The Roots of Agricultural Innovation: Patent Evidence of Knowledge Spillovers

Matt Clancy, Paul Heisey, Yongjie Ji, and GianCarlo Moschini *

Abstract

This chapter investigates the extent to which agricultural innovations draws on ideas originating outside of agriculture. We identify a large set of US patents for agricultural technologies granted between 1976 and 2018. To measure knowledge spillovers to these patents, we rely on three proxies: patent citations to other patents, patent citations to the scientific literature, and a novel text analysis to identify and track new ideas in the patent text. We find that more than half of knowledge flows originate outside of agriculture. The majority of these knowledge inflows, however, still originate in domains that are close to agriculture.

An earlier version of this paper was presented at the NBER conference on ***"Economics of Research and Innovation in Agriculture,"*** Washington, D.C., May 17, 2019.

This version: September 7, 2020

* Matt Clancy (mclancy@iastate.edu), Yongjie Ji (yongjiej@iastate.edu), and GianCarlo Moschini (moschini@iastate.edu) are with the Department of Economics and the Center for Agricultural and Rural Development, Iowa State University. At the time the research was conducted, Paul Heisey (pwheisey@gmail.com) was at the Economic Research Service, US Department of Agriculture.

0. Introduction

Changes in the technology of farming have profoundly affected U.S. production agriculture over the past century (Gardner, 2002). Myriad innovations adopted by farmers contributed to this transformation, including mechanization, vastly improved genetics for plants and animals, novel inputs such as fertilizers, pesticides, and antibiotics, and re-organization of farming activities to exploit specialization and scale economies. The results are impressive: between 1950 and 2015, for example, the total factor productivity index for U.S. agriculture increased by 167%, compared to 97% for the US non-farm private sector.[1]

Digging deeper into the causes of these waves of agricultural technical change uncovers the critical role played by past research and development (R&D) activities. Griliches' (1957) pioneering work on the yield improvements due to hybrid maize found a large payoff to the cumulated past research investment in this technology: a benefit–cost ratio of 7, or an internal rate of return of about 40%. More broadly, for a set of studies published over the 1965-2005 period, the median estimate of the internal rate of return of agricultural R&D was 45%, or a benefit-cost ratio of about 10 (Fuglie and Heisey 2007).

R&D explicitly focused on agriculture, conducted by firms and public organizations, is obviously essential to agricultural innovation. Non-agricultural R&D, however, may also play a role via so-called knowledge spillovers. The most immediate output of R&D is new knowledge, but it has long been recognized that the R&D performed by one entity (e.g., a public lab, or a firm) in a given industry may have substantial productivity impacts outside this entity or industry (Griliches 1992). At a positive level, spillovers create serious challenges to the task of inferring, from data, what R&D effort had which effect on outcomes of interest.

In this chapter, we focus squarely on assessing the extent to which knowledge spillovers may impact agricultural innovation. With some caveats, discussed later, we find that our proxies for

[1] Agricultural total factor productivity data based on input, output, and productivity data published by the Economic Research Service of the U.S. Department of Agriculture, https://www.ers.usda.gov/data-products/agricultural-productivity-in-the-us/. US non-farm private sector total factor productivity taken from Table XG4-2 from U.S. Bureau of Labor Statistics, https://www.bls.gov/mfp/special_requests/mfptablehis.xlsx.

knowledge flows—citations to patents, citations to scientific papers, and novel text—suggest more than 50% of knowledge spillovers originate in non-agricultural knowledge domains.

Knowledge spillovers have received limited attention in previous agricultural R&D studies. The typical econometric procedure has been to regress an estimate of agricultural productivity on relevant past R&D expenditures. To account for spillovers, some studies include broader measures of R&D expenditures. Attention has mostly concerned spillover between segments of agricultural R&D (Evenson 1989), or privileged spatial R&D spillovers, i.e., across states or countries (Latimer and Paarlberg, 1965; Khanna, Huffman, and Sandler, 1994). Alston (2002) concludes that such spillovers are sizeable: interstate or international R&D spillovers may account for more than half of the measured agricultural productivity growth. Consideration of vertical spillover effects in agriculture is rare. One exception is Wang, Xia, and Buccola (2009), who relate public research in three life-science fields (biology, agriculture, and medicine), and private research in two of these fields (agriculture and medicine), to research output (measured by patents) of private firms in agriculture and medicine.[2]

This paper's contribution is to provide new methods and data on the scope of knowledge spillovers in agriculture. In contrast to most studies in this area, we do not attempt to calculate the rate of return on R&D. Instead, we measure the extent of knowledge spillovers by directly observing proxies for knowledge flows and measuring the share of these that originate in non-agricultural knowledge domains. The goal is to provide new evidence on the extent to which agricultural technologies draw on knowledge originally developed outside of agriculture. We do so by developing various knowledge flow proxies embedded in US agricultural patents granted over the period 1976-2018.

Our initial step is to identify a set of relevant agricultural patents among the universe of US patents granted over this period. Note that while our analysis is restricted to US patents, these patents proxy global agricultural research. Depending on the subsector, we estimate 14%-49% of US agricultural patents are based on foreign research. We identify patents belonging to six distinct subsectors of agriculture: animal health, biocides, fertilizer, machinery, plants, and

[2] Consistent with our results, Wang, Xia, and Buccola (2009) find evidence of substantial spillovers from upstream biological to downstream agricultural and medical science, and from the public to the private sector in both downstream agriculture and downstream medicine.

agricultural research inputs. We chose these six subsectors because we can identify their patents with relative precision, and because, while not exhaustive, they span the major biological, chemical, and mechanical technology fields that have contributed to productivity growth in agriculture. Given significant differences in the size, organization, and scientific-technological knowledge base of these subsectors, our results are consistently presented separately for the six subsectors of interest. We then track the knowledge roots of each patent, using three proxies embedded in the patent document.

The first proxy we consider is citations to prior patents, which provides a measure of how agricultural innovations build on other (patented) technologies. When the cited patent is not an agricultural patent, this provides direct evidence of a knowledge spillover from outside agriculture. Furthermore, one advantage of studying citations to patents is that we can also identify the assignee of the cited patent. A major part of our work is to determine the "agricultural focus" of assignees' R&D, based on the share of agricultural patents in the assignee's recent patent portfolio. This permits us go beyond the binary classification of whether a cited patent is agricultural or not, and instead characterize it based on the agricultural focus of its assignee. For example, we can measure whether a cited patent belong to a firm that generally specializes in agricultural R&D, or belongs to an assignee that has zero agricultural patents.

The second proxy we employ is citations to scientific journal articles, which provides a measure of how agricultural innovations build on prior research. Citations to the scientific literature are important as a way of capturing the impact of public sector research, because public sector research frequently does not result in a patent. We create a classification system for scientific journals, identifying agricultural science, other biology, other chemistry, and "other" journals. We interpret a citation to a non-agricultural journal as evidence of a knowledge spillover to agriculture from outside of its natural knowledge domain.

Whereas citations to prior patents are generally acknowledged to contain both signal and noise, there is debate about the relative magnitude of each. For example, Chen (2017) finds the textual similarity of patents to their citations is much higher than to a control. An early survey by Jaffe, Trajtenberg, and Fogarty (2000), however, found only 38% of respondents were aware of the cited patent before or during the invention. Other papers have also found evidence that citations may not reflect genuine knowledge flows (Lampe 2012, Moser, Ohmstedt, and Rhode 2018).[3] For this reason, we also develop a third method of measuring knowledge flows, based on the text of patents.

The text analysis we develop identifies words and phrases that are new and important in agricultural patents applied for in the second half of our data sample (1996-2018). We call these words and phrases "text-novel concepts" and identify more than 100 in each subsector. We then scan the text of the entire patent corpus for prior patents (outside the agricultural subsector) that also mention these text-novel concepts. For example, in animal health, the word "pyrimethamine" and the phrase "equine protozoal myeloencephalitis" do not appear in any animal health patents prior to 1996, but are relatively common thereafter. In this case, we interpret prior mentions of "pyrimethamine" in human health patents prior to 1996 as evidence of a knowledge spillover from outside agriculture.

Our main finding is that knowledge spillovers from outside agriculture are important and influential for agricultural R&D, possibly as much as knowledge generated within agricultural science domains. In three of the subsectors studied—animal health, fertilizer, and machinery—every one of our proxies for knowledge flows originates outside of agriculture more than 50% of the time. In two additional sectors—biocides and research inputs—we have mixed evidence, but the majority of our proxies still originate outside of agriculture over 50% of the time. Only in the plants subsector do we typically find most knowledge flows point to agricultural technologies and research, though even this is not unanimous.

A second finding is that the non-agricultural domains that are important sources of knowledge spillover to agriculture are, in some sense, "close" to agriculture. It is typically more common for agricultural patents to cite or share text-novel concepts with the (non-agricultural) patents of firms that have at least one agricultural patent, even though the majority of patents belong to assignees with zero agricultural patents. Likewise, it is more common for citations to non-agricultural science journals to go to biology and chemistry journals than other journals, even though other journals account for the majority of journals.

[3] In general, there seems to be less cause for concern about bias in citations to the scientific literature (Roach and Cohen 2013).

Lastly, we demonstrate how text analysis can be a useful complement to citation-based measures of knowledge flows. In some cases, our text analysis suggests areas where citation-based results may be misleading. For example, in the biocides sector, the majority of patent citations flow to non-agricultural patents and journals. However, we find the majority of text-novel concepts for this sector (typically chemical names) have no prior mention outside of agriculture. It seems many of these chemicals appear for the first time in the patent corpus as part of a biocide patent.

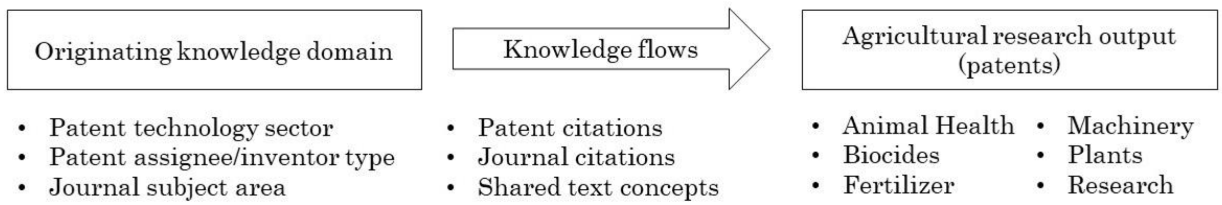
Our text analysis, in principle, has the ability to capture much deeper roots of knowledge than citations. It may be that an idea developed originally far outside of agriculture eventually enters agriculture via a long chain of citations. Because citation-based measure of knowledge spillovers only observe the last step in such chains, when an agricultural patents makes a citation, they may understate the role of non-agricultural knowledge spillovers. In contrast, our text analysis lets us track *all* prior mentions of important concepts used in agricultural R&D, including mentions that are many steps removed from agriculture. Consistent with this notion we find the share of text-novel concepts that originate in agricultural patents is smaller when measured using text than with citation (except in the biocides sector).

The rest of the paper is organized as follows. Section 1 describes our methodology for generating data on agricultural R&D output, knowledge flows, and originating knowledge domain and gives an example. Section 2 presents our main results. Section 3 discusses these results, and section 4 establishes that they are robust to a series of alternative assumptions. Section 5 concludes with some directions for future research.

1. Data

Our goal is to measure the extent of knowledge spill-ins for agricultural R&D. To accomplish this, we require three elements: a measure of agricultural research output, a measure of knowledge flows, and a measure of originating knowledge domain. These three components, plus our proxies for them, are illustrated in Figure 1.

Figure 1. Knowledge Spill-ins and Proxy Elements



Working from right to left, to measure agricultural research output we use patents with primarily agricultural application. Our paper focuses on six agricultural subsectors: animal health, biocides, fertilizers, machinery, plants, and research tools. We describe our method for identifying these patents in section 1.1. We measure knowledge flows in three ways: patent citations to other patents, patent citations to academic journals, and shared patent text. We describe how we generate these three proxies in section 1.2. We also define the originating knowledge domain in three ways: with patent technology classes, with assignee type, and with journal subject areas. We describe these methods in section 1.3. Section 1.4 provides some brief summary statistics for our data.

1.1. Measuring Agricultural Research Output

We use the universe of US patents granted between 1976 and 2018 for our analysis, though for some subsectors we only have data through 2015. Over this period, 5,886,981 patents were granted. While we use this entire dataset in our analysis, we are particularly interested in the subset of patents closely related to agriculture. Conceptually, our guiding principle is to identify patents over technologies used primarily in either agricultural production or agricultural research. We attempt to exclude patented technologies that have many applications, but where agriculture is not the primary use. For example, the CRISPR gene editing technology has applications in agriculture, but also many more applications in human medicine and fundamental research. We include only the subset of CRISPR patents closely related to agricultural research.

Our analysis is focused on six agricultural subsectors where we are able to identify related patents with relatively high precision: animal health, biocides, fertilizer, machinery, plants, and research inputs. While we feel these capture a large share of the major technological developments in agriculture over the last 40 years, we do not claim our analysis is exhaustive. In particular, the livestock genetics sector does not rely on patent protection to the same extent that the crop genetics sector does, and so we lack any information on this important sector. Another notable sector we are missing is information technology (e.g. software) applied to agriculture, for which we lack reliable means of identifying software with primarily agricultural application from others. Also, note that our analysis does not extend to the processing of agricultural products, either into food, feed, or biofuel.

With one exception (described below), our classification of patents starts with the cooperative patent classification (CPC) system. The CPC system is used by the US Patent and Trademark Office (USPTO) to classify patents into different technology categories, to facilitate USPTO patent examiners (and other interested parties) in finding relevant prior art. We use the *cpc_current* file, available on the USPTO's patentsview website, as our primary source. Patents are generally assigned multiple classifications, but we use only the primary classification for the purpose of allocating patents to a particular group.

For the biocide, fertilizer, and machinery subsectors, we identify CPC codes associated with the relevant sector and assign patents with identified codes as their primary classification to the relevant sector. Here we briefly describe our approach. A complete list of patents by subsector is available in the supplemental materials.

Biocides: This subsector includes fungicides, herbicides, insecticides, pesticides, and other chemicals meant to control biological pests. We start with CPC classification A01N, which includes these chemicals as well as chemicals for the preservation of bodies. We include any classifications under A01N related to biocides, but exclude classifications related to the preservation of bodies (which tend to begin with A01N 1/).

Fertilizer: This subsector includes chemical fertilizers. We use CPC classifications beginning with C05, which corresponds to chemical fertilizer technology.

Machinery: This subsector includes agricultural machinery, with a focus on mechanically powered machinery. Within the CPC classification A01, we include any classification related to agricultural machinery (e.g., harvesting, mowing, planting, milking, etc.), and exclude many other categories unrelated to machinery (e.g., structures, forestry, fishing, hunting, and most of the other agricultural subsectors considered). Most of our ag machinery patents are classified under A01B, A01C, A01D, and A01F. Within the machinery categories, we also exclude classifications related to hand tools and animal driven machinery.

These three subsectors require no additional processing. For the plant cultivar and ag research tools subsectors, the CPC classification system is not sufficiently precise for our purposes, so we supplement the CPC approach with manual cleaning.

Plants: This subsector includes utility patents for specific plant varieties/cultivars.[4] We begin with the set of patents assigned primary CPC code A01H, which includes both patented plant cultivars and plant modification and reproduction techniques, as well as related technologies. We exclude CPC codes related to non-agricultural plants and fungi. From the remaining set, we manually identify patents for plant cultivars by inspection of the patent title, abstract, and claims.

Biological Research Tools: This subsector (hereafter shortened to "research tools") includes technologies for conducting biological research, for example, genetic engineering and traditional breeding techniques. We begin with CPC classifications under the category A01H that are related to processes for modifying agricultural plants, and add some classifications under CPC class C12N (microorganisms and enzymes) that are specifically designated as being for the modification of plants. Note A01H also includes plant cultivar patents; we exclude any patents that are already classified in the plants subsector.

Animal Health: This subsector includes all patents associated with medical technologies approved for use in veterinary medicine by the FDA.

To obtain data on animal health patents, we adopt a different approach than for the other subsectors. While the CPC system suffices to identify patents related to medical technology, it does not distinguish between medical technologies for human versus non-human animal application. Instead, to identify patents for veterinary medicine technologies, we rely on US Food and Drug Administration (FDA) archival data. To facilitate generic competition in the animal health market, since 1989 the FDA has maintained a list of patents associated with all approved veterinary medicine products. Using archival records of this list, Clancy and

[4] Note that this subsector does not include "plant patents," a distinct form of intellectual property dating to 1930 and applicable to asexually reproduced plants (Clancy and Moschini, 2017).

