average value. Irregularities in the timing of rainfall are calculated as the (absolute) deviation of the start of the wettest quarter (expressed in dekads) from the household's long-term average value. See Appendix A.3.1 for more details.

FIGURE A.2

DISTRIBUTION OF HOUSEHOLD-LEVEL RAINFALL IRREGULARITIES

**A.3.2 Robustness.** To test the robustness of the link between irregular growing conditions and modern input use (as discussed in Section 5) with respect to clustering of standard errors, I repeat the regressions in Table 3 using spatially correlated standard errors with a bandwidth of 100 km (instead of 50 km), and for standard errors clustered at the district level (in Nigeria, the corresponding administrative unit is a state). In both cases, the estimated coefficients for deviations in the *timing* of rainfall remain significantly different from 0 at significance levels of at least 10% across all nine specifications (i.e., the nine columns in Table 3). For deviations in the *amount* of rainfall, all specifications that were significant in Table 3 remain significant at the 10% level when standard errors are clustered at the district level. When spatial standard errors with a bandwidth of 100 kilometers are used, the coefficient for deviations in the amount of rainfall remains significant at the 10% level in the specification in column (9), but ceases to be significantly different from 0 in the specification in column (6).

The results for the selected plot and household characteristics reported in Table 3 appear all to be very robust to alternative clustering of standard errors. Except for the coefficient of secondary education in the specification in column (6), which was insignificant already in the baseline, all estimates remain significantly different from 0 at the 10% level for both alternative ways of clustering standard errors.

It should be noted that using absolute deviations in rainfall patterns in the regression model specified in Equation (11) assumes that the relationship is symmetric in the direction of deviation (e.g., both early and late timing of rainfall makes it more difficult for farmers to se-

TABLE A.3  
RAINFALL IRRREGULARITIES AND MODERN FARM INPUT USE: ROBUSTNESS

Number of Modern Inputs Used on Plot:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	One or More Inputs	Two or More Inputs	Three or more inputs						
Rainfall timing (deviation):									
More than 2 dekads early	-0.114*** (0.035)	-0.105** (0.042)	-0.098*** (0.036)	-0.139*** (0.029)	-0.121*** (0.033)	-0.116*** (0.029)	-0.069*** (0.014)	-0.057*** (0.015)	-0.054*** (0.014)
1-2 dekads early	-0.055** (0.027)	-0.054** (0.027)	-0.046* (0.025)	-0.062** (0.026)	-0.060** (0.026)	-0.051** (0.024)	-0.026** (0.011)	-0.025** (0.011)	-0.021** (0.010)
Same dekad as past average	omitted								
1-2 dekads late	0.044 (0.030)	0.043 (0.029)	0.014 (0.025)	0.062** (0.029)	0.060** (0.028)	0.034 (0.025)	0.025* (0.015)	0.023 (0.015)	0.016 (0.014)
More than 2 dekads late	-0.041 (0.036)	-0.054 (0.054)	-0.098** (0.048)	-0.056* (0.031)	-0.080* (0.043)	-0.115*** (0.039)	-0.035** (0.014)	-0.051*** (0.019)	-0.061*** (0.017)
Rainfall timing: start of wettest quarter		0.002 (0.004)	0.003 (0.003)		0.003 (0.003)	0.005* (0.002)		0.002 (0.001)	0.002** (0.001)
Country fixed effects	yes	yes	yes	yes	yes	yes	yes	yes	yes
Household and plot controls									
Observations	21,584	21,584	21,584	21,584	21,584	21,584	21,584	21,584	21,584
R <sup>2</sup>	0.181	0.181	0.214	0.114	0.114	0.145	0.044	0.045	0.063
Mean dependent variable	0.533	0.533	0.533	0.222	0.222	0.222	0.055	0.055	0.055

NOTES: \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Conley spatial standard errors reported in parentheses are based on a Bartlett weighting kernel with 50 km bandwidth. Inputs included in the dependent variable are improved seed varieties, organic and inorganic fertilizer, agrochemicals, and irrigation. Included control variables are as follows: At the plot level, a dummy for distance to farm house larger than 1 km, plot size, and dummies for good soil quality, sloped, erosion, rented, and land title are included. At the household level, number of cultivated plots, farm size (cultivated area), number of household members, distance to road, distance to market, food expenditure, non-food expenditure, age of household head, and dummies for secondary education, rural household, and male household head are included. All regressions include an additional constant which is not reported. *Source:* Author's calculations based on nationally representative survey data from Malawi, Nigeria, Tanzania, and Uganda in the World Bank's Living Standards Measurement Study - Integrated Surveys on Agriculture (see Appendix A.3.1 for more details).

