

# The mathematical structure of innovation

Thomas M. A. Fink\* and Ali Teimouri†

London Institute for Mathematical Sciences, 35a South St, Mayfair, London W1K 2XF, UK

Despite our familiarity with specific technologies, the origin of new technologies remains mysterious. Are new technologies made from scratch, or are they built up recursively from new combinations of existing technologies? To answer this, we introduce a simple model of recursive innovation in which technologies are made up of components and combinations of components can be turned into new components—a process we call technological recursion. We derive a formula for the extent to which technological recursion increases or decreases the likelihood of making new technologies. We test our predictions on historical data from three domains and find that technologies are not built up from scratch, but are the result of new combinations of existing technologies. This suggests a dynamical process by which known technologies were made and a strategy for accelerating the discovery of new ones.

Where do new technologies come from? Despite our familiarity with specific technologies, such as wind turbines, touch screens and GPS, how new technologies come into being remains mysterious [1]. By definition, new technologies lie just beyond our imagination [2], and navigating an expanding space of possibilities is challenging [3, 4].

However, we have a better understanding of what technologies are. Technologies are made up of component building blocks [5–9, 23]: “a combination of components to some purpose” [10]. For example, GPS is made up of the components of atomic clocks, satellites and receivers, and wind turbines are made up of rotors, generators and towers.

When a technology—a combination of components—becomes a reliable functioning unit, it becomes more readily available, and obtaining the whole is easier than gathering the individual parts. It is given a name and used as a component in its own right. So technologies are combinations of components, and these technologies can themselves be adopted as components. In this way technologies are made up of more primitive technologies, in a hierarchical way, as shown in Fig. 1 for GPS. New technologies are not built up from scratch, but are the result of new combinations of existing technologies—a hypothesis described qualitatively by Brian Arthur in *The Nature of Technology* [11].

In the mind of the innovator, too, the combination is no longer a complex assembly, but a distinct unit, ready for use [14]. This alters the innovator’s perception of the adjacent possible: it biases it in favour of designs which use the combination, since now they are one step away, rather than the multiple steps needed for the constituent parts.

## Technological recursion

Turning a combination of components into a new component is a recursive process, and we call it technological recursion. How does it effect a technology sector? On the one hand, technological recursion makes a sector simpler. Technologies which were previously composed of many components become composed of fewer components. This makes them easier to make because, given access to a fixed number of components, we are more likely to have the components needed to make technologies with few components than those with many [7].

On the other hand, technological recursion makes a technology sector more complicated. There is another component to keep track of, and the number of combinations of components grows exponentially with the number of components. In this expanded search space, there are more candidate combinations to sift through when searching for new technologies.

Too much technological recursion and too little both inhibit

discovering new technologies. With too much, there is a surge of new components, most of which are nearly useless. This is similar to how, in language, if every sentence fragment becomes a new word, the numbers of words escalates. With too little recursion, technologies become ever more complex, with no apparent structure across different organizational length scales. For example, without technological recursion, GPS is made from at least 13 components rather than three (see Fig. 1). This is similar to how, in language, if no new words are introduced, the number of words needed to describe innovations multiplies; without the new word “GPS”, for instance, we would be stuck with “global positioning system”.

## In this Letter

The interplay between existing technologies and the components for making new ones is at the heart of understanding where new technologies come from. In this Letter we do four things. First, we introduce a simple model of recursive innovation in which technologies are made up of components, and combinations of components can be turned into new components—a process we call technological recursion. Second, we derive a formula for the expected benefit of technological recursion: the extent to which it increases or decreases the likelihood of making new technologies. Third, we apply our insights to historical data from language, gastronomy and drugs, and find that only a small fraction of technological recursions increase the chance of making new technologies. Remarkably, these are almost always complete technologies, confirming the hypothesis that technologies are built up recursively. Fourth, we show that repeated technological recursion can increase the likelihood of making new technologies by orders of magnitude. This suggests a dynamical process by which known technologies were made, and a strategy for accelerating the discovery of new ones [26].