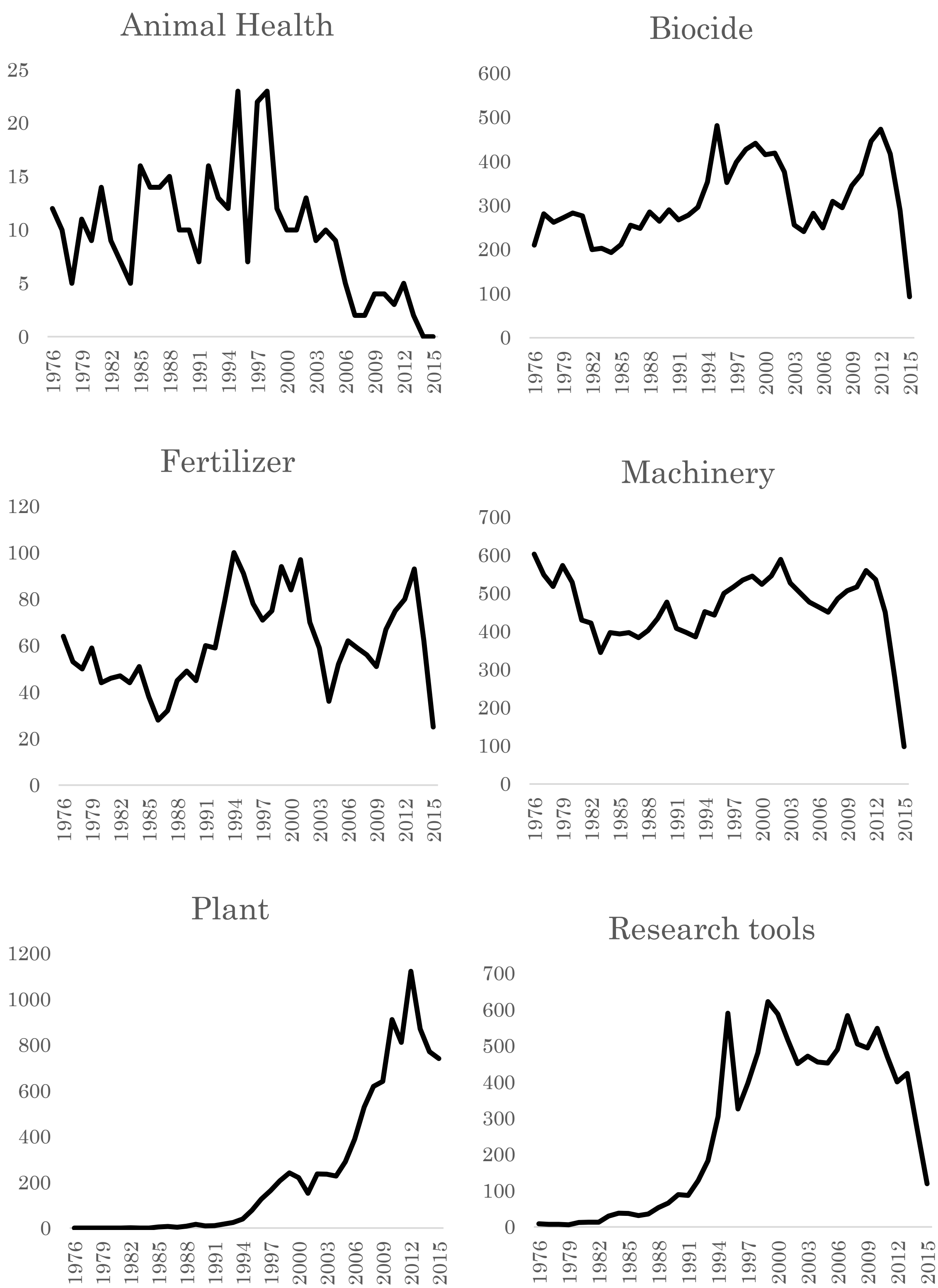
Sneeringer (2018) develop a list of all patents associated with approved veterinary medicine products.

It should be noted that the patents in the animal health subsector are subject to a selection effect that is not present in the other sectors. This is because animal health patents are only included if they are associated with veterinary drugs that eventually receive FDA approval. Drugs that are not approved may have associated patents, and we miss these. This selection effect may bias our results for this subsector in two ways. First, if successful and unsuccessful drugs enjoy spill-ins at differential rates, our results will only apply to successful drugs. In our robustness checks, however, we find little evidence in other subsectors that the most valuable patents differ dramatically in their citation patterns. Second, and perhaps more importantly, by omitting patents associated with unsuccessful drug applications, we will mis-classify citations to these patents as citations to non-agricultural patents. This may partially account for our finding that animal health relies more on non-agricultural knowledge flows than other agricultural subsectors (although there are, of course, other plausible explanations for such a finding).

Figure 2 illustrates the annual number of (granted) patents, by application year, in each of these subsectors. A few preliminary observations are in order. First, most subsectors exhibit a sharp decline in patents in the last few years of the sample. This is due to a truncation effect: we only observe patents if they are granted by 2016 in most sectors (we have data until 2018 for our plants and research tools subsectors) and few patents applied for in 2014 and 2015 are granted by 2016.

Second, the plants and research inputs subsectors exhibit a sharp increase from zero (or close to zero) in the 1980s. This is due to legal changes in the patentability of biological innovation in the wake of the 1980 Diamond v. Chakraborty Supreme Court case (Clancy and Moschini 2017). Prior to 1980, biological innovations such as new plant varieties were not patentable subject matter. It is important to note that any R&D related to biological innovation that occurs prior to 1980 is unlikely to be reflected in the patent record.

Figure 2. Number of Granted Patents, by Application Year and Subsector, 1979-2016



Finally, note that the scale of the vertical axis in Figure 2 varies substantially across sectors. In our dataset, the animal health sector has the smallest number of patents (414) and the machinery subsector has the most (19,362). Because of the variability in the size of subsector, how long innovation in the subsector has been eligible for patent protection, and the presence of selection effects in the animal health subsector, in this paper we always report disaggregated results by subsector.

1.2. Measuring Knowledge Flows

Our first measure of knowledge flows are patent citations to other patents. We use the USPTO patentsview dataset *uspatentcitation* as our source for patent citations. This provides the patent number of both the citing and cited patent, and identifies who added the citation (the applicant, examiner, or other parties), from 2002 onwards. Because we will be aggregating cited patents into different sectors and assignee-types, we limit ourselves to citations to patents granted between 1976 and 2016.

Our second measure of knowledge flows are patent citations to academic journals. We estimate public sector patents are just 2% of all patents granted in our observation period, far below the public sector's share of R&D (agricultural or otherwise). Accordingly, to measure the role of public sector R&D, it is important to supplement our patent citation analysis with journal citations. Analysis of citations to non-patent literature is complicated by the absence of standardized citation formatting. Patent applicants cite articles in a wide variety of ways: with or without abbreviations; using commas or periods to divide information; the order of author names, year, title, journal, volume number, etc. An emerging literature is attempting to match the raw citation text in patent documents to standardized journal entries in databases such as Clarivate (formerly Thompson Reuters) Web of Science, Elsevier Scopus, Google Scholar, Crossref, PubMed, and the Microsoft Academic Graph. We use Marx and Fuegi (2019), a dataset based on text analysis algorithms that matches raw patent text to entries in the Microsoft Academic Graph. Marx and Fuegi (2019) estimate they capture 90% of citations with 99% accuracy.

Our third measure of knowledge flows is a novel use of patent text, extending approaches pioneered by Packalen and Bhattacharya (2015) and Balsmeier et al. (2018). We identify a large set of "text-novel concepts," proxied by one-, two-, and three-word strings of text, that are popular in agricultural innovation in the second half of our dataset, but absent from the first half. We find all mentions of these text-novel concepts in other patents and use earlier mentions of the concept as a measure of potential knowledge flow. Because this approach is novel, we describe it in some detail here.

The goal of this approach is to identify strings of text in patents that proxy for concrete ideas and concepts with technological applications. Following Packalen and Bhattacharya (2015), we define a "concept" as a text string consisting of one, two, or three words, without separating punctuation between them (i.e., hyphens are permitted).

For a given agricultural subsector's patents, we break the text of the title, abstract and claims into concepts. This includes all individual words, as well as all sequences of two or three words, as long as the words are not divided by punctuation (with the exception of hyphens). We focus on the title, abstract, and claims because these likely are most informative as to the important concepts in a patent: titles and abstracts are meant to succinctly describe the innovation, while claims are legally binding.

We next clean the text of these concepts, using an approach similar to Packalen and Bhattacharya (2015). We convert all text into lowercase letters. We then exclude concepts with numbers as one of the words, concepts were words are divided by punctuation, or concepts which are unusually short and long (in terms of their total number of characters) [5].

This leaves us with a very large set of text, most of which does not correspond to ideas and concepts with technological application. To focus on new ideas in agriculture, we next divide our dataset in half. The concepts in patents applied for in the first half of our observation period (1976-1996) form a *baseline dictionary*. The concepts in patents applied for in the second half of our observation period (1996-2016) form a set of *recent concepts*. Any recent concept that is not contained in the baseline dictionary is considered a novel concept. Intuitively, this is a string of text that did not appear in any of the subsector's patent abstracts, titles, or claims prior to 1996, but does appear after 1996.

[5] Following Packalen and Bhattacharya (2015), we exclude one-word concepts shorter than 3 characters or longer than 29 characters, two-word concepts shorter than 7 characters or longer than 59, and three-word concepts shorter than 11 characters or longer than 89 characters. We also exclude concepts that include words in the python nltk.corpus stopwords list.

Next, we calculate the number of subsector patents that contain each novel concept in the abstract, title, or claim. We call these "mentions." For example, the word "trimethoprim" refers to an antibiotic. It does not appear in any animal health patents prior to 1996, but appears in 8 patents after 1996. We therefore say "trimethoprim" is a novel concept with 8 mentions.

Our goal is to identify a set of important agricultural concepts. To do this, we first identify the 200+ novel concepts with the most mentions. We frequently identify more than 200 concepts in this first pass, because mentions are necessarily integers and usually there are multiple concepts with the same number of mentions as the 200th concept. By construction, these are strings of text that did not appear in any of the sector's patent abstracts, titles, or claims prior to 1996, but which were relatively common after 1996.

To increase our confidence that our concepts are good proxies for concrete ideas and concepts with technological application, we go beyond Packalen and Bhattacharya (2015) and Balsmeier et al. (2018) and manually clean the set of candidate concepts using the following four guidelines. We exclude:

1. Concepts with numbers and measurements: These are unlikely to correspond to generalizable ideas or concepts, as they usually refer to specific measurements that are not good proxies in the absence of more context. Examples: "90 degrees", "1,500 ml"
2. Connective phrases: These are largely free of concepts and ideas with technological application, and instead likely reflect variation in preferred patent language. Examples: "combinations thereof", "one particular type"
3. Words with multiple context-dependent meanings: When a set of words can have significantly different meanings in different contexts, then it is a poor proxy for our purposes because it may be mentioned in multiple patents with no technological similarity. Example: "artificial" (which could be paired with "intelligence", "insemination", "sunlight")
4. Concepts including uninformative words: If some of the words in a concept appear to be valid (not excludable by any other criteria), but they only appear in conjunction with an additional word that is uninformative (e.g., "said" or "and"), we exclude the concept. In these cases, it is likely the concept is not really novel, but only the

conjunction of the concept and the uninformative word. Example: “said data structure”, “the database” (if “data structure” and “database” do not appear as novel concepts themselves, then they were in use in 1976-1996, only the exact formulation adding “said” or “the” was not).

Three of the coauthors independently examined the list of candidate concepts, based on the foregoing four criteria, and any concept excluded by at least two of the three coauthors was removed. This exclusion criteria removes 37% of the top 200 concepts overall, with a low of 11% in biocides and high of 47% in machinery. As a robustness check, we re-perform our analysis on the set of concepts that are retained unanimously by all these coauthors. What remains constitutes our set of “text-novel” concepts. They form a set of text proxies for concrete technological ideas that are important in agricultural innovation over the period 1996-2016, and are new at least in the sense that they were not used over 1976-1996 in patents. In some cases, the underlying concepts are not actually new, but represent one of two things: first, the discovery of new applications for ideas that had been in a state of dormancy over 1976-1996; and second, an expansion of the use of technological terms from the scientific literature to patent text. This latter phenomenon is often the result of an expansion of patentability, as in the case of utility patents for plant cultivars. For patents granted after 1996, depending on the subsector anywhere from 17% (in machinery) to 94% (in plants) of patents mention one of the associated text-novel concepts. See table 5 for the breakdown by subsector.

The top 10 text-novel concepts in each subsector are listed in Table 1. See the appendix for a complete list of top text-novel concepts in each subsector, broken down by those unanimously retained (the majority) and those retained only by two out of three reviewers (which are excluded in a robustness check). A cursory look at table 1 illustrates how text-concepts align with our intuitions about the knowledge base in different fields: animal health, plants, and research tools all involve biological terms; biocides is mostly chemical names; machinery includes different mechanical components, and so on. In our main specification we give equal weight to all concepts, but in our robustness checks we show our results are robust to the clustering of concepts into families of related concepts.

To identify potential knowledge flows, we identify any patents (whether agricultural or not), that mention these concepts. To do this, we again break the text of each patent’s title, abstract, and claims into concepts, clean the text of these concepts, and identify any concepts that match the set of text-novel concepts in agriculture. These form the set of all patents (agricultural and otherwise) that mention any text-novel concepts in agriculture. We interpret such mentions as informative (albeit noisily) of knowledge flows, and indicative that relevant research was ongoing in the sector to which agricultural researchers may have been exposed.

Table 1. Top ten text-novel concepts by patent subsector, 1996-2016

	Top ten text-novel concepts
Animal Health	Protozoal, trimethoprim, microbial, microbial infection, ear, preservative, terbinafine, penetration enhancer, kinase, bird
Biocides	Thiamethoxam, azoxystrobin, clothianidin, trifloxystrobin, spinosad, acetamiprid, thiacloprid, prothioconazole, pyraclostrobin, emamectin
Fertilizer	Selenium, itaconic, tea, canola, mean particle, chlorine dioxide, wetting agents, phosphite, ferrate, compost tea
Machinery	Controller configured, actuator configured, apparatus configured, antenna, dairy livestock, arm configured, flexible cutterbar assembly, controller operable, opening configured, gps receiver
Plants	Insect resistance, transgene, conversion, locus, trait selected, locus conversion, carbohydrate, backcross, metabolism, carbohydrate metabolism
Research tools	Clustal, one regulatory sequence, silencing, polynucleotide selected, isolated polynucleotides, chimeric gene results, polynucleotide operably linked, polynucleotide operably, polyunsaturated fatty acids, Rnai

1.3. Originating Knowledge Domains

To measure the source of knowledge flows, we define the originating knowledge domain in three ways. Our first approach is simply to leverage our work identifying patents in distinct agricultural subsectors. When a cited patent, or a patent linked by common text, belongs to one of our agricultural subsectors, we use the subsector as the originating knowledge domain. We find it useful, in general, to group these sectors by “own subsector” (for example, an animal health patent citing another patent belonging to animal health), “other agriculture” (for example, an animal health patent citing an agricultural research tools patent), and “not agriculture” (for example, an animal health patent citing a human health patent).

1.3.1. *Assignees*

Our second approach relies on the assignees and inventors associated with patents. Most patents have an assignee, usually corresponding to the employer of one of the patent's inventors, and all patents have an inventor (or inventors). We are interested in distinguishing between assignees that are specialized in agriculture, assignees that conduct agricultural R&D but for whom it is not their primary focus, and assignees that conduct no agricultural R&D.

The problem of *assignee disambiguation* and *inventor disambiguation* in patents is an active area of research. In brief, this is the challenge of determining when two patents belong to the same assignee or inventor. What makes this challenging is that the USPTO does not assign unique IDs to inventors and assignees. Instead, assignees and inventors are listed as text in the patent document. The same set of text (e.g. "John Smith") may refer to different individuals/assignees. Or, different text (e.g. "IBM" and "International Business Machines") may refer to the same individual/assignee.

We primarily rely on the disambiguation dataset built by Balsmeier et al. (2018). These authors begin with the hand-curated NBER patent data project, which matched patents granted between 1976-2006 with publicly traded companies in the compustat dataset. Balsmeier et al. (2018) then use a k-nearest neighbor clustering algorithm for the remaining patents. This algorithm identifies the five assignees "closest" to the unmatched patent's assignee, in terms of having similar inventors, CPC codes, locations, and cited patents. It compares the assignee name of the unmatched patent to the names of these five nearest assignees and takes the closest match, provided the similarity of this match exceeds a threshold. Otherwise, a new assignee is added to the dataset. A similar technique is used to disambiguate inventors.

We use Balsmeier et al. (2018) to differentiate between patents with assignees and those with individual inventors. However, assignees can take many forms: private firms, government agencies, non-profit organizations, and even individuals different from the inventor who are assigned the patent. Balsmeier et al. (2018) do not distinguish between different kinds of assignees. We attempt to separate public sector assignees from private sector ones, and then to characterize the extent of agricultural specialization for private sector assignees.

We adopt two approaches to identifying public sector assignees. First, the USPTO's patentsview *assignee* and *patent_assignee* files indicate whether an assignee is a government agency (state, federal or foreign). We classify the assignees of any patent with all government agency assignees as public sector assignees. Second, we use a list of keywords[6] to identify major non-governmental agency public sector assignees. Any assignee that includes one of these keywords is also classified as a public sector patent.