lect optimal cultivation practices). To assess the plausibility of this assumption, Table A.3 provides additional evidence on the shape of the link between rainfall irregularities and farmers' input choices, focusing on the timing of rainfall. In particular, the estimations in Table A.3 include five different indicators, allowing deviations with different directions (i.e., early and late) and magnitudes (below or above the median deviation of 2 dekads) to have differential effects. According to the results, both early and late timing of rainfall tends to be negatively related to modern input use (the omitted category comprises observations for which the wettest quarter started within the same dekad as the long-term average value). For deviations above the respective median of 2 dekads (around 20 days), the coefficients for negative and positive deviations are very similar in size across the main specifications (columns (3), (6), and (9)), which supports the use of absolute deviations in the regression model in Equation (11). For smaller deviations (1–2 dekads), the coefficients for early timing are always significantly negative, whereas the coefficients of late timing are not statistically different from 0 when controlling for household and plot characteristics (columns (3), (6), and (9)).

#### REFERENCES

- ABAY, K. A., G. BERHANE, A. S. TAFFESSE, B. KORU, and K. ABAY, "Understanding Farmers' Technology Adoption Decisions: Input Complementarity and Heterogeneity," IFPRI ESSP Working Paper 82, 2016.
- ACHARYA, S., and S. L. WEE, "Rational Inattention in Hiring Decisions," *American Economic Journal: Macroeconomics* 12 (2020), 1–40.
- AGHION, P., and P. HOWITT, "A Model of Growth through Creative Destruction," *Econometrica* 60 (1992), 323–51.
- AHUJA, A., M. KREMER, and A. P. ZWANE, "Providing Safe Water: Evidence from Randomized Evaluations," *Annual Review of Resource Economics* 2 (2010), 237–56.
- ALI, S. N., "Herding with Costly Information," *Journal of Economic Theory* 175 (2018), 713–29.
- AMBUEHL, S., A. OCKENFELS, and C. STEWART, "Attention and Selection Effects," Working Paper, 2018.
- ASHRAF, N., J. BERRY, and J. M. SHAPIRO, "Can Higher Prices Stimulate Product Use? Evidence from a Field Experiment in Zambia," *American Economic Review* 100 (2010), 2383–413.
- , D. KARLAN, and W. YIN, "Tying Odysseus to the Mast: Evidence from a Commitment Savings Product in the Philippines," *Quarterly Journal of Economics* 121 (2006), 635–72.
- ATKIN, D., A. CHAUDHRY, S. CHAUDRY, A. K. KHANDELWAL, and E. VERHOOGEN, "Organizational Barriers to Technology Adoption: Evidence from Soccer-Ball Producers in Pakistan," *Quarterly Journal of Economics* 132 (2017), 1101–64.
- BANDIERA, O., and I. RASUL, "Social Networks and Technology Adoption in Northern Mozambique," *Economic Journal* 116 (2006), 869–902.
- BANERJEE, A., and E. DUFLO, "Growth Theory through the Lens of Development Economics," in P. Aghion and S. Durlauf, eds., *Handbook of Economic Growth* (Amsterdam: Elsevier, 2005), 473–552.
- , and S. MULLAINATHAN, "Limited Attention and Income Distribution," *American Economic Review* 98 (2008), 489–93.
- BARDHAN, P., and C. UDRY, *Development Microeconomics* (Oxford: Oxford University Press, 1999).
- BARTOŠ, V., M. BAUER, J. CHYTILOVÁ, and F. MATĚJKA, "Attention Discrimination: Theory and Field Experiments with Monitoring Information Acquisition," *American Economic Review* 106 (2016), 1437–75.
- BEAMAN, L., D. KARLAN, B. THUYSBAERT, and C. UDRY, "Profitability of Fertilizer: Experimental Evidence from Female Rice Farmers in Mali," *American Economic Review* 103 (2013), 381–6.
- BERNHEIM, D., D. RAY, and S. YELTEKIN, "Poverty and Self-Control," *Econometrica* 83 (2015), 1877–911.
- BESLEY, T., and A. CASE, "Modeling Technology Adoption in Developing Countries," *American Economic Review* 83 (1993), 396–402.
- BHARGAVA, S., G. LOEWENSTEIN, and J. SYDNOR, "Choose to Lose: Health Plan Choices from a Menu with Dominated Option," *Quarterly Journal of Economics* 132 (2017), 1319–72.
- BLOOM, N., B. EIFERT, A. MAHAJAN, D. MCKENZIE, and J. ROBERTS, "Does Management Matter? Evidence from India," *Quarterly Journal of Economics* 128 (2013), 1–51.
- BORDALO, P., N. GENNAIOLI, and A. SHLEIFER, "Salience and Consumer Choice," *Journal of Political Economy* 121 (2013), 803–43.
- BRONNENBERG, B. J., J.-P. DUBÉ, M. GENTZKOW, and J. M. SHAPIRO, "Do Pharmacists Buy Bayer? Informed Shoppers and the Brand Premium," *Quarterly Journal of Economics* 130 (2015), 1669–726.