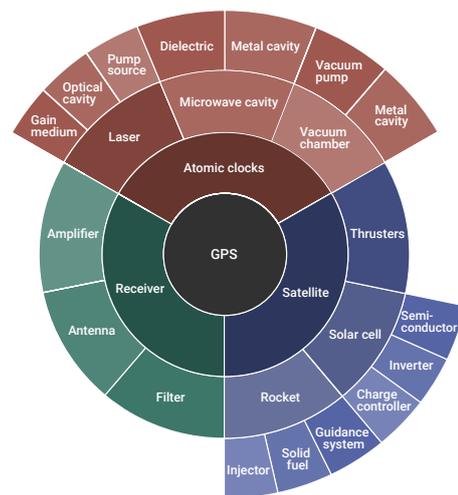


Figure 1: **The recursion structure of technology.** Technologies are made up of more primitive technologies in a hierarchical way. For example, GPS is made up of the components of atomic clocks, satellites and receivers. Atomic clocks, in turn, are made up of lasers, microwave cavities and vacuum chambers; and so on. We call this process of turning a combination of components into a new component technological recursion. Although we do not show it here, GPS is itself a component in subsequent technologies, such as guided missiles and earthquake monitors.

## Results

### Lego game

To illustrate our insights, consider three children, Eve, Frank and Grace, playing with three different Lego sets. The goal of each child is to make as many toys as possible. Eve has an ordinary set of Lego bricks. Frank has the same set, but he modified it by permanently gluing together pieces that frequently go together, such as wheels and axles, so that he has quick access to these modules when needed. Grace adopted Frank’s strategy, but took it further. Not only did she glue together pieces frequently used in the same toy, but she also glued pieces that are used in the same toy just occasionally, such as baseplates and doors. Each child scoops out a random bunch of pieces from their box of Lego bricks. All three have the same number, but whereas Eve has only individual bricks, some of Frank’s pieces are whole modules that he pre-assembled. Grace has more modules and fewer individual bricks.

Which child makes the most toys? While Eve assembles individual bricks, Frank has the advantage that when the toy he is making contains one of his pre-assembled modules, he’s sure to have all the pieces contained in it. But while Frank comes in first place, Grace comes in last. Despite having more modules than Frank, her approach backfires. Many of her modules belong to toys that she can’t complete. Because she has fewer individual bricks than the others, she’s often missing the right bricks to finish off toys she can partially make. As we shall see, turning the right combinations of components into new components can dramatically increase the number of makeable technologies, but turning too many combinations into components has the opposite effect.

### Components and technologies

We take technologies to be made up of distinct components [5–11]. A component can be a material object, like a capacitor, or a routine, like polymerase chain reaction, or a skill, like coding in Java. Once a component has been discovered, we do not have to worry about running out; there are no capacity constraints. Any subset of our components can be combined, but a combination either is, or is not, a technology, according to some universal recipe book of technologies. Suppose further that there are a total of  $N$  possible components in “God’s own cupboard” but that, at any given stage  $n$ , we only have in our basket  $n$  of these  $N$  possible components. At every stage, we pick a new component to add to our basket, increasing  $n$  by 1.

The size  $c$  of a technology is the number of distinct components a technology is made of. The order of components is irrelevant, and multiple occurrences of a component count once, so the word “innovation” has  $c = 6$  components, not 10.

### Size of the technology space

Consider a specific basket of  $n$  components, which we call  $\mathbf{n}$ . Notice how we differentiate between a particular set of components  $\mathbf{n}$  and their number  $n$ . Let the size of the technology space  $p(\mathbf{n})$  be the number of technologies that we can make from  $\mathbf{n}$ . Of course,  $p(\mathbf{n})$  depends on the specific components in our basket, and not just on how many we have; for example, we can make more words from the first five letters of the alphabet than the last five. To bypass this, we take the average over all possible baskets of size  $n$  drawn from the  $N$  possible components. We call this the expected size of the technology space,  $\bar{p}(n)$ . We denote the number of makeable technologies of size  $c$  by  $p(\mathbf{n}, c)$ , so that summing  $p(\mathbf{n}, c)$  over  $c$  gives  $p(\mathbf{n})$ .