Patents not classified as belonging to the public sector or individual inventors belong mostly to private sector firms. We are interested in characterizing the extent to which these firms R&D focus is agricultural. We face two challenges here: ascertaining the extent of agricultural R&D, and determining how to classify assignees that change their research focus over time. Some major firms dramatically reinvented themselves as agricultural companies over our observation period (Monsanto is a notable example), and so we need a way to distinguish between different phases of the firm's existence.

We use the share of patents classified as belonging to one of our agricultural subsectors to determine an assignee's agricultural focus. To capture the fact that assignees may change their research focus over time, we use only patents granted in the preceding five years to construct a time-varying, assignee-specific agricultural focus.[7] While we use this continuous measure of agricultural R&D focus, we also construct three types of assignee, where types can change year-to-year:

Specialized Agricultural Assignee: A firm for which 50% or more of their patents, granted in the last five years, belong to one of our 6 agricultural subsectors.

---

[6] Keywords include: university, universities, college, colleges, institute of technology, foundation, school, polytechnic institute, virgina tech, Argonne, Tulane education, board of regents, universita, universitat, universite, universidad
We find these keywords largely match the number of patents granted to US colleges and universities, as reported by the USPTO and NSF, in 2011 (USPTO 2019).

[7] When we do not have data on five prior years of patenting (i.e., in the first four years after an assignee begins to patent, or the first four years in our dataset), we use the patents granted in the first available five years or the maximum number of years available if five are not available. For example, for a patent granted in 1977, we use patents granted in 1976-1980 to determine the assignee type in 1977.

Minority Agricultural Assignee: A firm that has at least one agricultural patent in the last five years, but for which less than 50% of their patents, granted in the last five years, belong to one of our 6 agricultural subsectors.

Non Agricultural Assignee: A firm with no patents granted in the last five years that belong to one of our 6 agricultural subsectors.

Our choice of five years balances two competing desires. A shorter time window introduces more noise into our estimates. A longer time frame is slow to recognize when a firm reorients its R&D focus. To assign firms a position in technology space, it is common to use the entire period under observation (see for example, Greenstone, Hornbeck, and Moretti 2010 and Bloom, Schankerman, and Van Reenen 2013), and so our time five-year lag is relatively short. We find that using a longer time-window results in fewer firms that we classify as specialized ag firms. Therefore, if we used a longer time frame, it would likely strengthen our conclusion that non-agricultural firms are a major source of knowledge flows in agriculture.

Approximately 5% of patents lack disambiguated assignee data in Balsmeier et al. (2018) and we assign these to an "unclassifiable" category. When a patent has multiple assignees spanning different types, we fractionally allocate the patent across different assignee types. Lastly, note that there is no concordance between assignees in the USPTO patentsview data and the Balsmeier et al. (2018) dataset. In the rare case (less than 1.5%) where a patent has multiple assignees, and some but not all are indicated as government agencies by the USPTO datasets, we cannot determine which of the assignees in Balsmeier et al. (2018) are the government agencies (text similarity matching fails). We allocate this small number of patents to the unclassifiable category.

Based on these criteria, 55% of all patents over our observation period belong to non-agricultural assignees, 23% belong to minority agricultural assignees, 15% to individuals, 5% are unclassifiable, 2% belong to public sector firms, and 0.5% belong to specialized ag firms. For comparison, patents in any of our agricultural subsectors account for 1% of all patents granted over the period. Note this implies the agricultural patents of minority ag firms account for slightly more than 3% of their patents.

Table 2 displays the four assignees with the most patents in each agricultural subsector. As expected, they largely correspond to well-known firms.

Table 2. Top four patent-holding assignees by subsector, 1976-2016

	Top four assignees by patent holdings
Animal Health	Pfizer Inc., Eli Lilly and Company, Alza Corporation, Hoechst Aktiengesellschaft
Biocides	Hoechst Aktiengesellschaft, BASF Aktiengesellschaft, Sumitomo Chemical Company Limited, CIBA Geigy Corporation
Fertilizer	Union Oil Company of California, Tennessee Valley Authority, OMS Investments Inc., Allied Signal Inc.
Machinery	Deere & Company, CNH America LLC, Unisys Corporation, J I Case Company
Plants	Pioneer Hi Bred International Inc., Monsanto Technology LLC, Stine Seed Farm Inc., Syngenta Participation AG
Research tools	Pioneer Hi Bred International Inc., E I Du Pont De Nemours and Company, Monsanto Technology LLC, The Regents of the University of California

1.3.2. Journal Classification

Our first two approaches to defining the originating knowledge domain are only appropriate for knowledge flows that are proxied by patents (i.e., either cited patents or patents with shared text concepts). Here, we develop a third approach—appropriate for our journal citation proxy of knowledge flows—based on the classification of cited journals into broad academic categories. We create four main categories: agricultural science journals, other biology/biochemistry journals, other chemistry journals, and other journals.

Our list is based on the SCImago portal for the Scopus abstract and citation database for peer-reviewed literature. [8] Journals are placed in broad "subject areas," and within each subject area are more narrowly defined "subject categories." Journals can be placed in more than one subject category, and for that matter, in more than one subject area. To create the "agricultural science"

[8] https://www.scimagojr.com/

category, we start with two SCImago subject areas: (1) Agricultural and Biological Sciences; (2) Veterinary Sciences. Table 3 lists the subject categories within these two areas, and how the journals of each subject category are treated.

Table 3. Defining the set of Agricultural Sciences Journals

Agricultural and Biological Sciences	
Agricultural and Biological Sciences (misc)	Journals manually inspected
Agronomy and Crop Science	All journals included
Animal Science and Zoology	Journals manually inspected
Aquatic Science	Journals not inspected
Ecology, Evolution, Behavior and Systematics	Journals not inspected
Food Science	Journals not inspected
Forestry	Journals not inspected
Horticulture	All journals included
Insect Science	Journals manually inspected
Plant Science	Journals manually inspected
Soil Science	All journals included
Veterinary Science	
Equine	Journals not inspected
Food Animals	All journals included
Small Animals	Journals not inspected
Veterinary (misc.)	Journals manually inspected

Note that because journals can be cross-listed in several categories, it is possible for a journal to be designated an agricultural science journal, even if it belongs to one of the subject categories whose journals we do not inspect. This can occur, for example, if the journal is also listed in a category we do inspect. Eliminating duplicate entries results in a set of 981 journals classified as "agricultural sciences."

To create our set of "other biology/biochemistry" journals, we begin with all journals in the SCImago Agricultural and Biological Sciences area and Veterinary Sciences area that ended up not being included in the aforementioned agricultural sciences category. To this, we add all journals classified by SCImago in the "Biochemistry, Genetics, and Molecular Biology" subject area, and which were not already classified as Agricultural Sciences by us. This results in a set of 3,029 journals classified as "all other biology/biochemistry."

To create the “other chemistry” journal list, we combine all journals (not already classified in the preceding steps) from the “Chemistry” and “Chemical Engineering” subject areas in the SCImago set. This results in a set of 995 journals classified as “other chemistry.”

Lastly, all remaining journals in SCImago are classified as “Other.” It is important to note this category contains several prominent multidisciplinary journals, such as *Science*, *Nature*, and *PNAS*. We show our results are robust to separating out these three major journals into their own category. In all cases, we retain journals, book series, and trade journals, but mostly exclude conferences and proceedings volumes. This results in a set of 21,166 other journals.

A final challenge remains. Our source for journal citations is Marx and Fuegi (2019), which links the raw text in patents to entries in the Microsoft Academic Graph. We match journal titles in the Microsoft Academic Graph to journal titles in our SCImago classification system by a Levenshtein distance text-matching algorithm (we retain matches above 90% confidence). For Agricultural Sciences, we further manually check all journal matches. Table 4 illustrates the share of Microsoft Academic Graph journals that we successfully match to journals in the SCImago.

Table 4. Journal Match Performance

	Matched to SCImago Journals	Matched in MSAG to other Journals	Not Matched in MSAG to Journals
Animal Health	75.6%	16.9%	7.5%
Biocides	79.6%	10.2%	10.2%
Fertilizer	74.1%	11.9%	14.0%
Machinery	60.9%	10.1%	29.0%
Plants	73.0%	1.6%	25.4%
Research tools	92.4%	3.5%	4.1%

Note: MSAG denotes Microsoft Academic Graph. Column 1 is the share of patent citations to journals in the MSAG that we match to journals in SCImago. Column 2 is the share of citations in the MSAG that Microsoft indicates correspond to journals, but for which we are unable to match the entry to a journal in SCImago. Column 3 is the set of citations that Microsoft lacks enough information to match to a journal.

As indicated by Table 4, we always match the majority of journals and typically match approximately 75%. Our performance is worse in the machinery subsector (60.9%)—this is probably due to the fact that this is a field where citations to academic journals is rare and citations to conference proceeding papers (which we mostly exclude) are common. In the plants subsector, the Microsoft Academic graph is unable to match 25% of non-patent citations to journals. Manual inspection of a sample of these citations indicate they mostly accrue to books, which are also not in our dataset.

1.4. Summary

Table 5 provides a summary of our data.

Table 5. Summary Statistics

	Patents	Share Top 4 Assignees	Avg. Patent Cites Made	Avg. Non-Patent Cites Made	Share Patents w/ Text Concepts
Animal Health	414	24.9%	9.4	8.5	76.3%
Biocide	12,774	13.7%	8.3	6.5	24.2%
Fertilizer	2,554	3.7%	10.7	3.4	32.9%
Machinery	19,362	16.8%	13.2	1	16.7%
Plants	10,216	67.0%	7.6	9.2	94.4%
Research Tools	10,872	21.5%	7.5	37.3	41.6%

Note: Patents is the number of patents in the subsector. Share top 4 assignees is the share of these patents assigned to the four largest assignees. Avg. Patent Cites Made is the mean number of citations made to other patents, per patent. Avg. Non-Patent Cites Made is the mean number of non-patent references per patent. Share patents w/ text concepts is the share of patents granted after 1996 that mention one of the top text-concepts included in our text analysis.

Note the subsectors vary significantly in their propensity to cite, especially with respect to non-patent references (the majority of which are to academic journals). The machinery and fertilizer subsectors, for example, cite more patents than any other subsector, but the fewest non-patent references. Meanwhile, the research tools subsector cites non-patent literature at more than four times the rate of the next highest subsector.

Subsectors also vary in their concentration. Whereas fertilizer patents are dispersed among a plethora of small assignees, plant patents are highly concentrated in a small number of firms (with Monsanto and Pioneer alone accounting for more than half of all patents). Table 5 also highlights how our text analysis approach varies in how representative it is for different subsectors. Whereas the majority of patents granted after 1996 in Animal Health and Plants carry one of our text-novel concepts, only 17% of such patents in Machinery do (although, as the largest single subsector, the small share translates into thousands of patents).

While not a major focus of our paper, the extent of foreign research in our data is also of interest and can be proxied by the share of foreign inventors on a given patent. Using the US PatentsView inventors location dataset, we classify patents as being derived from foreign research if all the inventors have non-US addresses and as being derived from domestic research if all the inventors have US addresses. For patents with some, but not all inventors residing abroad, we classify the patent as fractionally foreign, based on the share of its inventors that reside abroad (i.e., a patent with 1 out of 2 inventors residing abroad is listed as 0.5 foreign patents). By this measure, plants have the fewest foreign patents (14%) and Biocides the most (49%). Table A1 provides a full breakdown by subsector.

1.5. An Example

As an example, consider patent 5,747,476, titled "Treatment of equine protozoal myeloencephalitis." The patent was applied for in July 1996 and granted in May 1998, and assigned to the Mortar & Pestle Veterinary Pharmacy, Inc. in Des Moines, Iowa. We classify this patent as an "animal health" patent. As the title suggests, it describes a novel treatment for equine protozoal myeloencephalitis (EPM), a debilitating neurologic disease that affects horses. At the time of the patent application, EPM was commonly treated by crushing two different kinds of tablets intended to treat humans – one with the active ingredient pyrimethamine and another with a trimethoprim-sulfonamide combination – and suspending the mixture in solution. This was fed to the horse prior to feeding, often for 90 days. The patent describes a new therapy, designed specifically for EPM that involves a compound of pyrimethamine and a sulfonamide ("preferably sulfadiazine"), but with a much smaller dose of trimethoprim (or none at all).

Such an innovation obviously builds on ideas developed outside of agriculture. Pyrimethamine was discovered in 1952 and developed into an anti-malarial treatment (for humans) in 1953 but has many applications in treating parasitic diseases. Sulfanomides have an even older history, forming part of the first set of antibiotics widely used (again, for humans) in the 1930s. However, their joint application in treating EPM is novel.

Patent 5,747,476 reflects these deep roots in several ways. It cites 10 patents, most of which have little to do with veterinary medicine (the oldest being US patent 4,293,547: Method of Treating Malaria, granted in 1981). We identify patents as belonging to agriculture if they belong in one of our agricultural patent datasets (which the cited ones do not), or if the assignees of these cited patents have other agricultural patents within the last five years. Where they do we find the share of agricultural patents is quite small. To take one example, patent 4,293,547 belongs to the Upjohn company, and only 1.1% of its patents were agricultural in 1981 (over the preceding 5 years).

Only one cited patent belongs to a publicly owned entity, patent 5,486,535 – Method of treating Toxoplasmosis, which is assigned to the regents of the University of California. To understand the patent's use of publicly funded knowledge, we instead turn to its 13 citations to journals. The cited references include the American Journal of Veterinary Research, the Canadian Veterinary Journal and the Journal of Parasitology. Of these, we classify the first two as agricultural science journals, and the last as a biology/biochemistry journal, suggesting this patent draws on both specific agricultural research and basic biology.

Finally, the text of the patent itself contains important concepts. The word "pyrimethamine" is absent in our animal health dataset for the first half of our observation period, but relatively common in the second half, so that it is one of our top text-novel concepts. The words "equine protozoal myeloencephalitis" are another concept that is absent over 1976-1995, but relatively common in animal health patents after 1995.

When we search the broader patent corpus for patents including the word "pyrimethamine" (in the title, abstract, or claims), we find many examples that predate its use in animal health (hardly surprising, given its history) not among the patents cited. These patents provide a third indicator that this patent draws on knowledge developed outside of agriculture. In contrast, the phrase "equine protozoal myeloencephalitis" appears for the first time in any US patent in patent 5,747,746. Beginning with this example, it goes on to appear in several other patents in animal health. In contrast to "pyrimethamine", the concept of (treating) "equine protozoal myeloencephalitis" is one that was born in agriculture, reflecting the primarily agricultural research base upon which it is based.

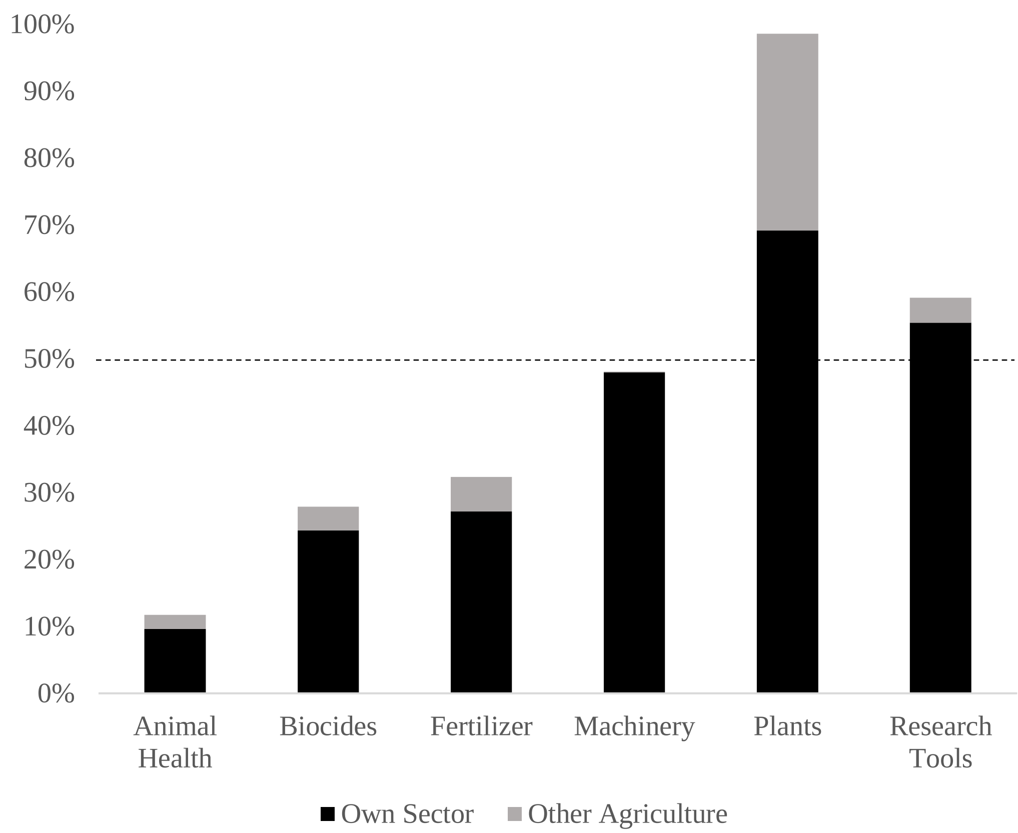
2. Main Results

We here present five different measures of knowledge spill-ins to agriculture. We begin with results that use patent citations, then present results that rely on citations to non-patent literature, and then results that use shared text concepts.