- CAPLIN, A., and M. DEAN, "Rational Inattention and State Dependent Stochastic Choice," Working Paper, 2013.
- CASABURI, L., M. KREMER, S. MULLAINATHAN, and R. RAMRATTAN, "Harnessing ICT to Increase Agricultural Production: Evidence from Kenya," Working Paper, 2014.
- CASELLI, F., and W. J. COLEMAN, "Cross-Country Technology Diffusion: The Case of Computers," *American Economic Review* 91 (2001), 328–35.
- CHRISTIAENSEN, L., L. DEMERY, and J. KUHLMANN, "The (Evolving) Role of Agriculture in Poverty Reduction: An Empirical Perspective," *Journal of Development Economics* 96 (2011), 239–54.
- COLE, S., and A. N. FERNANDO, "The Value of Advice: Evidence from Mobile Phone-Based Agricultural Extension," HBS Working Paper 13-047, 2012.
- COMIN, D., and B. HOBIJN, "Cross-Country Technology Adoption: Making the Theories Face the Facts," *Journal of Monetary Economics* 51 (2004), 39–83.
- CONLEY, T. G., "GMM Estimation with Cross Sectional Dependence," *Journal of Econometrics* 92 (1999), 1–45.
- , and C. R. UDRY, "Learning about a New Technology: Pineapple in Ghana," *American Economic Review* 100 (2010), 35–69.
- COVER, T. M., and J. A. THOMAS, *Elements of Information Theory* (New York: John Wiley and Sons, 1991).
- DATA, S., and S. MULLAINATHAN, "Behavioral Design: A New Approach to Development Policy," *Review of Income and Wealth* 60 (2014), 7–35.
- DELLAVIGNA, S., "Psychology and Economics: Evidence from the Field," *Journal of Economic Literature* 47 (2009), 315–72.
- DERCON, S., and L. CHRISTIAENSEN, "Consumption Risk, Technology Adoption and Poverty Traps: Evidence from Ethiopia," *Journal of Development Economics* 96 (2011), 159–73.
- DUFLO, E., M. KREMER, and J. ROBINSON, "How High Are Rates of Return to Fertilizer? Evidence from Field Experiments in Kenya," *American Economic Review* 98 (2008), 482–8.
- , "Why Don't Farmers Use Fertilizer? Evidence from Field Experiments in Kenya," Working Paper, 2010.
- , "Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya," *American Economic Review* 101 (2011), 2350–90.
- , and R. PANDE, "Dams," *Quarterly Journal of Economics* 122 (2007), 601–46.
- DUPAS, P., "What Matters (and What Does Not) in Households' Decision to Invest in Malaria Prevention?," *American Economic Review* 99 (2009), 224–30.
- EASTERLY, W., and R. LEVINE, "What Have We Learned from a Decade of Empirical Research on Growth? It's Not Factor Accumulation: Stylized Facts and Growth Models," *World Bank Economic Review* 15 (2001), 177–219.
- EVENSON, R., and D. GOLLIN, "Assessing the Impact of the Green Revolution, 1960 to 2000," *Science* 300 (2003), 758–62.
- FALKINGER, J., "Limited Attention as a Scarce Resource in Information-Rich Economies," *Economic Journal* 118 (2008), 1596–620.
- FAO, *The State of Food and Agriculture: Innovation in Family Farming* (Rome: Food and Agriculture Organization of the United Nations, 2014).
- , *The International Code of Conduct for the Sustainable Use and Management of Fertilizers* (Rome: Food and Agriculture Organization of the United Nations, 2019).
- FEDER, G., "Farm Size, Risk Aversion and the Adoption of New Technology under Uncertainty," *Oxford Economic Papers* 32 (1980), 263–83.
- FOSTER, A. D., and M. R. ROSENZWEIG, "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture," *Journal of Political Economy* 103 (1995), 1176–209.
- , "Microeconomics of Technology Adoption," *Annual Review of Economics* 2 (2010), 395–424.
- GABAIX, X., D. LAIBSON, G. MOLOCHE, and S. WEINBERG, "Costly Information Acquisition: Experimental Analysis of a Boundedly Rational Model," *American Economic Review* 96 (2006), 1043–68.
- GEROSKI, P. A., "Models of Technology Diffusion," *Research Policy* 29 (2000), 603–25.
- GOECKE, H., W. J. LUHAN, and M. W. ROOS, "Rational Inattentiveness in a Forecasting Experiment," *Journal of Economic Behavior and Organization* 94 (2013), 80–9.
- GOLLIN, D., "Agricultural Productivity and Economic Growth," in R. Evenson and P. Pingali, eds., *Handbook of Agricultural Economics* (Amsterdam: North Holland, 2010), 3825–66.
- , and C. UDRY, "Heterogeneity, Measurement Error, and Misallocation: Evidence from African Agriculture," NBER Working Paper 25440, 2019.
- GREENWOOD, P., and R. PARASURAMAN, "Effects of Aging on the Speed and Attentional Cost of Cognitive Operations," *Developmental Neuropsychology* 7 (1991), 421–34.
- GROSSMAN, G. M., and E. HELPMAN, *Innovation and Growth in the Global Economy* (Cambridge, MA: MIT Press, 1993).