In previous work [7, 22], we proved that  $\bar{p}(n, c)$ —the average of  $p(\mathbf{n}, c)$  over all possible baskets  $\mathbf{n}$ —satisfies an exact

conservation law: for two different stages  $n$  and  $n'$ ,

$$\bar{p}(n, c) / \binom{n}{c} = \bar{p}(n', c) / \binom{n'}{c}, \quad (1)$$

where  $\binom{n}{c}$  is the binomial coefficient. We will use this later when we determine the effect of technological recursion.

### Combination usefulness

Some combinations of components show up in more technologies than others. Consider some combination of  $k$  components which appear together in one or more technologies; we call the  $k$  components a module and label it  $\theta$  for convenience. The combination usefulness  $u_\theta(\mathbf{n})$  is the number of technologies makeable from the components  $\mathbf{n}$  that the combination  $\theta$  shows up in. As with the size of the technology space  $p$ ,  $u_\theta(\mathbf{n}, c)$  denotes the number of technologies of size  $c$  that  $\theta$  shows up in.

We proved (see Methods) that  $\bar{u}_\theta(n, c)$ —the average of  $u_\theta(\mathbf{n}, c)$  over all possible baskets  $\mathbf{n}$ —satisfies an exact

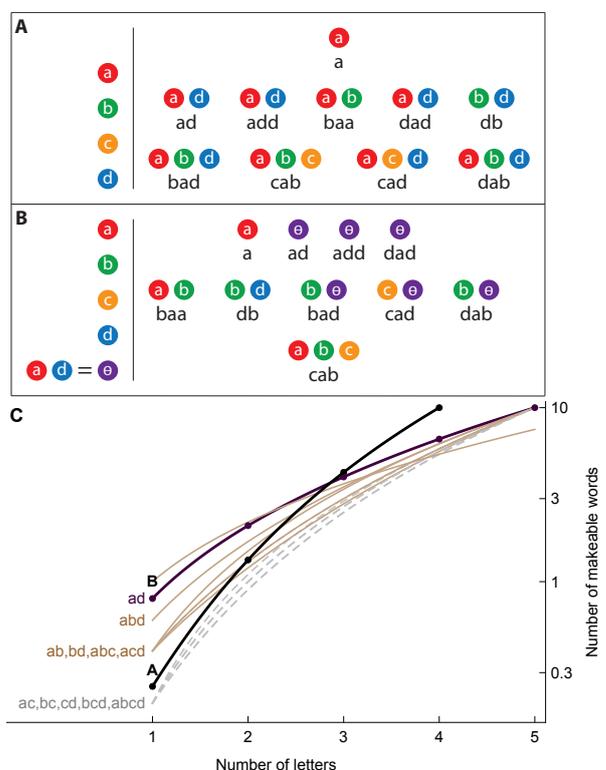


Figure 2: **Turning a technology into a component can increase the number of makeable technologies.** **A** In a simple example from language, the components are the letters a, b, c and d and the technologies are the 10 English words that can be made from them. The order and repetition of letters in the words are irrelevant. One word has 1 different kind of component, five have 2 kinds and four have 3 kinds. **B** When we expand our repertoire of components by turning one of the words, ad, into a component—call it  $\theta$ —some words are made up of fewer components than before. Now four words have 1 different kind of component, five have 2 kinds and one has 3 kinds. **C** How many words can we make, on average, when we pick  $n$  random components? When we draw from the original repertoire of four components in **A**, the result is given by the black curve. When we draw from the new repertoire of five components in **B**, it is given by the purple curve. Turning ad into a component increases the expected number of makeable words 3.2-fold when we draw one component and 1.6-fold when we draw two, but when we draw three or more components it decreases. The solid tan curves show the result for turning other words into components, all of which offer some advantage when the size of the draw is small. The dashed gray curves show the result for all other combinations, none of which offer any advantage.