2.1. Patent Citations

In figure 3, we show the share of citations made by each agricultural subsector that originate in their own subsector (i.e., animal health patents citing animal health patents) and other subsectors (i.e., animal health patents citing research tool patents).

Figure 3. Share of Patent Citations to Agricultural Subsectors



Note: The citing sector is on the horizontal axis. Only cited patents granted between 1976 and 2016 are included. Each citation is counted once, even if multiple citations point to the same patent. Own sector gives the share of these citations to patents in the same subsector. Other agriculture gives the share of these citations to any other agricultural subsector. The remaining share of citations accrue to patents not contained in any of our agricultural subsectors.

It is apparent that for the first four agricultural subsectors, more than half of citations accrue to patents not classified as agricultural patents. This indicates a substantial role for knowledge spill-ins from outside agriculture. In these four sectors, the second most cited subsector is the own subsector. There is very little knowledge flow between different agricultural subsectors.

In contrast, the majority of citations in the plants and research tools subsectors accrue to patents that belong to these subsectors. While the research tools subsector still cites a substantial number of patents outside of agriculture (40.9%), in the plants subsector citations to other plant patents and to research tool patents account for almost 100% of all citations made.

Table 6 breaks down the share of citations from each subsector to the type of assignee/inventor associated with the cited patent. As noted in section 1.3, we divide non-individual assignees into four categories: assignees (mostly firms) specializing in agricultural R&D, assignees (mostly firms) that conduct some agricultural R&D, but for whom such activities are the minority, assignees (mostly firms) conducting no agricultural R&D, and the public sector (mostly government, universities, and not for profit organizations). We omit the patents of unclassified assignees, which never receive more than 1.5% of citations.

Table 6. Share of Patent Citations to Assignee Types

	Ag Specialized	Ag Minority	Non Ag	Public Sector	Individuals
Animal Health	1.8%	69.1%	18.4%	4.1%	6.2%
Biocides	8.6%	65.1%	13.2%	4.6%	7.8%
Fertilizer	17.4%	33.7%	20.7%	4.5%	23.5%
Machinery	33.5%	29.1%	8.8%	1.1%	27.5%
Plants	80.6%	5.4%	0.3%	12.8%	0.6%
Research tools	28.1%	38.2%	12.8%	13.6%	5.8%

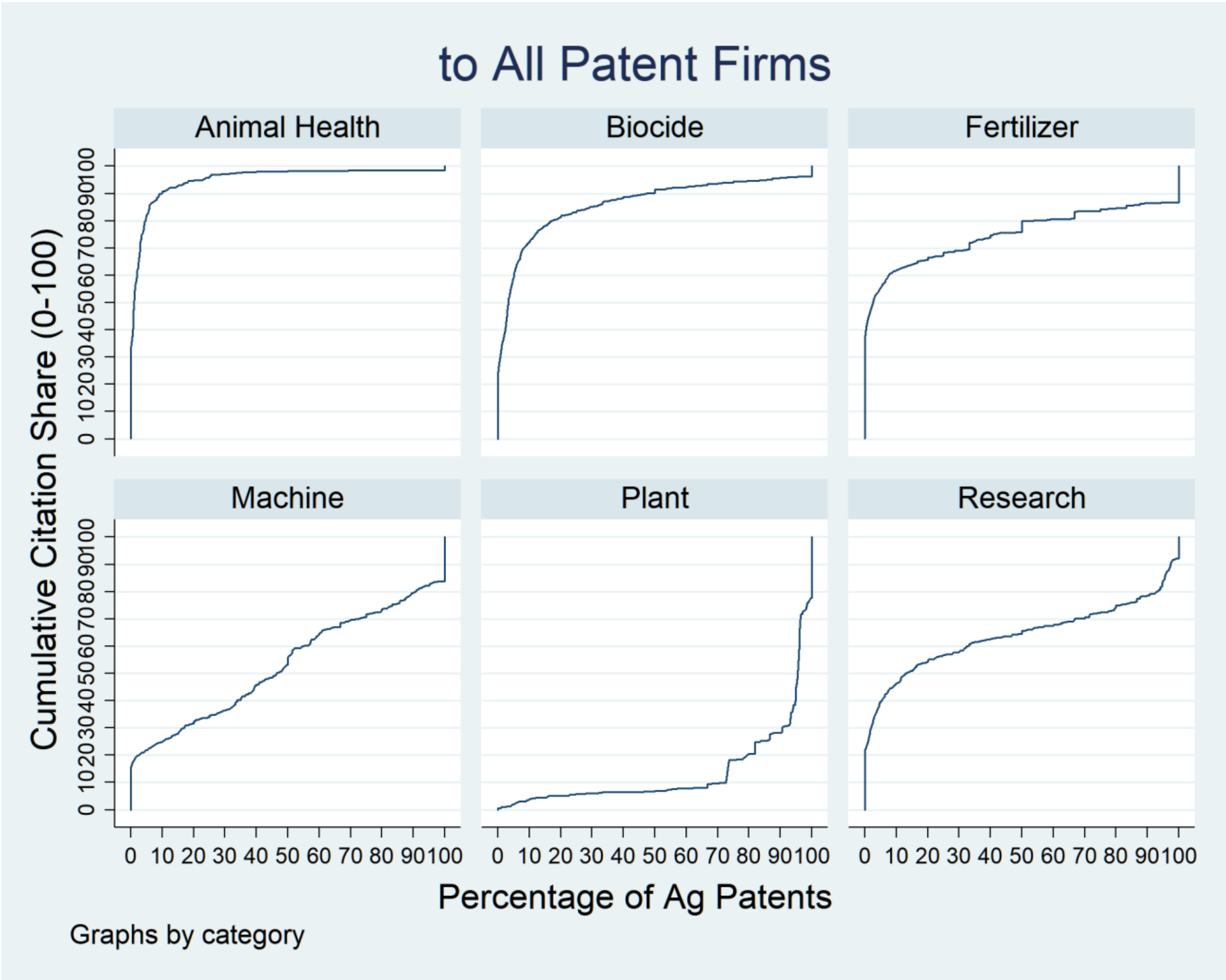
Note: The rows indicate the citing agricultural subsector and columns the assignee and inventor type to which the cited patents belong. Specialized ag assignees have more than 50% of their patents belonging to an agricultural subsector in the last 5 years. Minority ag assignees have more than zero patents belonging to an agricultural subsector in the last 5 years, but less than 50%. Non-ag assignees have no patents belonging to agricultural subsectors. Public sector assignees correspond to government and non-profit organizations. Individuals refers to patents owned by individual inventors. Rows do not add up to 100% - the remainder of patent citations are made to unclassified assignees (see section 1.3.1).

Only in the plants subsector do the majority of cited patents belong to assignees that specialize in agriculture. A plurality of patent citations in the machinery subsector also originate with assignees that specialize in agriculture. For animal health, biocides, fertilizer, and research, either a plurality or majority of patent citation originate in ag minority firms. In no sector do more than 21% of patent citations originate with assignees that do not conduct any agricultural research (even though these assignees account for 55% of all patents over this period). Public sector research is disproportionately important for all firms (considering that it accounts for just 2% of all patents), and especially important for plant and research tools patents.

Figure 4 presents more granular information on the agricultural focus of cited patents. For each point (x,y), share y of all citations made by the subsector accrue to patents belonging to firms with x% or less agricultural patents over the past 5 years. Note that this sample is conditional on the citation going to an assignee, and not a public sector organization or individual inventor.

The concave-to-convex curves in most of these figures tell us most citations go to firms that are either very specialized in agriculture (i.e., a very large share of the assignees patents are classified as agricultural) or have only a tiny agricultural R&D operation (i.e., a very small share of the assignee's patents are agricultural). Only the machine subsector is an outlier, with an approximately linear curve. No curve has a convex-to-concave "S" shape, which would characterize the presence of many cited assignees with agricultural focus near 50%. This suggests our division of assignees into ag-minority and ag-specialized is a reasonable one. It also suggests most of the ag-minority patents have only a tiny footprint in agriculture.

Figure 4. Share of Citations to Assignees by Agriculture-Specialization



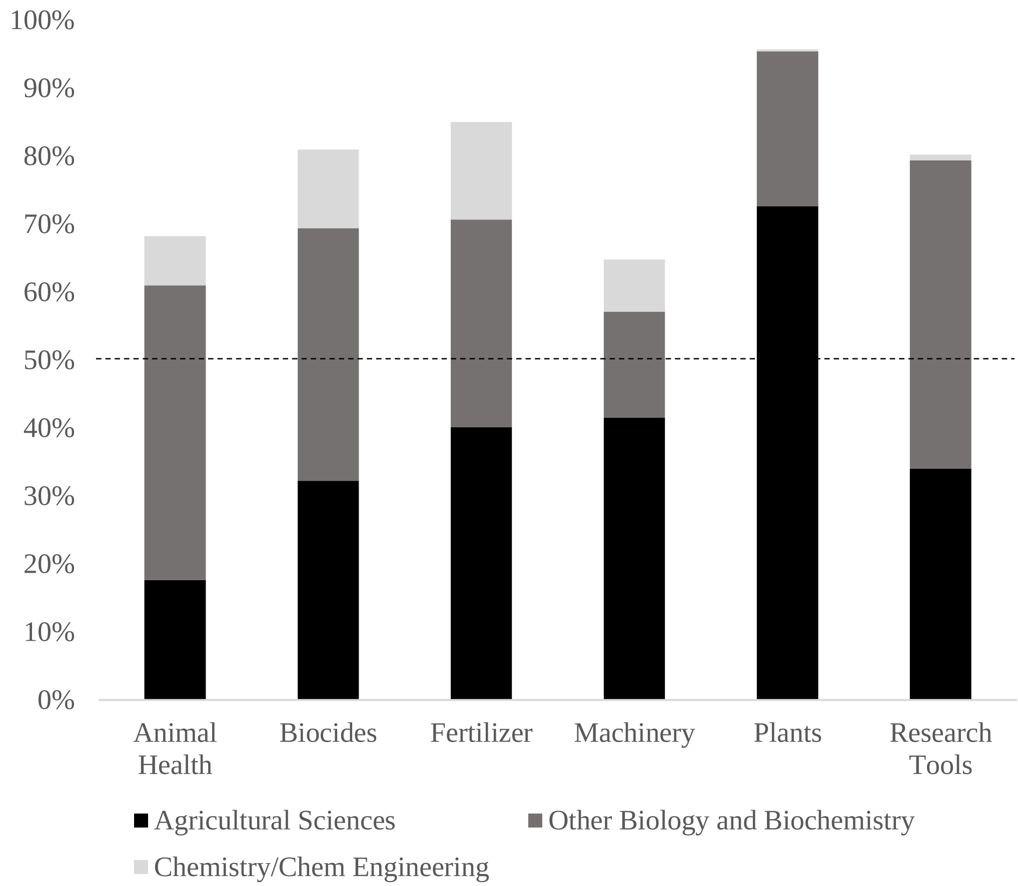
Note: Cumulative distribution function for citations, by agricultural focus of cited assignee. For each point (*x,y*), share *y* of all citations made by the subsector accrue to patents belonging to firms with *x*% or less agricultural patents over the past 5 years. Note that this sample is conditional on the citation going to an assignee, and not a public sector organization or individual inventor.

2.2.Journal Citations

In figure 5, we present the share of matched SCImago journal citations belonging to different journal categories.

Only in the plants subsector do the majority of cited journals belong to the agricultural sciences category. In the fertilizer and machinery subsectors, a plurality of cited journals belong to the agricultural sciences sector. With the exception of machinery, the other biology and biochemistry category is either the most or next-most important category of cited journals. In the machinery subsector, other journals are the second-most important source.

Figure 5. Share of Journal Citations to Journal Categories



Note: Citing agricultural subsector is listed on the horizontal axis. Shares are given conditional on matching journal title to the SCImago database. The remaining share of citations to journals accrue to other journals in SCImago that we do not classify as one of the above categories.

2.3. Shared Text Concepts

Our shared text concept results are designed to detect the sources of important new (or at least recently reawakened) concepts in agriculture. An important difference compared to the foregoing analysis is that whereas citations track knowledge flows "one step removed", our text approach can accurately track the "deep roots" of knowledge spill-ins. For example, an idea originating in a distant technology sector may pass through a long sequence of citations before finally being cited by an agricultural patent. To generate the following results, we perform the following calculation for each text-novel concept (see section 1.2) in each subsector. First, we identify the earliest subsector patent that mentions the concept. We use the application date of this patent as the date this text-novel concept is first applied in that subsector.

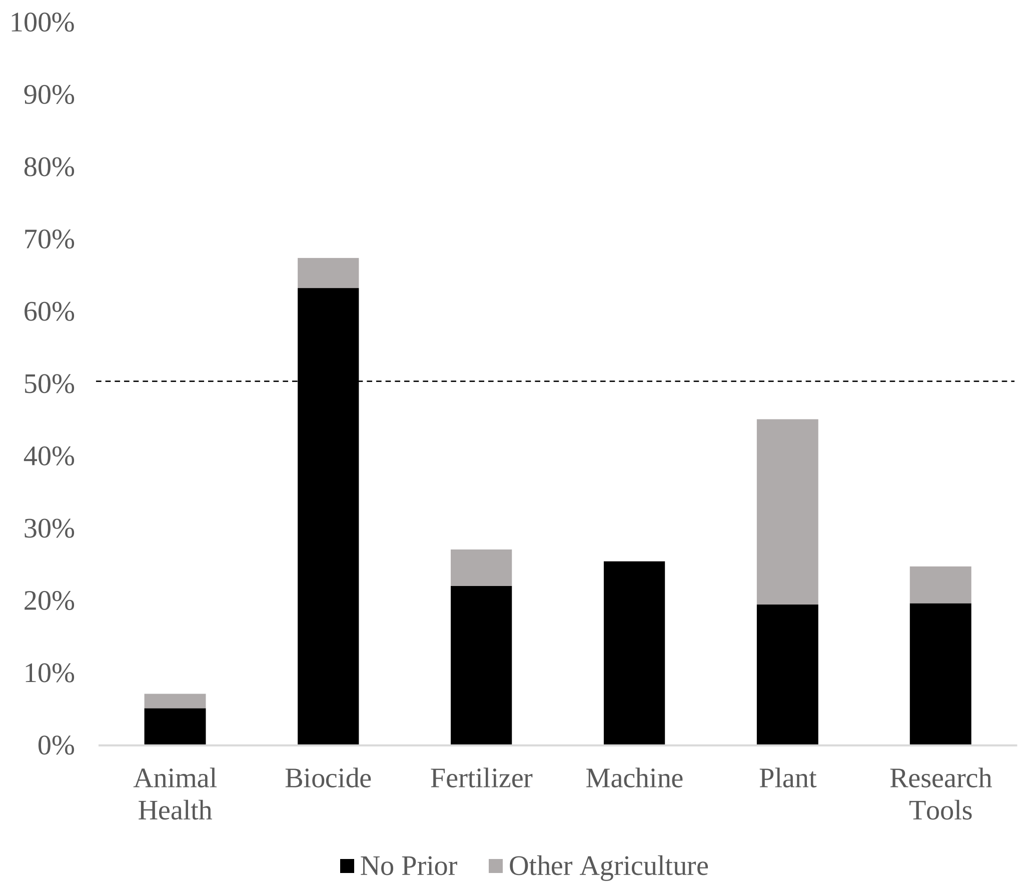
Next, we look for any mention of the concept in patents granted prior to this date. By construction, none of these patents will be in the "own subsector" prior to this date, but they may have been used in other agricultural subsectors, or outside of agriculture. If there are any antecedent patents mentioning the concept, we compute the share of these that belong to each originating knowledge domain. Denote the share of concept c's prior mentions originating in knowledge domain i by $s_i(c)$. If no prior patents mention the concept, we say the concept has no prior mentions ($s_i(c) = 1$, with i denoting "no prior mentions"). We then take the average share across all text-novel concepts:

$$p_i = \frac{1}{n}\sum_{c=1}^{n} s_i(c) \quad (1)$$

Intuitively, the interpretation of p_i is the probability a randomly selected knowledge flow from a randomly selected text-novel concept c originates in sector i. Figure 6 depicts the probability a random knowledge flow from a concept originate in agriculture.

In the biocides sector, fully 63% of top text-novel concepts appear for the first time in the patent corpus as part of the title, abstract, or claims of a biocide patent. This turns out to be an exception. Other than the biocides sector, the majority of text-novel concepts in each subsector are mentioned in earlier patents. The majority of these are mentioned by patents outside of agriculture. Again, there is little transfer of knowledge from within agriculture, with the exception of the plant subsector, where 20% of prior mentions come from the research tools subsector and 5% from the biocides subsector.

Figure 6. Probability of Antecedent Text-novel Concept Mentions across Agricultural Subsectors



Note: Each bar gives the probability a randomly selected patent mentioning a randomly mentioned text-novel concept originates in a given sector. Antecedent mentions are all those made by patents applied for prior to the first patent in the given subsector that mentions the concept. The remaining share of antecedent mentions accrue to patents not classified as agriculture.