- GRUBB, M. D., and M. OSBORNE, "Cellular Service Demand: Biased Beliefs, Learning, and Bill Shock," *American Economic Review* 105 (2015), 234–71.
- HALL, B., "Innovation and Diffusion," in J. Fagerberg, D. Mowery, and R. Nelson, eds., *Oxford Handbook of Innovation* (Oxford: Oxford University Press, 2005), 459–84.
- HALL, R. E., and C. I. JONES, "Why Do Some Countries Produce So Much More Output per Worker than Others?," *Quarterly Journal of Economics* 114 (1999), 83–116.
- HANDEL, B., and J. SCHWARTZSTEIN, "Frictions or Mental Gaps: What's Behind the Information We (Don't) Use and When Do We Care?," *Journal of Economic Perspectives* 32 (2018), 155–78.
- HANDEL, B. R., "Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts," *American Economic Review* 103 (2013), 2643–82.
- HANNA, R., S. MULLAINATHAN, and J. SCHWARTZSTEIN, "Learning Through Noticing: Theory and Evidence from a Field Experiment," *Quarterly Journal of Economics* 129 (2014), 1311–53.
- HAUSHOFER, J., and E. FEHR, "On the Psychology of Poverty," *Science* 344 (2014), 862–7.
- HELLWIG, C., and L. VELDKAMP, "Knowing What Others Know: Coordination Motives in Information Acquisition," *Review of Economic Studies* 76 (2009), 223–51.
- HIRSHLEIFER, D., S. S. LIM, and S. H. TEOH, "Driven to Distraction: Extraneous Events and Underreaction to Earnings News," *Journal of Finance* 64 (2009), 2289–325.
- HORTAÇSU, A., and C. SYVERSON, "Product Differentiation, Search Costs, and Competition in the Mutual Fund Industry: A Case Study of S&P 500 Index Funds," *Quarterly Journal of Economics* 119 (2004), 403–56.
- HOSSAIN, T., and J. MORGAN, "...Plus Shipping and Handling: Revenue (Non) Equivalence in Field Experiments on Ebay," *Advances in Economic Analysis & Policy* 5 (2006).1–27.
- IFAD, *Smallholders, Food Security, and the Environment* (Rome: International Fund for Agricultural Development, 2013).
- IRZ, X., L. LIN, C. THIRTLE, and S. WIGGINS, "Agricultural Productivity Growth and Poverty Alleviation," *Development Policy Review* 19 (2001), 449–66.
- JOVANOVIĆ, B., and Y. NYARKO, "Learning by Doing and the Choice of Technology," *Econometrica* 64 (1996), 1299–310.
- KACPERCZYK, M., S. VAN NIEUWERBURGH, and L. VELDKAMP, "A Rational Theory of Mutual Funds' Attention Allocation," *Econometrica* 84 (2016), 571–626.