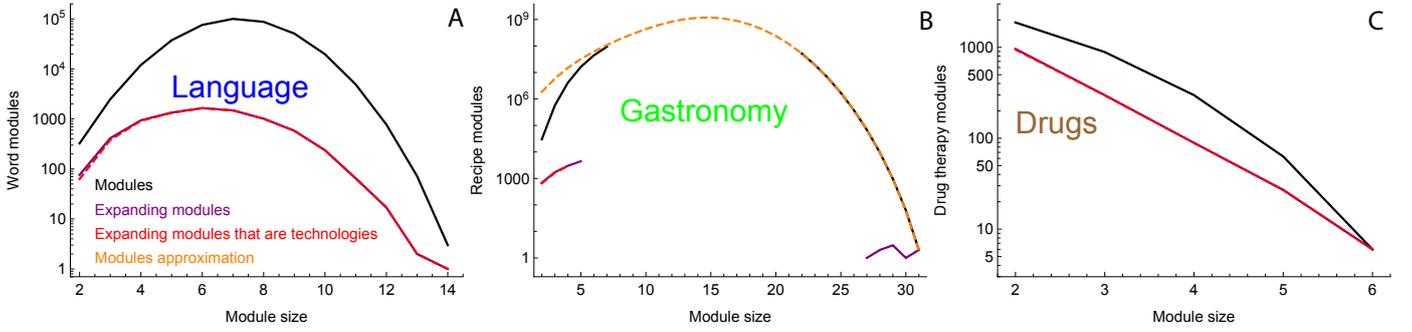


Figure 3: **Expanding modules are rare but almost always complete technologies.** ABC In our three domains, the top curves are the number of distinct combinations of  $c$  components: pairs, triples, quadruples, and so on. The purple solid curve is the number expanding modules: combinations that increase the likelihood of making new technologies. The red dashed curves are the number of expanding modules that are complete technologies, rather than parts of technologies. Remarkably, almost all expanding modules are complete technologies.

conservation law: for two different stages  $n$  and  $n'$ ,

$$\bar{u}_{\theta}(n', c) \binom{n'}{k} / \binom{n'}{c} = \bar{u}_{\theta}(n, c) \binom{n}{k} / \binom{n}{c}. \quad (2)$$

With these two conservation laws in hand, we are ready to determine the effect of technological recursion.

### Effect of technological recursion

The likelihood that a random combination of  $c$  components is a technology is  $\bar{p}(n, c) / \binom{n}{c}$ . Therefore increasing the likelihood of making a technology corresponds to increasing the expected number of technologies  $\bar{p}(n, c)$  that we can make, which we now address.

In a technological recursion, we replace a combination of  $k$  components  $\theta$  with a new individual component  $\theta$  in all of the technologies in which the combination  $\theta$  shows up. We indicate this as  $\theta \rightarrow \theta$ . Notice that we differentiate between the combination of components  $\theta$  and the new individual component  $\theta$  that replaces them. An example is shown in Fig. 2ab.

When we apply the technological recursion  $\theta \rightarrow \theta$ , some technologies of size  $c + k - 1$  are reduced to size  $c$ , and some technologies of size  $c$  are reduced to size  $c - k + 1$ . So we both add and subtract to the original number of makeable technologies of size  $c$ . The effect of a technological recursion depends on whether this net change, summed over  $c$ , is positive or negative. We obtained a formula, derived in the Methods, that gives the expected number of technologies we can make after applying  $\theta \rightarrow \theta$ :

$$\bar{p}_{\theta}(n) \simeq \left(1 - \frac{x}{N+1} \frac{d}{dx}\right) (\bar{p}(n) + (x - x^k) \bar{u}_{\theta}(n)), \quad (3)$$

where  $x = n/N$ . Because  $p(n)$  and  $u_{\theta}(n)$  are unbiased estimates of their means, we can estimate  $\bar{p}(n)$  and  $\bar{u}_{\theta}(n)$  by  $p(n)$  and  $u_{\theta}(n)$ . We do not assume knowledge of the technology recipe book for components that we do not possess; there is no omniscience.