Table 7 performs the same exercise for the type of assignee/inventor. Most text-novel concepts are mentioned before their use in agriculture by patents that do not specialize in agricultural R&D. This is consistent with figure 6, which establishes that most text-novel concepts are not mentioned in other agricultural sectors prior to their appearance in a given subsector. A large share of these concepts are mentioned, however, in firms with some agricultural research. The plurality of mentions occurs in minority ag assignees in four of the six sectors, whereas the plurality occurs in non-agricultural assignees in the other two (machinery and research tools).

Table 7. Share of Antecedent Text-novel Concept Mentions across Assignee-Type

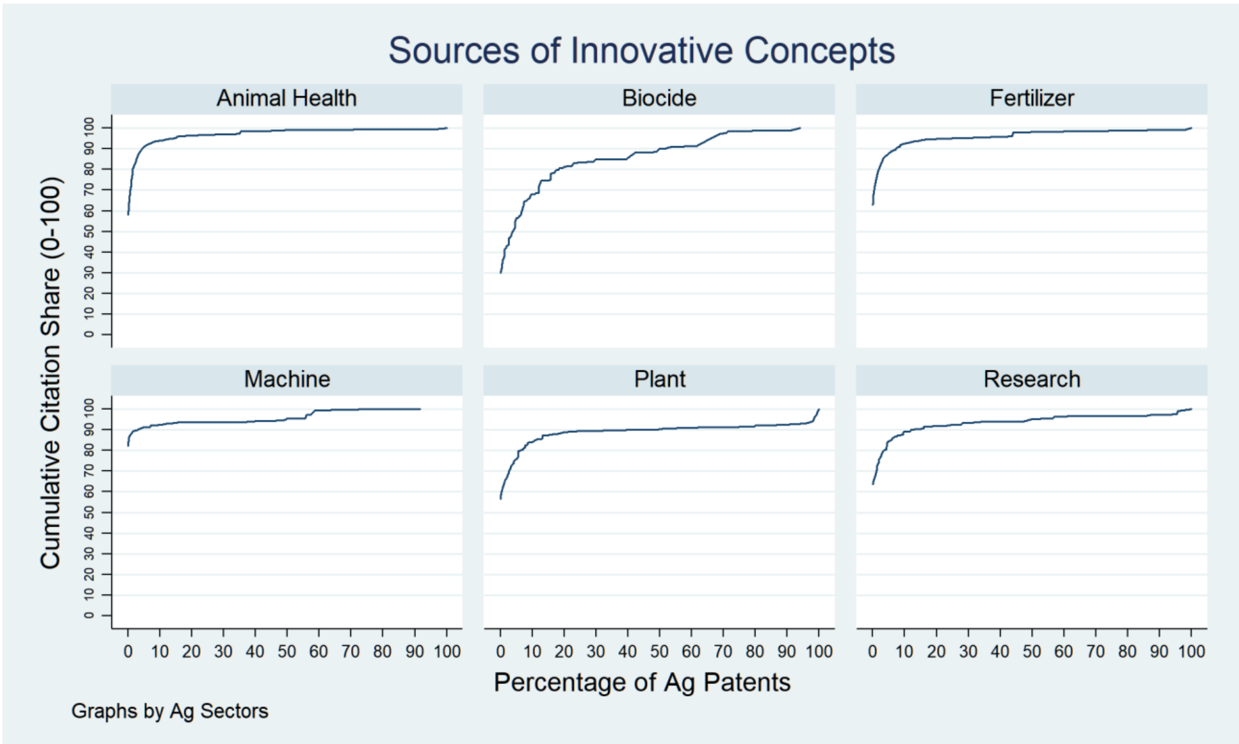
	Ag Specialized	Ag Minority	Non-Ag	Public	Individuals	No Prior
Animal Health	1.2%	44.1%	31.2%	7.0%	9.4%	5.1%
Biocide	3.5%	26.3%	4.8%	0.9%	0.3%	63.3%
Fertilizer	2.5%	29.8%	29.0%	4.3%	11.2%	22.1%
Machine	2.8%	16.1%	42.3%	1.0%	11.8%	25.5%
Plant	10.8%	28.7%	23.3%	10.4%	5.9%	19.5%
Research tools	2.1%	25.4%	30.3%	13.3%	7.2%	19.7%

Note: An entry gives the probability a randomly selected patent mentioning a randomly selected text-novel concept originates with a given assignee type. Antecedent mentions are all those made by patents applied for prior to the first patent in the given subsector that mentions the concept. Specialized ag assignees have more than 50% of their patents belonging to an agricultural subsector in the last 5 years. Minority ag assignees have more than zero patents belonging to an agricultural subsector in the last 5 years, but less than 50%. Non-ag assignees have no patents belonging to agricultural subsectors. Public sector assignees correspond to government and non-profit organizations. Individuals refers to patents owned by individual inventors. No prior indicates the concept has no prior mentions. Rows do not add up to 100% - remainder of patent mentions (0.1-1.5%) made to unclassified assignees (see section 1.3.1).

Figure 7 again presents more granular information on the agricultural focus of patents mentioning text-novel concepts. Any point (x,y) in figure 7 gives the cumulative probability x a randomly selected knowledge flow containing a randomly selected concept belongs to a patent with agricultural focus y or less. Note that this sample is even more restricted than figure 6, since it excludes the patents of public sector firms and individuals, as well as text-concepts that have *no* prior mentions.

Unlike figure 6, these shapes are mostly just concave, rather than concave-to-convex (the plant subsector being the only one showing a significant concave ending). This suggests that prior mentions by minority-ag assignees are mostly assignees with only a small agricultural focus – much less than 50%. For important text-novel concepts that are not born in agriculture, they tend to come from firms with either no history in agriculture or a very minor one.

Figure 7. Cumulative probability of Antecedent Text-novel Concept Mentions by Assignee specialization in agriculture



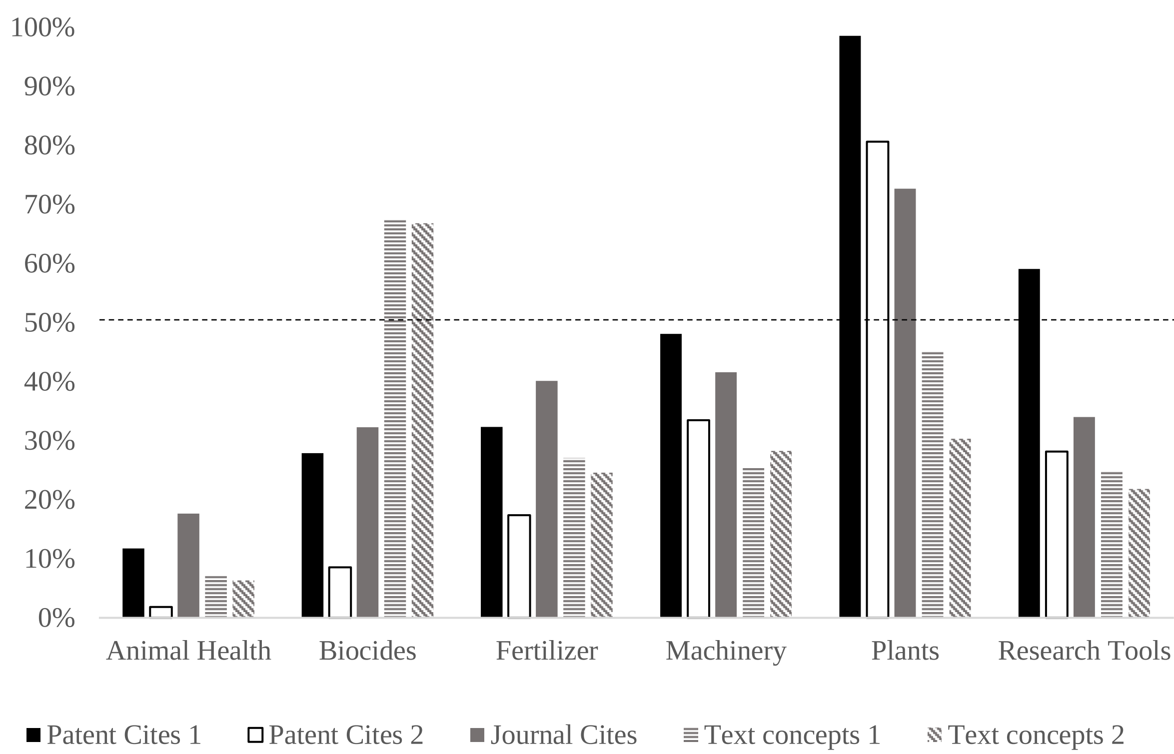
Note: Cumulative distribution function for prior mentions of text-concepts, by agricultural focus of cited assignee. Any point (*x*,*y*) in figure 7 gives the cumulative probability *x* a randomly selected knowledge flow containing a randomly selected concept belongs to a patent with agricultural focus *y* or less. Note that this sample is even more restricted than figure 6, since it excludes the patents of public sector firms and individuals, as well as text-concepts that have *no* prior mentions.

3. Discussion

Section 2 describes five different measures of the extent of knowledge spill-ins to agriculture. Each measure emphasizes a different potential aspect of spill-ins. Section 2.1 emphasizes the flow of knowledge in the space of patented technologies across our entire time period. Section 2.3 also focuses on the space of patented technologies, but focuses specifically on a subset of "concepts" that arose to prominence in agriculture during the second half of our observation period. It measures the extent of prior R&D (potentially many citations removed) related to these concepts outside of the particular agricultural subsector. Section 2.2, in contrast, examines the flow of knowledge from the primarily academic sector to patented technology.

Summarizing this heterogenous set of proxies is challenging, but one of our over-arching conclusions is that knowledge spill-ins from outside agriculture are likely as important as knowledge generated within agricultural domains. This conclusion is bolstered by figure 8, which indicates whether the share of knowledge flows that originate in an agricultural knowledge domain, defined below.

Figure 8. Share of Knowledge Flows Originating Within Agriculture



Notes: Patent cites 1 is the sum of own-sector and other-agriculture bars from figure 3. Patent cites 2 is the share of citations going to specialized-ag assignees in table 6. Journal cites is the share of journal citations to agricultural science journals from figure 5. Text concepts 1 is the sum of no-prior and other-agriculture bars from figure 6. Text concepts 2 is the sum of no-prior and ag-specialized categories in Table 7.

In this figure, we pull together proxies for the share of knowledge flows originating in agriculture:

- Patent Cites 1. Share of patent citations to agricultural subsectors in figure 3.
- Patent Cites 2. Share of patent citations to specialized ag assignees in table 6.
- Journal cites. Share of journal citations to agricultural sciences journals in figure 5.
- Text concepts 1. Probability a text-concept either has no-prior mention or a knowledge flow originating with an agricultural patent in figure 6.
- Text concepts 2. Sum of the no-prior mention and specialized ag columns in table 7.

By these definitions, the animal health, fertilizer, and machine subsectors source the majority (more than half) of their ideas from outside agriculture, as measured by any proxy.

The evidence is more mixed for the research tools and biocide subsectors. For research tools, 55% of patent citations refer back to other research tools patents and another 4% originate with other agricultural patents. However, most of these patents are assigned to firms that are not specialized in agriculture, and most of the text-novel concepts in research tools patents are mentioned in patents that lie outside agriculture. Moreover, research tools patents cite academic journals at four times the rate of any other sector, but only 34% of citations flow to agricultural science journals.

Biocide patent and journal citations primarily flow to non-agricultural firms, patents and journals. However, the strong majority of text-novel text concepts in biocides have no prior mention and appear for the first time in the patent corpus in a biocide patent. The majority of these concepts are chemical names, suggesting the subsector develops many chemicals for application in agriculture that appear nowhere else in the patent corpus. This is an observation that would be missed if we relied solely on citations.

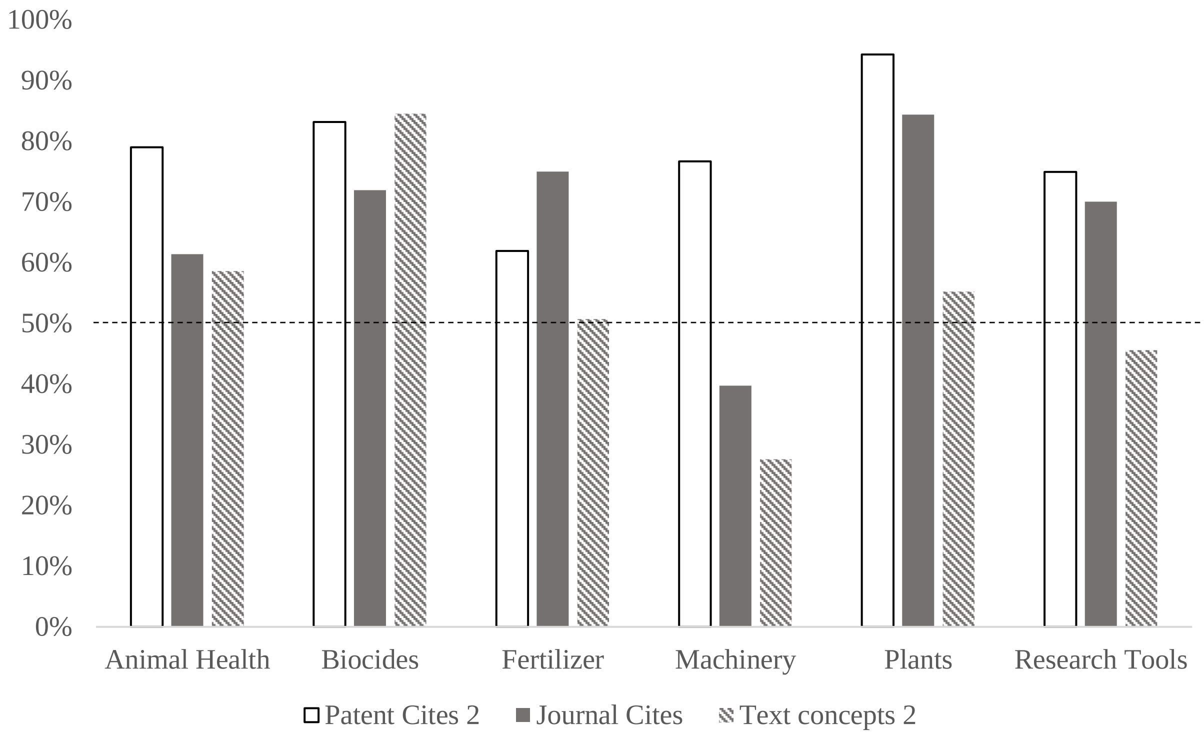
Finally, plants seem to be different. The majority of citations flow to specialized ag firms, agricultural patents, and agricultural science journals. For text concepts, the majority are mentioned in non-agricultural patents before their appearance in patent for plant varieties, but not by an overwhelming number (55%). It is important to note that utility patents for plants differ from other utility patents in more than just their subject matter. This field is dominated by an unusual extent by a small number of firms, with some evidence that they use a standardized template for new patents (Moser, Ohmsteadt, and Rhode 2018).

Taken together, in no field do all our knowledge flow proxies agree that agriculture is the main source of inputs. Rather, spill-ins from outside agriculture appear to matter, and to matter a great deal in most subsectors. We now turn to the nature of these non-agricultural spill-ins.

Whereas our paper does not try to rigorously define the "distance" between different knowledge domains, our results do provide some evidence that knowledge flows from outside of agriculture do not originate "too far" from agriculture. In figure 9, we present an attempt to measure whether knowledge flows originate "far" from agriculture, by resorting to some reasonable but perhaps *ad hoc* assumptions. We assume research originating in "non-ag" assignees (tables 6 and 9) is farther from agriculture than research originating in "minority ag" firms. This would be the case, for example, if an assignee's knowledge capital has some agricultural applications, as well as many others. In this case, the fact that the assignee also patents in agriculture is a signal that it has recognized the agricultural application of its knowledge capital. The animal health sector would seem to be a good example of this kind of dynamic. Much of the basic research on health for humans or animals is similar at the cellular level, even though the human health market is vastly larger than the veterinary health market (Clancy and Sneeringer 2019). That said, caution is warranted, because an assignee may also be a conglomerate with many parallel research operations that effectively embody separate knowledge capital stocks.

We feel it is also reasonable to assume biology and chemistry are scientific disciplines that are among the closest to agriculture, and so citations to biological and chemistry journals is an indicator that fields "close" to agriculture matter. Agriculture is typically classified as one of the life sciences (for example, by the NSF), and agricultural science has deep roots in chemistry (Huffman and Evenson 2006). Figure 9 uses these notions to provide some evidence that knowledge from outside agriculture is not "too far" away.

Figure 9. Share of Non-Agricultural Knowledge Flows Originating "Close" to Agriculture



Notes: Patent Cites 2 is the share of citations to non-specialized ag assignees that are classified as minority ag. Journal cites 2 is the share of non-agricultural journal citations classified as biology/biochemisty or chemical/chemical engineering. Text concepts 2 is the share of prior text mentions by non-specialized ag assignees that are classified as minority ag.

In this figure, we pull together very rough proxies for the distance from agriculture of non-agricultural knowledge flows.