- KARLAN, D., R. OSEI, I. OSEI-AKOTO, and C. UDRY, "Agricultural Decisions after Relaxing Credit and Risk Constraints," *Quarterly Journal of Economics* 129 (2014), 597–652.
- KELLER, W., "International Technology Diffusion," *Journal of Economic Literature* 42 (2004), 752–82.
- KOCH, B., R. KHOSLA, W. FRASIER, D. WESTFALL, and D. INMAN, "Economic Feasibility of Variable-Rate Nitrogen Application Utilizing Site-Specific Management Zones," *Agronomy Journal* 96 (2004), 1572–80.
- KŐSZEGI, B., and F. MATĚJKA, "Choice Simplification: A Theory of Mental Budgeting and Naive Diversification," *Quarterly Journal of Economics* 135 (2020), 1153–207.
- , and A. SZEIDL, "A Model of Focusing in Economic Choice," *Quarterly Journal of Economics* 128 (2013), 53–104.
- KREMER, M., G. RAO, and F. SCHILBACH, "Behavioral Development Economics," in D. Bernheim, S. DellaVigna, and D. Laibson, eds., *Handbook of Behavioral Economics*, Vol. 2 (North Holland: Elsevier, 2019).
- LAMBERT, D., J. LOWENBERG-DEBOER, and G. MALZER, "Economic Analysis of Spatial-Temporal Patterns in Corn and Soybean Response to Nitrogen and Phosphorus," *Agronomy Journal* 98 (2006), 43–54.
- LUO, Y., "Consumption Dynamics under Information Processing Constraints," *Review of Economic Dynamics* 11 (2008), 366–85.
- MAĆKOWIAK, B., F. MATĚJKA, and M. WIEDERHOLT, "Dynamic rational inattention: Analytical results," *Journal of Economic Theory* 176 (2018), 650–92.
- , and ———, "Rational Inattention: A Review," Working Paper, 2021.
- , and M. WIEDERHOLT, "Optimal Sticky Prices under Rational Inattention," *American Economic Review* 99 (2009), 769–803.
- , "Business Cycle Dynamics under Rational Inattention," *Review of Economic Studies* 82 (2015), 1502–32.
- MACROBERT, J., P. SETIMELA, J. GETHI, and M. WORKU, *Maize Hybrid Seed Production Manual* (Mexico City: International Maize and Wheat Improvement Center, 2014).
- MANI, A., S. MULLAINATHAN, E. SHAFIR, and J. ZHAO, "Poverty Impedes Cognitive Function," *Science* 341 (2013), 976–80.
- MANKIW, N. G., and R. REIS, "Sticky Information Versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve," *Quarterly Journal of Economics* 117 (2002), 1295–328.
- MARENDA, P., and C. B. BARRETT, "State-Conditional Fertilizer Yield Response on Western Kenyan Farms," *American Journal of Agricultural Economics* 91 (2009), 991–1006.