### Expanding modules

When  $\bar{p}_{\theta}(n) > \bar{p}(n)$ , we can make more technologies from  $n$  components drawn from the updated repertoire of  $N + 1$  components than we can from  $n$  components drawn from the original repertoire of  $N$  components. We call a combination  $\theta$  that increases the number of makeable technologies an expanding module. From (3),  $\bar{p}_{\theta}(n) > \bar{p}(n)$  implies

$$\bar{v}_{\theta}(n) > \frac{x}{N+1} (\bar{p}'(n) + \bar{v}'_{\theta}(n)), \quad (4)$$

where  $x = n/N$ ,  $\bar{v}_{\theta}(n) = (x - x^k) \bar{u}_{\theta}(n)$  and  $'$  indicates the derivative with respect to  $x$ .

### Language, gastronomy and drugs

With a test for expanding modules in hand, we now apply our theory to real systems. We gathered data from three domains: language, gastronomy and drugs. In language, the technologies are the 38,970 English words and the components are the 26 letters used to make them. In gastronomy, the technologies are 56,498 recipes and the components are the 381 ingredients used to make them [24]. In drugs, the technologies are the 1363 known drug combination therapies and the components are the 901 individual drugs used to make them [25].

We did the following experiment for each domain. We randomly selected  $n$  components to put into our basket and counted how many technologies we could make from them. Using the conservation law in eq. (1), we were able to average this over all possible baskets of size  $n$ . This gives  $\bar{p}(n)$ .

We then picked a combination of  $k$  components  $\theta$ , and replaced it with the new individual component  $\theta$  everywhere it showed up in a technology. This raises the total number of possible components from  $N$  to  $N + 1$ . We randomly selected  $n$  components drawn from this new repertoire to put into our basket and counted how many technologies we could make. As before, we averaged this over all possible baskets using eq. (3). This gives  $\bar{p}_{\theta}(n)$ .

### Expanding combinations are rare

We repeated the above experiment for all possible combinations of  $k$  components  $\theta$  that show up in at least one technology: pairs, triples, quadruples and so on. (We did this for all  $k$  in language and drugs but for  $k \leq 5$  for gastronomy for computational reasons.) We found that only a small fraction of these combinations increase the number of makeable technologies. For  $k \leq 5$  across all domains, we found that in language, there are 128,413 letter combinations, of which 4,426 are expanding; in gastronomy, there are 65,157,341 ingredient combinations, of which 9,839 are expanding; and in drugs, there are 3,130 drug combinations, of which 1,381 are expanding. These are plotted in Fig. 3 as a function of the module size  $k$ .

### Expanding modules are complete technologies

Our hypothesis is that technologies are not built up from scratch, but are the result of new combinations of existing technologies [10, 11]. In other words, the components of technologies are themselves complete technologies, and not just parts thereof.

To test this this hypothesis, we checked to see how many of these expanding modules are complete technologies, and not just parts of technologies. Remarkably, almost all expanding modules are complete technologies: 98.70%, 96.88% and 98.62% in language, gastronomy and drugs.



the space by bringing regions that were once far apart closer together. When a technology of multiple components becomes a single component, it is easier to use that technology to make a new technology because using it is now one step away, rather than the multiples steps required to reach all of its constituent components. In this way technological recursion alters the map of the adjacent possible [3, 4]. The innovator, especially one with limited resources, is biased towards technologies nearby in possibility space because there are fewer parts that he might be lacking [17]. Our insights provide a prescription for which technological recursions maximise the likelihood of discovering new technologies [18].

### New technologies shape their own environment

In a recursive process, the output of the process becomes the input to the same process. Recursive processes are hard to understand because they can shape their own environment; we can no longer assume that the environment is constant, or at least slowly changing, as is commonly assumed in models of evolution, for example. In technological recursion, the process is combining components, and the output—a potential technology—can become part of the input—the repertoire of components for making new technologies.

An important consequence of technological recursion is that technologies shape their own environment: the likelihood of new technologies coming into being can be substantially altered by the technology itself. While the technologies of next year will be made up in part of the technologies of today, the technologies two years hence will be made up of those of next year. Forecasting technology thus involves a hierarchy of prediction problems in an expanding space of building blocks, which helps explain the poor track record of long-term technological prediction [2, 15].