- Patent Cites 2. Share of citations to assignees, but not specialized ag assignees, that are classified as minority ag (as opposed to non-agricultural).
- Journal Cites. Share of non-agricultural science journal citations to journals classified as biology/biochemistry or chemical/chemical engineering (as opposed to "other").
- Text concepts 2. Share of prior text mentions by assignees, but not specialized ag assignees, that are classified as minority ag (as opposed to non-agricultural).

In contrast to figure 9, most proxies now clear the 50% line. Where we can reasonably rank knowledge domains as being closer or farther from agriculture, non-agricultural knowledge flows in animal health, biocides, fertilizer, and plants are more likely to come from knowledge domains close to agriculture than from afar. For machinery and research tools, text concepts tend to be mentioned more often in non-agricultural assignees than minority ag ones. Machinery is also more likely to cite other journals than biology or chemistry ones, which is not surprising. Note, however, that the machinery sector cites by far the fewest journal publications.

Together, figure 8 and 9 suggest, while non-agricultural knowledge sources are very important, some non-agricultural knowledge domains are clearly more relevant than others. Whereas we view this conclusion as more tentative than our first one, it has relevance for science policy in agriculture.

4. Robustness Checks

In this section we conduct a wide array of robustness checks. To prevent the main paper from becoming too long, we report tables in the appendix, and merely summarize important details in the text.

4.1. Patent Citations

We investigate three potential sources of bias in our patent citation figures. First, that our results are driven by assignee's self-citation of their own patents. Second, that our results are robust to the exclusion of examiner-added citations. And third, that our results are robust when we restrict attention only to the most valuable patents (those receiving a high number of citations themselves).

There is a debate about the extent to which patent citations may be biased by a tendency for firms to cite their own work, or by the additional citations added by patent examiners (Lampe 2010, Moser, Ohmsteadt, and Rhode 2018). To assess whether our results are driven by self-citation, we first remove all citations from assignees to their own patents. Because so many individual inventors have a single patent, and because it is harder to accurately disambiguate inventor names, we restrict attention to assignee self-citation. The results are presented in Tables A2 and A3.

Excluding self-citations does not materially change the distribution of patent citations across different agricultural sectors, with one exception. In figure 3, the share of citations from plant patents to plant patents is 69%, but when we exclude self-citations, this falls to 56%. Similarly, in table 6, the share of citations to specialized ag firms is 81%, but when we exclude self-citations this falls to 69%. Moser, Ohmsteadt, and Rhode (2018), studying a sample of hybrid corn patents granted between 1985 and 2002 find that self-citations frequently reflect genuine cumulative innovation, as firms build on the prior genetic stock of their earlier patented plant cultivars. Therefore, it is not at all clear that the smaller share of 56% should be preferred to our baseline estimate of 69%.

Next, we remove all examiner-added citations. This is only possible for the period 2002 onward, when patents begin to identify who added a citation. There is some debate about whether examiner-added citations are good proxies for knowledge flows. If applicants seek to avoid citing relevant prior art for strategic reasons, examiner-added citations can correct this bias (Lampe 2010). Moreover, Chen (2017) finds examiner-added citations are more textually similar to the patent than other patents. That said, there is a large literature that highlights potential issues with examiner-added citations: for example, Moser, Ohmsteadt, and Rhode (2018) find examiners of hybrid corn patents are biased towards adding from their set of preferred patents, and that patents will tend to be added more for physical similarity of plants rather than genetic heritage. Jaffe and Rassenfosse (2018) summarize a number of other studies that describe potential distortions examiner-added citations may introduce. Tables A4 and A5 present the distribution of patent citations for patents granted after 2002, excluding examiner-added citations.

Removing examiner-added citations leaves our results largely unchanged, with one exception. In the machinery subsector, in figure 3 we found 48% of patents citations originated in the machinery subsector and 52% originated outside of agriculture. In Table A4, we instead find 56% of citations originate in the machinery subsector and 44% originate outside of agriculture. It turns out, however, that this has little to do with examiners and is instead driven by restricting patents to those granted after 2002. If we restrict attention to patents granted after 2002 (Table A6), 56% of patent citations in the machinery subsector originate in the same sector. Indeed, across all subsectors, there is a slight increase in patents originating from within the same subsector when we restrict attention to more recent patents.

Our final robustness check relates to the heterogenous value of patents. Many studies (see Nagaoka, Motohashi, and Goto 2010 for an overview) have shown that the value of patents is highly skewed. A small number of patents account for a disproportionately large share of value. Our results may be misleading if the minority of valuable patents differ in the sources of their knowledge, compared to patents as a whole. To check this, we identify the set of most valuable patents in agriculture, defined as those receiving 8 or more citations[9] in the 5 years following the date they are granted (this necessarily means we do not include patents from the last five years of our sample). Patents receiving 8 or more citations are in the top 5% for all agricultural patents. Tables A7 and A8 repeat our patent citation analysis for this subset of elite patents.

Restricting our attention to only the citations made by "elite" patents, we find a significantly higher share of citations originate from within the same subsector for the fertilizer, machinery, and research subsectors. Indeed, for machinery, the effect is large enough to tip the share of citations originating in the machinery subsector above 50%, from 48% in figure 3 to 64%, in table A7. In no other sector, however, does the share of citations from a given sector cross the 50% threshold, and so the conclusions drawn from our figures 8 and 9 remain valid. Turning to the share of citations received by different assignee types, restricting attention to only the most highly cited patents has the largest impact for the plant subsector, where the share of citations to specialized ag firms drops from 81% to 67%, and the share of citations to public sector patents rises from 13% to 25%.

4.2. Journal Citations

Flagship multidisciplinary journals such as *Science*, *Nature*, and *PNAS* present a challenge to our journal citation analysis. We classify these journals as "other", but citations to these journals could conceivably be to top articles in agricultural science, biology, or chemistry. In table A9, we break out citations to these three journals as a separate category. In the research tools subsector, these three journals account for 14.2% of all journal citations. However, even if the cited articles

[9] Citations received is a common proxy for the value of patents. See Nagaoka, Motohashi, and Goto (2010).

are all agricultural science articles, we still find about 50% of all journal citations would be to agricultural science. In the other subsectors, these three journals account for 1.3-3.9% of citations to academic work, suggesting that the main conclusions presented in section 3 are robust.

4.3. Text Concepts

We check the robustness of our text concept analysis to three alternative specifications. First, we impose a stricter criteria to our manual cleaning of concepts in agriculture. Second, we use an alternative weighting scheme that controls for the possibility that some of our concepts are duplicates that refer to the same underlying idea. Third, we use an alternative weighting scheme that puts more weight on clusters of concepts that are used in more future patents.

Tables A10 and A11 impose stricter criteria to our manual cleaning of text-novel concepts in agriculture. To manually clean concepts, three coauthors independently apply four exclusion rules (see section 1.2) to all concepts in our data. There is some subjectivity in these rules, for example, in judging what is an “uninformative” word and what “connective phrases” are. In the main specification, we retain a concept when at least two of the three judges retain it. In our robustness check, we require all three inspectors to agree for a concept to be retained. Depending on the subsector, this leads to us excluding an additional 10-20% of the original 200 concepts (the set of included and excluded concepts is available in tables A16-A21 of the appendix). Our core results, however, are not substantively changed by this stricter exclusion policy. No entries in tables 11 and 12 are changed.

Tables A12 and A13 summarize our text data in a different way. One possible concern with our text analysis approach is that we may be “double-counting” some concepts. This could occur, for example, if two concepts both refer to the same underlying idea. For example, suppose pyrimethamine is exclusively used to treat variants of the disease myeloencephalitis. Whenever the concept pyrimethamine appears in a patent, so too does the phrase myeloencephalitis, and vice versa, although perhaps not in the same sentence (or paragraph). Section 2.3 treats these two phrases as distinct concepts. There, we compute the share of prior mentions for each of these concepts, and then average over all these shares. But it could be argued the two concepts “pyrimethamine” and “myeloencephalitis” only really refer to one underlying idea (treating the disease with the antibiotic), since they are always and everywhere used together. If this is correct, then we are giving too much weight to the shares of prior patents mentioning these concepts by counting each concept separately.

Here, we consider an alternative approach that creates "families" of related concepts. For each concept, we look for its first appearance in a given agricultural subsector, which we call an originating patent. All concepts in the same originating patent constitute a family of related concepts.

For example, if pyrimethamine and myeloencephalitis are always used together, then they will both appear for the first time in animal health in the same patent and therefore will belong to the same family. For each of these families, we find the set of unique patents applied for before the originating patent with *any* concepts in the family. We compute the share of these patent originating in different knowledge domains. Denote the share of patents with concepts from family f that originate in knowledge domain i by $s_i(f)$.

We then average these shares over all families:

$$p_i = \frac{1}{n}\sum_{f=1}^{n} s_i(f) \tag{2}$$

This methodology uses originating patents to define families of related concepts, and give each family the same weight, ensuring we do not double-count concepts referring to the same concept. The trade-off with this approach is that a concept with no prior mentions may belong to a family of concepts that do have prior mentions. This methodology obscures the fact, because it treats families of concepts as units of observation.

This alternative methodology does have some significant impacts on our results, but none large enough to alter the conclusions in figures 8 and 9. Indeed, our major conclusion that ideas from outside of agriculture are important is actually strengthened. Under this alternative weighting scheme, the share of concepts originating in patents outside agriculture rises in every subsector, as does the share of concepts originating in the patents of non-agricultural assignees.

Lastly, we weight families of concepts by the number of agricultural patents that end up using any concepts in the family. Let $w(f)$ denote the number of patents in a subsector that use any concept in family f. Our final weighting scheme is:

$$p_i = \frac{\sum_{f=1}^{n} w(f) s_i(f)}{\sum_{f=1}^{n} w(f)} \tag{3}$$

Intuitively, this puts more weight on families of concepts that subsequently end up being used more heavily in the agricultural subsector. The results, presented in Tables A14 and A15 do not differ materially from Tables A12 and A13, although they again tend to increase the weight put on families of concepts originating outside of agriculture.

5. Conclusions

Agricultural total factor productivity grew enormously over the past century. In the years to come, continued increases in agricultural productivity will be essential for meeting the challenge of feeding a rising world population amid the challenges of climate change. There is widespread recognition that past R&D investments were crucial to develop the new and improved agricultural technologies that have mediated these celebrated productivity gains. This paper presents new evidence on the structure of knowledge underpinning agricultural R&D, with an emphasis on the role of knowledge spillovers across scientific and technological domains.

Using agricultural patents in animal health, biocides, fertilizer, machinery, plants, and research tools as measures of agricultural research outputs, we track knowledge flows into agriculture in five different ways. We start with citations to patents in agricultural subsectors, and across different types of inventive organizations and individuals. To capture knowledge flows from academia, we also track citations to journal articles across different journal categories. Finally, we complement these citation-based approaches with text analysis, where we identify text-concepts that are new (in text) and important in agriculture in the second half of our observation period. We then track the appearance of these text-concepts in earlier patents.

Our results indicate a major role for ideas that originate outside of agriculture, perhaps a role as important as R&D conducted within agriculture. In the animal health, fertilizer, and machinery subsectors across every measure we find the majority of knowledge flows originate in non-agricultural knowledge domains. In the remaining three subsectors, we find mixed evidence: some of our indicators suggest the majority of knowledge originates outside agriculture, and some from within. Amid these sets, the strongest case for knowledge originating primarily from within agriculture is the plant subsector, which primarily cites other agricultural patents and agricultural science journals. But even this subsector has the majority of its text concepts appearing outside of agriculture prior to their appearance in plant patents.

We also present some evidence that these "outside agriculture" knowledge domains remain predictably close to agriculture. Whereas agricultural science journals do not account for the majority of journal citations in most subsectors, together with biology and chemistry journals they do. Moreover, our other measures of knowledge flows indicate organizations with at least some agricultural patents do R&D more relevant to agriculture than organizations with no agricultural patents.

The novelty of this paper is to use information contained in patents, through patent citations and text analysis, to study agricultural knowledge flows, and this work suggests a number of possible avenues for future research. First, our text-concept approach can be easily extended to the corpus outside of patents. In particular, academic journals are a promising avenue to explore. For example, we find the biocide sector originates the majority of its text concepts, and that these concepts tend to be chemical names. At the same time, the sector heavily cites chemistry journals and it would be interesting to see if these chemical names appear first in chemistry journals. Outside of agriculture, Li et al. (2017) track knowledge flows from life science patents to basic research and find 30% of NIH grants result in publications that are subsequently cited by life science patents. Our text analysis approach could help identify cases where NIH funding results in ideas that are used in life science patents without a direct citation. More generally, this approach can be extended to books, company filings, and so on.

Second, the combination of text-novel concepts and citations represent a clear opportunity to track the diffusion of specific ideas through technology space. Are citations a channel through which text-concepts flow, and if so, can we track the movement of an idea originating in one technology field through a chain of linked citations to an eventual application in a distant technology field? This would allow one to examine the factors that most facilitate the transfer of ideas. Lastly, the analysis we have presented can be brought to bear on work linking agricultural

R&D to agricultural productivity measures. Patents may serve as new proxies for knowledge capital, proxies with more detailed information about the relevant R&D spending, both in agriculture and beyond.

Albeit preliminary, we may attempt to draw some normative implications of the results presented in this paper. The early work of Schultz (1956) and Griliches (1958) underscored agriculture's leading position in identifying the role of technical progress on productivity. A large and varied literature has since established the fundamental role that investments in science and technological R&D have on innovation and economic growth. The many market failures that beset the innovation process suggest a critical role for public policies to fund and support the R&D enterprise. Evidence of past remarkable successes have fostered the belief that scientific research is underfunded, and that a renewed investment impetus is needed to sustain growth. The argument is particularly pressing for U.S. agriculture, where public R&D investments have substantially declined, in real terms, over the last decade.[10] Meritorious calls for increased public agricultural R&D inevitably meet the reality of declining availability of public funds. In this age of scarcity, science policy needs to be mindful of the complexity and connectedness of the research enterprise. As highlighted in the model of Akcigit, Hanley, and Serrano-Velarde (2016), the spillover effects from basic research are critical. In our context, the knowledge spillovers we have identified suggest that agricultural science policy might best support agricultural productivity growth if it retains a holistic perspective. Attention to the broader research agenda, and in particular to areas that, while not being strictly agriculture oriented have traditionally been connected with agricultural innovation, is of paramount importance. Priorities that rely on narrowly defined measures of past returns to R&D may not provide the most productive use of scarce public R&D funds.

[10] https://www.ers.usda.gov/data-products/agricultural-research-funding-in-the-public-and-private-sectors/

References

Akcigit, U., Hanley, D. and Serrano-Velarde, N. "Back to Basics: Basic Research Spillovers, Innovation Policy and Growth." Working Paper, December 10 2016.

Alston, J.M., 2002. "Spillovers." *Australian Journal of Agricultural and Resource Economics*, *46*(3), pp. 315-346.

Arrow KJ. 1962. "Economic Welfare and the Allocation of Resources for Inventions." In: *The Rate and Direction of Inventive Activity: Economic and Social Factors*, ed. Nelson RR. Princeton, JN: Princeton University Press.

Balsmeier, B., Assaf, M., Chesebro, T., Fierro, G., Johnson, K., Johnson, S., Li, G.C., Lück, S., O'Reagan, D., Yeh, B. and Zang, G., 2018. "Machine learning and natural language processing on the patent corpus: Data, tools, and new measures." *Journal of Economics & Management Strategy*, 27(3), pp. 535-553.

Bloom, Nicholas, Mark Schankerman, and John Van Reenen. 2013. Identifying Technology Spillovers and Product Market Rivalry. *Econometrica* 81(4): 1347-1393.

Chen, Lixin. 2017. "Do patent citations indicate knowledge linkage? The evidence from text similarities between patents and their citations." *Journal of Informetrics* 11: 63-79.

Clancy, M.S., and G. Moschini. 2017. Intellectual Property Rights and the Ascent of Proprietary Innovation in Agriculture. *Annual Review of Resource Economics* 9:53-74.

Clancy, M., and Sneeringer, S., Forthcoming. "How Much Does it Cost Induce R&D in Animal Health?" *Applied Economic Perspectives and Policy.*

Evenson, R.E. 1989. "Spillover Benefits of Agricultural Research: Evidence from U.S. Experience." *American Journal of Agricultural Economics* 71(2): 447-452.

Fuglie, K.O. and Heisey, P.W., 2007. *Economic returns to public agricultural research*. Economic Brief No. 10, U.S. Department of Agriculture, Economics Research Service.

Gardner, BL. 2002. *American agriculture in the twentieth century: How it flourished and what it cost.* Cambridge, MA: Harvard University Press.

Greenstone, M., Horbeck., Moretti, E., 2010. "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings." *Journal of Political Economy* 118(3): 536-598.

Griliches, Z., 1957. "Hybrid corn: An exploration in the economics of technological change." *Econometrica*, 25(4), 501-522.

Griliches, Z. 1992. "The Search for R&D Spillovers." *Scandinavian Journal of Economics*, 94 (Supplement), 29-47.