- MARTIN, D., "Strategic Pricing with Rational Inattention to Quality," *Games and Economic Behavior* 104 (2017), 131–45.
- MATĚJKA, F., and A. MCKAY, "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model," *American Economic Review* 105 (2015), 272–98.
- MIGUEL, E., and M. KREMER, "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities," *Econometrica* 72 (2004), 159–217.
- MORRIS, M., V. KELLY, R. KOPICKI, and D. BYERLEE, *Fertilizer Use in African Agriculture: Lessons Learned and Good Practice Guidelines* (Washington, DC: World Bank, 2007).
- MORRIS, S., and M. YANG, "Coordination and Continuous Stochastic Choice," Working Paper, 2019.
- MOSE, C. M., and C. B. BARRETT, "The Complex Dynamics of Smallholder Technology Adoption: The Case of SRI in Madagascar," *Agricultural Economics* 35 (2006), 373–88.
- MUNSHI, K., "Social Learning in a Heterogeneous Population: Technology Diffusion in the Indian Green Revolution," *Journal of Development Economics* 73 (2004), 185–213.
- NAEHER, D. & M. SCHÜNDELN, "The Demand for Advice: Theory and Empirical Evidence from Farmers in Sub-Saharan Africa," *World Bank Economic Review*, forthcoming, 2021. <https://doi.org/10.1093/wber/lhab001>
- NYANGENA, W., and O. M. JUMA, "Impact of Improved Farm Technologies on Yields: The Case of Improved Maize Varieties and Inorganic Fertilizer in Kenya," Environment for Development Discussion Paper 14-02, 2014.
- PARENTE, S. L., "Technology Adoption, Learning-by-Doing, and Economic Growth," *Journal of Economic Theory* 63 (1994), 346–69.
- , and E. C. PRESCOTT, "Barriers to Technology Adoption and Development," *Journal of Political Economy* 102 (1994), 298–321.
- PAXSON, C. H., "Using Weather Variability to Estimate the Response of Savings to Transitory Income in Thailand," *American Economic Review* 82 (1992), 15–33.
- PLACE, F., C. B. BARRETT, H. A. FREEMAN, J. J. RAMISCH, and B. VANLAUWE, "Prospects for Integrated Soil Fertility Management Using Organic and Inorganic Inputs: Evidence from Smallholder African Agricultural Systems," *Food Policy* 28 (2003), 365–78.
- PRESCOTT, E. C., "Needed: A Theory of Total Factor Productivity," *International Economic Review* 39 (1998), 525–51.
- ROSENBERG, N., "Factors Affecting the Diffusion of Technology," *Explorations in Economic History* 10 (1972), 3–33.
- SALLEE, J. M., "Rational Inattention and Energy Efficiency," *Journal of Law and Economics* 57 (2014), 781–820.
- SCHULTZ, T. W., *Transforming Traditional Agriculture* (New Haven: Yale University Press, 1964).
- SHEAHAN, M., and ———, "Ten Striking Facts about Agricultural Input Use in Sub-Saharan Africa," *Food Policy* 67 (2017), 12–25.
- , R. BLACK, and T. JAYNE, "Are Kenyan Farmers Under-utilizing Fertilizer? Implications for Input Intensification Strategies and Research," *Food Policy* 41 (2013), 39–52.
- SIMS, C. A., "Implications of Rational Inattention," *Journal of Monetary Economics* 50 (2003), 665–90.
- SMITH, L., and P. SØRENSEN, "Pathological Outcomes of Observational Learning," *Econometrica* 68 (2000), 371–98.
- STEINER, J., C. STEWART, and F. MATĚJKA, "Rational Inattention Dynamics: Inertia and Delay in Decision-Making," *Econometrica* 85 (2017), 521–53.
- SURI, T., "Selection and Comparative Advantage in Technology Adoption," *Econometrica* 79 (2011), 159–209.
- TUTINO, A., "Rationally Inattentive Consumption Choices," *Review of Economic Dynamics* 16 (2013), 421–39.
- UDRY, C., and S. ANAGOL, "The Return to Capital in Ghana," *American Economic Review* 96 (2006), 388–93.
- VAN NIEUWERBURGH, S., and L. VELDKAMP, "Information Acquisition and Under-Diversification," *Review of Economic Studies* 77 (2010), 779–805.
- VELDKAMP, L. L., *Information Choice in Macroeconomics and Finance* (Princeton, NJ: Princeton University Press, 2011).
- VERHAEGHEN, P., D. STEITZ, M. SLIWINSKI, and J. CERELLA, "Aging and Dual-Task Performance: A Meta-Analysis," *Psychology and Aging* 18 (2003), 443–60.
- Bank World, *World Development Report 2008: Agriculture for Development* (Washington, DC: World Bank, 2008).
- , *Living Standard Measurement Study - Integrated Surveys on Agriculture (LSMS-ISA)* (2018), uRL: <https://www.worldbank.org/en/programs/lsms/initiatives/lsms-ISA> (last accessed on August 31, 2021).