### Drug discovery

Drug combination therapies, one of the three domains that we analyzed, use multiple drugs simultaneously to treat a disease. But designing new combination therapies is difficult because there are a thousand individual drugs which can be used as components [25]. So there are on the order of a million possible two-drug therapies, a billion three-drug therapies, and so on. While there are some rules of thumb for which drugs tend to co-appear in successful therapies, thereby reducing the space of possibilities, drug combination therapy lacks systematic design principles [25]. Similar challenges are found in combinatorial chemistry, which has played a key role in the discovery of many drugs. But such brute force approaches, with their low odds of success, are expensive, and the way in which many pharmaceutical companies approach drug discovery is thought to be unsustainable [29].

Using our insights, pharmaceutical companies can increase the chance of hitting upon new drugs and drug combination therapies by strategically redefining the repertoire of component building blocks to include existing drugs and drug therapies [27, 28]. This is analogous to how, in language, phonemes can be a more successful repertoire of components for building words than letters. Our analysis could be enhanced by access to the proprietary positive and negative results held by drug firms. This would lead to higher discovery hit rates and fewer wasted trials.

## Methods

### Data

Our three data sets were obtained as follows. In language, our list of common English words is from the built-in WordList library in Mathematica 11.3. Of the 39,176 words in WordList, we only considered the 38,970 made from the 26 letters a–z,

ignoring case: we excluded words containing a hyphen, space, and so on. In gastronomy, the 56,498 recipes can be found in the supplementary material in [24]. In drugs, the 1,363 therapies can be found in [25].

### Module invariant

Let  $\mathbf{N}$  be the set of  $N$  possible components and  $\theta$  be a subset of  $k$  components, or module, chosen from  $\mathbf{N}$ . Let  $\mathbf{N}_k$  be the set of  $N - k$  other components not including  $\theta$ , let  $\mathbf{n}_k$  be a subset of  $n - k$  components chosen from  $\mathbf{N}_k$ , and let  $\mathbf{c}_k$  be a subset of  $c - k$  components chosen from  $\mathbf{n}_k$ . The module usefulness  $u_\theta(\mathbf{n}, c)$  is the number of technologies of size  $c$  makeable from the components  $\mathbf{n}$  that the  $k$  components  $\theta$  show up in. We can also think of the module usefulness in a different way. It is how many more technologies of size  $c$  we can make from the components  $\mathbf{n}_k$  together with  $\theta$ , than from the components  $\mathbf{n}_k$  alone:

$$u_\theta(\mathbf{n}, c) = \sum_{\mathbf{c}_k \subseteq \mathbf{n}_k} \mathcal{P}(\theta \cup \mathbf{c}_k) - \mathcal{P}(\mathbf{c}_k),$$

where  $\mathcal{P}(\theta \cup \mathbf{c}_k)$  takes the value 0 if the combination of components  $\theta \cup \mathbf{c}_k$  form no technologies of size  $c$  and 1 if  $\theta \cup \mathbf{c}_k$  form one technology of size  $c$ .

The mean usefulness of the  $k$  components  $\theta$ ,  $\bar{u}_\theta(n, c)$ , is the average of  $u_\theta(\mathbf{n}, c)$  over all subsets  $\mathbf{n}_k \subseteq \mathbf{N}_k$ ; there are  $\binom{N-k}{n-k}$  such subsets. Therefore

$$\begin{aligned} \bar{u}_\theta(n, c) &= 1/\binom{N-k}{n-k} \sum_{\mathbf{n}_k \subseteq \mathbf{N}_k} u_\theta(\mathbf{n}, c) \\ &= 1/\binom{N-k}{n-k} \sum_{\mathbf{n}_k \subseteq \mathbf{N}_k} \sum_{\mathbf{c}_k \subseteq \mathbf{n}_k} \mathcal{P}(\theta \cup \mathbf{c}_k) - \mathcal{P}(\mathbf{c}_k). \end{aligned}$$