Hall, B.H., J. Mairesse, and P. Mohnen. 2010. "Measuring the Returns to R&D." Chapter 24 in B.H. Hall and N. Rosenberg (eds.), *Handbook of the Economics of Innovation*, Volume 2. Amsterdam: North-Holland, pp. 1033-1082.

Huffman, W.E., and R.E. Evenson. 2006. *Science for Agriculture: A Long-Term Perspective* (2nd edition). Ames, IA: Blackwell Publishing.

Jaffe, Adam B., Manuel Trajtenberg, and Michael S. Fogarty. 2000. "The Meaning of Patent Citations: Report on the NBER/Case-Western Reserve Survey of Patentees." NBER Working Paper 7631.

Jaffe, A. B., and de Rassenfosse, G., 2017. "Patent Citation Data in Social Science Research: Overview and Best Practices." *Journal of the Association for Information Science and Technology* 68(6): 1460-1374.

Khanna, J., Huffman, W.E. and Sandler, T., 1994. Agricultural research expenditures in the United States: a public goods perspective. *Review of Economics and Statistics*, pp.267-277.

Lampe, Ryan. 2012. "Strategic Citation." *Review of Economic Studies* 94(1): 320-333.

Latimer, R., and D. Paarlberg. 1965. "Geographic Distribution of Research Costs and Benefits." *Journal of Farm Economics* 47(2): 234-241.

Li, Danielle, Pierre Azoulay, and Bhaven N. Sampat. 2017. "The applied value of public investments in biomedical research." *Science* 356: 6333, 78-81.

Mansfield, E., J. Rapoport, A. Romeo, S. Wagner, and G. Beardsley. 1977. "Social and Private Rates of Return from Industrial Innovations." *Quarterly Journal of Economics* 91(2): 221-240.

Marx, M. and Fuegi, A., 2019. Reliance on Science in Patenting. Working paper, Boston University, 20 February 2019. Available at SSRN.

Moser, Petra, Joerg Ohmstedt, Paul W. Rhode. 2018. "Patent Citations - An Analysis of Quality Differences and Citing Practices in Hybrid Corn." *Management Science* 64(4): 1926-1940.

Nagaoka, S., K. Motohashi, and A. Goto. 2010. Patent Statistics as an Innovation Indicator. In the *Handbook of the Economics of Innovation*, eds., B.H. Hall and N. Rosenberg. North Holland Publishing.

Packalen, M. and Bhattacharya, J., 2015. "New ideas in invention." NBER Working Paper No. w20922. National Bureau of Economic Research.

Roach, Michael, Wesley M. Cohen. 2013. "Lens or Prism? Patent Citations as a Measure of Knowledge Flows from Public Research." *Management Science* 59(2): 504-525.

Schultz, T.W., 1956. "Reflections on agricultural production, output and supply." *Journal of Farm Economics*, 38(3), pp.748-762.

United States Patent and Trademark Office, Patent Technology Monitoring Team. 2019. U.S. Colleges and Universities – Utility Patent Grants, Calendar Years 1969-2012. <https://www.uspto.gov/web/offices/ac/ido/oeip/taf/univ/org_gr/all_univ_ag.htm>.

Wang, C., Y. Xia, and S. Buccola. 2009. "Public Investment and Industry Incentives in Life-Science Research." *American Journal of Agricultural Economics* 91(2): 374-388.

Appendix Tables

Table A1. Share of Patents Derived from Domestic and Foreign Research

	Derived from	
	Domestic Research	Foreign Research
Animal Health	65.4%	34.6%
Biocides	51.2%	48.8%
Fertilizer	59.8%	40.2%
Machinery	64.5%	35.5%
Plants	85.9%	14.1%
Research Tools	59.7%	40.3%

Note: Patents are fractionally classified as derived from domestic or foreign research based on the share of inventors listing a US (domestic) or non-US (foreign) address.

Table A2. Share of Patent Citations to Agricultural Subsectors, excluding assignee self-citations

	Own Sector	Other Agriculture	Not Agriculture
Animal Health	5.8%	2.4%	91.8%
Biocides	23.1%	3.6%	73.3%
Fertilizer	26.3%	5.0%	68.7%
Machinery	46.3%	0.1%	53.6%
Plants	56.1%	41.8%	2.1%
Research tools	53.3%	3.5%	43.2%

Note: The rows indicate the citing agricultural subsector and columns the subsector to which cited patents belong. Only cited patents granted between 1976 and 2016 are included. Each citation is counted once, even if multiple citations point to the same patent. Own sector gives the share of these citations to patents in the same subsector. Other agriculture gives the share of these citations to any other agricultural subsector. Not agriculture gives the share of citations to patents not contained in any of our agricultural subsectors. We exclude citations made by assignees to their own patents.

Table A3. Share of Patent Citations to Assignee Types, excluding self-citations

	Ag Specialized	Ag Minority	Non Ag	Public Sector	Individuals
Animal Health	1.8%	63.1%	22.2%	4.8%	7.5%
Biocides	8.5%	62.5%	14.8%	4.6%	8.8%
Fertilizer	17.0%	32.0%	21.7%	4.4%	24.6%
Machinery	32.0%	27.7%	9.5%	1.1%	29.6%
Plants	69.2%	8.5%	0.5%	20.4%	1.0%
Research tools	26.1%	37.9%	14.1%	14.0%	6.4%

Note: The rows indicate the citing agricultural subsector and columns the assignee and inventor type to which the cited patents belong. Specialized ag assignees have more than 50% of their patents belonging to an agricultural subsector in the last 5 years. Minority ag assignees have more than zero patents belonging to an agricultural subsector in the last 5 years, but less than 50%. Non-ag assignees have no patents belonging to agricultural subsectors. Public sector assignees correspond to government and non-profit organizations. Individuals refers to patents owned by individual inventors. Rows do not add up to 100% - the remainder of patent citations (0.1-1.4%) are made to unclassified assignees (see section 1.3.1). We exclude citations made by assignees to their own patents.

Table A4. Share of Patent Citations to Agricultural Subsectors (2002 and later), excluding examiner-added citations

	Own Sector	Other Agriculture	Not Agriculture
Animal Health	6.9%	2.4%	90.7%
Biocides	24.4%	4.7%	70.8%
Fertilizer	29.3%	6.6%	64.1%
Machinery	56.4%	0.2%	43.5%
Plants	67.0%	31.8%	1.2%
Research tools	55.9%	3.3%	40.9%

Note: The rows indicate the citing agricultural subsector and columns the subsector to which cited patents belong. Only cited patents granted between 1976 and 2016 are included and only citing patents granted after 2002 are presented. Each citation is counted once, even if multiple citations point to the same patent. Own sector gives the share of these citations to patents in the same subsector. Other agriculture gives the share of these citations to any other agricultural subsector. Not agriculture gives the share of citations to patents not contained in any of our agricultural subsectors. We exclude citations made by patent examiners.

Table A5. Share of Patent Citations to Assignee Types (2002 and later), excluding examiner-added citations

	Ag Specialized	Ag Minority	Non Ag	Public Sector	Individuals
Animal Health	1.2%	64.0%	25.0%	4.7%	4.9%
Biocides	9.3%	62.6%	14.9%	4.4%	7.7%
Fertilizer	18.0%	31.3%	23.1%	5.4%	22.0%
Machinery	35.7%	28.5%	9.5%	1.3%	25.0%
Plants	79.4%	5.5%	0.3%	14.0%	0.6%
Research tools	28.0%	38.7%	13.5%	13.2%	5.2%

Note: The rows indicate the citing agricultural subsector and columns the assignee and inventor type to which the cited patents belong. Only cited patents granted between 1976 and 2016 are included, and only citing patents granted after 2002. Specialized ag assignees have more than 50% of their patents belonging to an agricultural subsector in the last 5 years. Minority ag assignees have more than zero patents belonging to an agricultural subsector in the last 5 years, but less than 50%. Non-ag assignees have no patents belonging to agricultural subsectors. Public sector assignees correspond to government and non-profit organizations. Individuals refers to patents owned by individual inventors. Rows do not add up to 100% - the remainder of patent citations (0.1-1.4%) are made to unclassified assignees (see section 1.3.1). We exclude citations made by patent examiners.

Table A6. Share of Patent Citations to Agricultural Subsectors (2002 and later)

	Own Sector	Other Agriculture	Not Agriculture
Animal Health	8.7%	2.5%	88.8%
Biocides	26.0%	4.5%	69.5%
Fertilizer	31.0%	6.2%	62.8%
Machinery	55.7%	0.1%	44.1%
Plants	69.9%	28.8%	1.2%
Research tools	57.0%	3.5%	39.5%

Note: The rows indicate the citing agricultural subsector and columns the subsector to which cited patents belong. Only cited patents granted between 1976 and 2016 are included and only citing patents granted after 2002 are presented. Each citation is counted once, even if multiple citations point to the same patent. Own sector gives the share of these citations to patents in the same subsector. Other agriculture gives the share of these citations to any other agricultural subsector. Not agriculture gives the share of citations to patents not contained in any of our agricultural subsectors.

Table A7. Share of Patent Citations from Highly Cited Patents to Agricultural Subsectors

	Own Sector	Other Agriculture	Not Agriculture
Animal Health	0.0%	0.0%	100.0%
Biocides	25.8%	7.0%	67.2%
Fertilizer	41.1%	1.9%	57.0%
Machinery	63.7%	0.1%	36.2%
Plants	61.3%	37.2%	1.5%
Research tools	68.1%	2.2%	29.7%

Note: The rows indicate the citing agricultural subsector and columns the subsector to which cited patents belong. Only cited patents granted between 1976 and 2016 are included and only citing patents that receive 8 or more citations in the five years after their grant dates. Each citation is counted once, even if multiple citations point to the same patent. Own sector gives the share of these citations to patents in the same subsector. Other agriculture gives the share of these citations to any other agricultural subsector. Not agriculture gives the share of citations to patents not contained in any of our agricultural subsectors.

Table A8. Share of Patent Citations from Highly Cited Patents to Assignee Types

	Ag Specialized	Ag Minority	Non Ag	Public Sector	Individuals
Animal Health	0.0%	88.9%	11.1%	0.0%	0.0%
Biocides	10.3%	72.7%	9.1%	1.9%	5.1%
Fertilizer	23.9%	30.2%	24.3%	2.4%	19.1%
Machinery	42.1%	27.4%	5.3%	1.1%	24.0%
Plants	67.5%	6.8%	0.2%	24.9%	0.5%
Research tools	30.7%	47.8%	7.8%	8.5%	4.0%

Note: The rows indicate the citing agricultural subsector and columns the assignee and inventor type to which the cited patents belong. Only cited patents granted between 1976 and 2016 are included, and only citing patents receiving 8 or more citations within the first five years after being granted. Specialized ag assignees have more than 50% of their patents belonging to an agricultural subsector in the last 5 years. Minority ag assignees have more than zero patents belonging to an agricultural subsector in the last 5 years, but less than 50%. Non-ag assignees have no patents belonging to agricultural subsectors. Public sector assignees correspond to government and non-profit organizations. Individuals refers to patents owned by individual inventors. Rows do not add up to 100% - the remainder of patent citations (up to 1.1%) are made to unclassified assignees (see section 1.3.1).

Table A9. Share of Journal Citations to Journal Categories, separating out Science, Nature, and PNAS

	Agricultural Sciences	Other Biology and Biochemistry	Chemistry/Chem Engineering	Science, Nature, PNAS	Other Scimago
Animal Health	17.6%	43.4%	7.2%	3.4%	28.3%
Biocides	32.3%	37.2%	11.6%	3.9%	15.1%
Fertilizer	40.1%	30.6%	14.3%	1.3%	13.7%
Machinery	41.6%	15.6%	7.7%	2.7%	32.5%
Plants	72.7%	22.8%	0.3%	3.8%	0.5%
Research Tools	34.0%	45.4%	0.8%	14.2%	5.6%

Note: Each entry is the share of identified journal citations originating in patents in the row subsector that go to journals in the column category.

Table A10. Share of Antecedent Text-novel Concept Mentions to Agricultural Subsectors, Strict Inclusion Criteria

	No Prior Mention	Other Agriculture	Not Agriculture
Animal Health	4.9%	2.0%	93.1%
Biocide	65.7%	4.3%	30.0%
Fertilizer	20.2%	4.2%	75.6%
Machine	32.9%	0.0%	67.1%
Plant	17.0%	28.8%	54.2%
Research tools	23.8%	5.4%	70.8%

Note: An entry gives the probability a randomly selected patent mentioning a randomly mentioned text-novel concept originates in a given sector. Antecedent mentions are all those made by patents applied for prior to the first patent in the given subsector that mentions the concept. This table includes a concept only if it is included by all three co-author inspectors.

Table A11. Share of Antecedent Text-novel Concept Mentions to Assignee-Type, Strict Inclusion Criterion

	Ag Specialized	Ag Minority	Non-Ag	Public	Individuals	No Prior
Animal Health	1.2%	46.0%	29.9%	7.4%	9.1%	4.9%
Biocide	3.7%	25.2%	3.7%	0.6%	0.1%	65.7%
Fertilizer	2.8%	30.4%	29.4%	4.9%	11.3%	20.2%
Machine	1.3%	16.8%	35.4%	0.9%	12.0%	32.9%
Plant	12.9%	31.1%	21.4%	10.7%	5.3%	17.0%
Research tools	1.9%	25.9%	27.4%	12.7%	6.8%	23.8%

Note: An entry gives the probability a randomly selected patent mentioning a randomly selected text-novel concept originates with a given assignee type. Antecedent mentions are all those made by patents applied for prior to the first patent in the given subsector that mentions the concept. Specialized ag assignees have more than 50% of their patents belonging to an agricultural subsector in the last 5 years. Minority ag assignees have more than zero patents belonging to an agricultural subsector in the last 5 years, but less than 50%. Non-ag assignees have no patents belonging to agricultural subsectors. Public sector assignees correspond to government and non-profit organizations. Individuals refers to patents owned by individual inventors. No prior indicates the concept has no prior mentions. Rows do not add up to 100% - the remainder of patent mentions (up to 0.8%) are made to unclassified assignees (see section 1.3.1). This table includes a concept only if it is included by all three co-author inspectors.

Table A12. Share of Antecedent Text-novel Concept Mentions to Agricultural Subsectors, Weighted by Concept Family

	No Prior Mention	Other Agriculture	Not Agriculture
Animal Health	2.5%	3.7%	93.8%
Biocide	56.8%	6.5%	36.7%
Fertilizer	4.2%	7.7%	88.1%
Machine	16.7%	0.1%	83.3%
Plant	9.4%	27.4%	63.2%
Research tools	17.6%	4.5%	77.9%

Note: An entry gives the probability a randomly selected patent mentioning a text-novel concept from a randomly selected family of concepts originates in a given sector. Antecedent mentions are all those made by patents applied for prior to the first patent in the given subsector that mentions the concept.

Table A13. Share of Antecedent Text-novel Concept Mentions to Assignee-Type, Weighted by Concept Family

	Ag Specialized	Ag Minority	Non-Ag	Public	Individuals	No Prior
Animal Health	0.9%	44.3%	34.2%	5.3%	10.3%	2.5%
Biocide	5.3%	29.6%	6.2%	1.1%	0.5%	56.8%
Fertilizer	1.8%	38.4%	36.3%	5.3%	12.8%	4.2%
Machine	3.3%	17.6%	46.5%	1.2%	14.2%	16.7%
Plant	9.6%	33.2%	28.5%	9.4%	7.5%	9.4%
Research tools	3.2%	24.0%	31.4%	13.1%	8.4%	17.6%

Note: An entry gives the probability a randomly selected patent mentioning a text-novel concept from a randomly selected concept family originates with a given assignee type. Antecedent mentions are all those made by patents applied for prior to the first patent in the given subsector that mentions the concept. Specialized ag assignees have more than 50% of their patents belonging to an agricultural subsector in the last 5 years. Minority ag assignees have more than zero patents belonging to an agricultural subsector in the last 5 years, but less than 50%. Non-ag assignees have no patents belonging to agricultural subsectors. Public sector assignees correspond to government and non-profit organizations. Individuals refers to patents owned by individual inventors. No prior indicates the concept has no prior mentions. Rows do not add up to 100% - the remainder of patent mentions (up to 1.6%) are made to unclassified assignees (see section 1.3.1).

Table A14. Share of Antecedent Text-novel Concept Mentions to Agricultural Subsectors, Weighted by Concept Family and Subsequent Patents

	No Prior Mention	Other Agriculture	Not Agriculture
Animal Health	1.6%	3.1%	95.3%
Biocide	52.3%	8.1%	39.6%
Fertilizer	3.8%	7.4%	88.9%
Machine	14.5%	0.0%	85.5%
Plant	4.9%	21.4%	73.7%
Research tools	14.4%	4.2%	81.4%

Note: An entry gives the probability a randomly selected patent mentioning a text-novel concept from a randomly selected family of concepts originates in a given sector, where the probability of selecting a concept family is weighted by the number of ag subsector patents using concepts belonging to the family. Antecedent mentions are all those made by patents applied for prior to the first patent in the given subsector that mentions the concept.

Table A15. Share of Antecedent Text-novel Concept Mentions to Assignee-Type, Weighted by Concept Family and Subsequent Patents

	Ag Specialized	Ag Minority	Non-Ag	Public	Individuals	No Prior
Animal Health	0.8%	44.8%	34.5%	5.1%	10.8%	1.6%
Biocide	5.9%	32.4%	6.8%	1.1%	0.6%	52.3%
Fertilizer	1.8%	38.9%	36.6%	5.0%	12.2%	3.8%
Machine	4.1%	17.2%	48.2%	1.3%	14.0%	14.5%
Plant	7.8%	35.4%	32.4%	9.0%	8.6%	4.9%
Research tools	3.1%	22.6%	33.0%	15.2%	8.4%	14.4%

Note: An entry gives the probability a randomly selected patent mentioning a text-novel concept from a randomly selected concept family originates with a given assignee type, where the probability of selecting a concept family is weighted by the number of ag subsector patents using concepts belonging to the family. Antecedent mentions are all those made by patents applied for prior to the first patent in the given subsector that mentions the concept. Specialized ag assignees have more than 50% of their patents belonging to an agricultural subsector in the last 5 years. Minority ag assignees have more than zero patents belonging to an agricultural subsector in the last 5 years, but less than 50%. Non-ag assignees have no patents belonging to agricultural subsectors. Public sector assignees correspond to government and non-profit organizations. Individuals refers to patents owned by individual inventors. No prior indicates the concept has no prior mentions. Rows do not add up to 100% - the remainder of patent mentions (up to 1.6%) are made to unclassified assignees (see section 1.3.1).

Table A16. Top 117 Text-novel Animal Health Concepts

UNANIMOUS (included in robustness check)		
protozoal	sarcocystis	physiologically active
protozoal myeloencephalitis	cyclooxygenase-2	volatile liquid
equine protozoal myeloencephalitis	milbemycin oxime	bovine respiratory
myeloencephalitis	releasing hormone	bovine respiratory disease
equine protozoal	gonadotropin releasing hormone	respiratory disease
trimethoprim	gonadotropin	swine respiratory
microbial	oral mucosa	pharmacologically active compound
microbial infection	equimolar	pharmacologically active
ear	propofol	diabetes
preservative	prodrug	weaning
terbinafine	cyclooxygenase	without heat
penetration enhancer	bacterial protozoa	heat detection
dermal penetration	bacterial protozoa infections	without heat detection
dermal	protozoa infections	sow
dermal penetration enhancer	kinases	c1-c6alkyl
kinase	polymorph	isoxazoline-substituted
janus kinase	transmucosal	insemination
janus	felis	buprenorphine
bird	ctenocephalides felis	spinosad
injection site	ctenocephalides	linoleic
single injection	hydrate	linoleic acid
hydrophilic surfactant	octyl	animal selected
stearoyl groups	octyl salicylate	catalytic
independently stearoyl groups	succinic	alkyl substituted
stearoyl	succinic acid	tobramycin
palmitoyl groups	xanthan gum	hydroxypropylcellulose
palmitoyl	xanthan	folic acid
asthma	equines	folic
fentanyl	furoate	ethanesulfonic
hydroxypropylmethylcellulose	mometasone furoate	ethanesulfonic acid
hydroxypropylmethylcellulose dissolved	mometasone	methanesulfonic acid
pyrimethamine	gnrh	methanesulfonic
epm	buccal	hydroxypropyl cellulose
prophylactic	cox-2	phenol
CONSENSUS (excluded in robustness check)		
mediated	containing hydroxypropylmethylcellulose	thickener
independently stearoyl	gonadotropin releasing	breeding
pharmaceutically active agent	synchronizing	daily dosage
veterinary applications	transmucosal administration	sweetener
controlled-	gum	sweeteners

Table A17. Top 177 Text-Novel Biocide Concepts (1/2)

UNANIMOUS (included in robustness check)		
thiamethoxam	boscalid	spirodiclofen
azoxystrobin	ethiprole	asulam
clothianidin	methoxyfenozide	noviflumuron
trifloxystrobin	cinosulfuron	thifluzamide
spinosad	penoxsulam	strobilurin
acetamiprid	flonicamid	halofenozide
thiacloprid	triflumuron	oxasulfuron
prothioconazole	neonicotinoid	quinoxyfen
pyraclostrobin	benoxacor	diofenolan
emamectin	isoxaflutole	ethaboxam
emamectin benzoate	tebufenpyrad	trifloxysulfuron
fluquinconazole	sulfosulfuron	gamma-cyhalothrin
dinotefuran	novel active compound	cyazofamid
lufenuron	metaflumizone	dioxygenase
imazamox	dimoxystrobin	fenpyroximate
controlling animal pests	isoxadifen-ethyl	milbemectin
nitenpyram	spiromesifen	cloquintocet
kresoxim-methyl	metosulam	zeta-cypermethrin
mesotrione	pyridaben	bromobutide
ipconazole	teflubenzuron	halosulfuron-methyl
fluoxastrobin	florasulam	thifensulfuron-methyl
sulfentrazone	chlorfluazuron	c1-c4-alkoxy
hexaflumuron	cyclosulfamuron	mefenoxam
chlorfenapyr	imazapic	chlorantraniliprole
cloquintocet-mexyl	protoporphyrinogen	pyroquilon
flumioxazin	protoporphyrinogen oxidase	fluxofenim
tebufenozide	fenclorim	fenhexamid
indoxacarb	orysastrobin	tritosulfuron
famoxadone	penthiopyrad	oxabetrinil
c1-c4-alkyl	spirotetramat	mepronil
mefenpyr-diethyl	flutolanil	tricyclazole
picoxystrobin	isoxaben	thenylchlor
novaluron	carfentrazone-ethyl	acibenzolar-s-methyl
pymetrozine	propaquizafop	aminopyralid
flumetsulam	foramsulfuron	flubendiamide
spinetoram	simeconazole	flufenacet
boxh	pyridalyl	etoxazole
ethoxysulfuron	fenamidone	metominostrobin
diafenthiuron	pyrifenox	isoprothiolane
spiroxamine	tau-fluvalinate	iodosulfuron

Table A17: Top 177 Text-Novel Biocide Concepts (2/2)

triazamate	iprovalicarb	moxidectin
daimuron	fosamine	fosthiazate
iminoctadine	oxadiargyl	diflufenzopyr
fenoxaprop-p	furametpyr	c1-c4-haloalkyl
carfentrazone	doramectin	macrocyclic
phytopathogenic harmful fungi	flufenerim	cyometrinil
phytopathogenic harmful	probenazole	nithiazine
fluopyram	trinexapac-ethyl	bixafen
pyridin-3-yl	diclosulam	isotianil
chromafenozide	bifenazate	saflufenacil
cyhalofop-butyl	mandipropamid	fluopicolide
pyributicarb	cyprosulfamide	flupyrsulfuron
kinoprene	cyflufenamid	metalaxyl-m
triazoxide	mepanipyrim	pyriprole
nanoparticles	metrafenone	benthiavalicarb
clodinafop	proquinazid	cyantraniliprole
		tolfenpyrad
CONSENSUS (excluded in robustness check)		
extenders and/or surfactants	preventively	whereinr1
ch3	r14	no2
plant essential	fully halogenated	

Table A18: Top 213 Text-Novel Fertilizer Concepts (1/2)

UNANIMOUS (included in robustness check)		
selenium	inorganic substrate component	biomass particles
itaconic	inorganic substrate	organic drying
itaconic moieties	cell component	compound drying
itaconic acid	corn steep	organic compound drying
itaconic anhydride	corn steep liquor	compound drying agent
compost tea	fertigation	biotic
canola	bactericide	co2
canola oil	maleic moieties	vermicompost
particle domain	bioorganic	recurring polymeric
mean particle domain	inorganically-augmented bioorganic fertilizer	polymeric subunits
water-dispersible particle	bioorganic fertilizer	recurring polymeric subunits
particle dispersion	inorganically-augmented bioorganic	sulfate nitrate
polymer-containing composition	inorganically-augmented	ammonium sulfate nitrate
soil amendment compositions	animal manures	wood ash
chlorine dioxide	hydrolyzed animal	plant nutrient content
wetting agents	animal hair	mycorrhizal fungi
phosphite	hydrolyzed animal hair	seed meal
ferrate	urea-formaldehyde polymer	soy meal
sodium ferrate	vinylic polymer	triple super phosphate
calcium ferrate	vinylic	dried residue
potassium ferrate	vinylic polymers	industrial molasses
decompose potassium	polycarboxylated polymer	pharmaceutical fermentation
potassium minerals	polycarboxylated	threonine
decompose potassium minerals	municipal biosolids	ellipsoideus
decompose potassium compounds	biochar	delbrueckii
partial salt	meat meal	saccharomyces delbrueckii
copolymer salt	cerevisiae	green waste
block copolymer	saccharomyces cerevisiae	toxins
yeast cell	saccharomyces cerevisiae hansen	heat source
yeast cells	hansen	abiotic
carbon-skeleton energy	cerevisiae hansen	dissolved materials
carbon-skeleton	calcium hypochlorite	phosphorus minerals
carbon skeleton energy	adenosine	decompose phosphorus minerals
skeleton energy	adenosine triphosphate	decompose phosphorus
complex carbon	triphosphate	biostimulant
carbon compounds	atp	radical polymerization
complex carbon compounds	nh4	free radical polymerization
binder component	ester groups	swine manure
water-soluble binder	environmentally friendly	bio
substrate component	biomass feedstock	dissolved oxygen

Table A18: Top 213 Text-Novel Fertilizer Concepts (2/2)

metal silicate	saccharomyces uvarum beijer	saccharomyces ludwigii
electrical conductivity	uvarum	saccharomyces willianus
heat-dried biosolids	uvarum beijer	willianus
heat-dried	saccharomyces uvarum	saccharomyces rosei
lower alcohol	beijer	rosei
pva	mellis	rouxii
bactericidal	saccharomyces mellis	saccharomyces rouxii
neodymium	saccharomyces microellipsoides	saccharomyces sake
bifenthrin	microellipsoides	sake
c1-c4 alcohols	oviformis	exiguus
electromagnetic field	saccharomyces oviformis	saccharomyces exiguus
decompose phosphorous	saccharomyces fermentati	carlsbergensis
decompose phosphorous compounds	fermentati	saccharomyces carlsbergensis
aluminum phosphate	saccharomyces logos	chevalieri
organic alcohols	logos	saccharomyces chevalieri
sylvinite	ludwigii	saccharomyces sp.
CONSENSUS (excluded in robustness check)		
tea	energy component	bimodal vinylic
mean particle	skeleton energy component	bimodal vinylic polymer
particle domain size	decompose complex carbon	ch2
polymer-containing	decompose complex	overproduce
biological fertilizer composition	convert complex	overproduce growth
domain size ranges	convert complex carbon	paste-like
mean particle size	binder component present	paste-like material
amendment compositions	steep liquor	dust control
enhancing soil	hemp	drying agent selected
soil fertility	fertilizer marketplace	quick drying
salt form	agricultural fertilizer marketplace	drying properties
partial salt form	agricultural crop	organic drying agent
form granulated particles	commercial agricultural fertilizer	quick drying properties
form granulated	polymer made	integrated system
yeast cell component	polymer composition also	mgso

Table A19: Top 106 Text-Novel Machine Concepts

UNANIMOUS (included in robustness check)		
aeration apparatus	operating travel	controller communicatively
axle driving apparatus	operating travel direction	controller communicatively coupled
antenna	wheel configured	air cart
robotic arm	operative control	perimeter wall
flexible cutterbar assembly	rolling basket	upright axes spaced
modular disc cutterbar	teatcup liner	residue spreader
modular disc	foot platform	crop residue spreader
cutterbar assembly flexes	position based	rotary milking
fore-and-aft draper	receiving data	pin configured
flexible draper	meter roller	aeration tine
trimmer head assembly	robotic attacher	aeration pockets
non-transitory computer	axle driving unit	energy storage device
computer readable	agricultural row unit	wireless communication
computer readable medium	forward working direction	crop throughput
non-transitory computer readable	zero radius turning	residue chopper
computer program product	approximate zero	aeration tines
product tank	grain cart	tool coupled
gps receiver	ecu	imaging device
seed metering system	pump mounting surface	inductor box
location-determining receiver	rotary cutting deck	output device
location-determining	computer-readable	distribution lines
gnss receiver	location information	horizontal cutter disks
gnss	motor mounting surface	generally horizontal cutter
		positions spaced transversely
CONSENSUS (excluded in robustness check)		
controller configured	system based	module configured
actuator configured	controllably operable	plate configured
apparatus configured	control unit configured	belt configured
dairy livestock	vehicle position	sensor arrangement
arm configured	agricultural working machine	processor configured
cutterbar assembly attached	valve configured	manner selected
assembly flexes	motor configured	conveyor configured
sickle assembly supported	characteristic data	units configured
controller operable	controller receiving	adjustment mechanism configured
opening configured	chamber configured	controller controlling
harvesting header operable	headland	crop inputs
readable medium	executable	evaluate

Table A20: Top 118 Text-Novel Plant Concepts

UNANIMOUS (included in robustness check)		
transgene	modified carbohydrate metabolism	genetic material
transgene encoding	acid metabolism	glyphosate
transgene encodes	phenoxy proprionic acid	glufosinate
locus conversion	phenoxy	sulfonylurea
single locus	phenoxy proprionic	transgenic
single locus conversion	proprionic	benzonitrile
backcross conversion	proprionic acid	triazine
backcross	nucleic acid	backcrossing
backcross progeny	nucleic	tissue cultures
progeny plants	altered fatty	bacillus
selected progeny	altered	bacillus thuringiensis
trait selected	altered phosphorus	thuringiensis
selected progeny plants	phosphorus	bacillus thuringiensis endotoxin
herbicide selected	altered carbohydrates	endotoxin
selected backcross	carbohydrates	thuringiensis endotoxin
produce selected	altered fatty acids	pest resistance
selected backcross progeny	fatty acids	dicamba
backcross progeny plants	altered antioxidants	herbicide resistant
higher backcross	antioxidants	imidazolinone
higher backcross progeny	altered essential amino	transgenes
transformation	amino	insect resistant
f1 progeny	amino acids	fungal
insect resistance	essential amino	waxy starch
plant derived	essential amino acids	pistil
soybean hulls	modified protein	root tip
modified fatty acid	protein concentrate	bacterial
modified fatty	protein isolate	viral disease
fatty acid metabolism	herbicide tolerance	hypocotyl
metabolism	abiotic stress	introgressed
carbohydrate metabolism	abiotic	traits introgressed
carbohydrate	abiotic stress tolerance	
modified carbohydrate	herbicide resistance	
CONSENSUS (excluded in robustness check)		
transgene confers	produce backcross progeny	isolate
encoding	locus confers	subsequent generation
transgene conferring	single locus confers	environmental conditions
conversion	plant product	site-specific recombination
locus	commodity plant	recombination
locus conversion confers	commodity plant product	waxy
desired trait	hulls	tip
produce backcross	concentrate	corn variety

Table A21: Top 122 Text-Novel Research Input Concepts

UNANIMOUS (included in robustness checks)		
clustal alignment method	hairpin rna	abiotic stress tolerance
clustal method	amplicon	seed-preferred
novel nucleotide	elongase	hemp
single nucleotide	antibody compositions	digestibility
sequence identity based	nitrogen use efficiency	pufas
identity based	increased biomass	biofuel
gene silencing	increased seed yield	colloid
transcribable polynucleotide	increased oil	agr
transcribable polynucleotide molecule	switchgrass	schizochytrium
isolated polynucleotides	agrobacterium -mediated	transcription factors
chimeric gene results	#NAME?	lyophilization
pesticidal polypeptide	agrobacterium -mediated transformation	poaceae
polyunsaturated fatty acids	olive	sirna
oilseed plant	isolated polypeptides	salinity
plant biomass	diacylglycerol	epa
nucleic acid segments	diacylglycerol acyltransferase	dalapon
eicosapentaenoic acid	mirna	fescue
eicosapentaenoic	salix	thraustochytrium
docosahexaenoic acid	salix species	pathogen-inducible promoter
docosahexaenoic	crucifers	dehalogenase
acid metabolism	heterologous nucleotide sequences	hppd
acid segments	molecular markers	castor bean
fatty acid metabolism	stress-related protein	coconut palm
wild type variety	carbohydrate metabolism	snp
rnai	fluorescent protein	silage
turfgrass	green fluorescent	starch branching
double-stranded rna	green fluorescent protein	frt
rna interference	vicia species	cosmetics
CONSENSUS (excluded in robustness checks)		
clustal	polynucleotide operably	nitrogen use
clustal v method	isolated polynucleotides encoding	polypeptides encoded
clustal v	coding nucleic	food product
alignment method	coding nucleic acid	primer pair
clustal alignment	acid molecules encoding	stress-related
pairwise alignment	acid molecule operably	gene involved
one regulatory sequence	type variety	full complement
provides recombinant expression	corresponding wild	element operably linked
silencing	full-length complement	agronomic interest
polynucleotide selected	representative seed	recombination sites
isolated polynucleotide selected	encodes seq	orientation relative
one polynucleotide	encodes seq id	increasing resistance
polynucleotide operably linked	use efficiency	