Consider some particular combination of components  $\mathbf{c}'_k$ . The double sum above will count  $\mathbf{c}'_k$  once if  $c = n$ , but multiple times if  $c < n$ , because  $\mathbf{c}'_k$  will belong to multiple sets  $\mathbf{n}_k$ . How many? In any set  $\mathbf{n}_k$  that contains  $\mathbf{c}_k$ , there are  $n - c$  free elements to choose, from  $N - c$  other components. Therefore the double sum will count every combination  $\mathbf{c}_k$  a total of  $\binom{N-c}{n-c}$  times, and

$$\begin{aligned} \bar{u}_\theta(n, c) &= \binom{N-c}{n-c} / \binom{N-k}{n-k} \sum_{\mathbf{c}_k \subseteq \mathbf{N}_k} \mathcal{P}(\theta \cup \mathbf{c}_k) - \mathcal{P}(\mathbf{c}_k) \\ &= \binom{n}{c} / \binom{N}{c} \binom{N}{k} / \binom{N}{k} u_\theta(\mathbf{N}, c), \end{aligned} \quad (5)$$

noting that  $\bar{u}_\theta(N, c) = u_\theta(\mathbf{N}, c)$ . Solving both equations for  $\bar{u}_\theta(N, c)$  and equating them, we obtain the exact conservation law for the module usefulness for any two stages  $n'$  and  $n$ :

$$\bar{u}_\theta(n', c) \binom{n'}{k} / \binom{n'}{c} = \bar{u}_\theta(n, c) \binom{n}{k} / \binom{n}{c}.$$

### Technological recursion equation

Setting  $n' = N$  in eq. (1),  $\bar{p}(n, c)$  can be expressed in terms of the number of makeable technologies when we have access to all  $N$  of the possible components,

$$\bar{p}(n, c) = p(\mathbf{N}, c) \binom{n}{c} / \binom{N}{c}, \quad (6)$$

noting that  $\bar{p}(N, c) = p(\mathbf{N}, c)$ .

Now let us apply the technological recursion  $\theta \rightarrow \theta$ , that is, let us replace the combination of  $k$  components  $\theta$  with the single new component  $\theta$  every time the  $k$  components show up in the same technology. Let  $\bar{p}_\theta(n, c)$  be the expected number of technologies of size  $c$  that can be made from  $n$  components drawn from  $\mathbf{N} \cup \theta$ , the union of the  $N$  components  $\mathbf{N}$  and the new component  $\theta$ , giving  $N + 1$  components in total. Just as in eq. (6), we can write

$$\bar{p}_\theta(n, c) = p(\mathbf{N} \cup \theta, c) \binom{n}{c} / \binom{N+1}{c}. \quad (7)$$

When we replace  $\theta$  with  $\theta$ , some technologies of size  $c + k - 1$  are reduced to size  $c$ , and some technologies of size  $c$  are reduced to size  $c - k + 1$ . It is the net change that counts, so the relation between  $p(\mathbf{N} \cup \theta, c)$  and  $p(\mathbf{N}, c)$  is

$$p(\mathbf{N} \cup \theta, c) = p(\mathbf{N}, c) - u_{\theta}(\mathbf{N}, c) + u_{\theta}(\mathbf{N}, c + k - 1).$$

Substituting this into eq. (7) gives

$$\begin{aligned} \bar{p}_{\theta}(n, c) &= (p(\mathbf{N}, c) - u_{\theta}(\mathbf{N}, c) + u_{\theta}(\mathbf{N}, c + k - 1)) \\ &\times \left(1 - \frac{c}{N+1}\right) \binom{n}{c} / \binom{N}{c}. \end{aligned}$$

Substituting eqs. (6) and (5) into this, and approximating  $\binom{n}{c}$  and  $\binom{N}{c}$  by  $n^c$  and  $N^c$  for  $n, N \gg c$ , we find

$$\begin{aligned} \bar{p}_{\theta}(n, c) &\simeq \left(1 - \frac{c}{N+1}\right) (\bar{p}(n, c) - x^k \bar{u}_{\theta}(n, c) + x \bar{u}_{\theta}(n, c + k - 1)) \\ &= \left(1 - \frac{x}{N+1} \frac{d}{dx}\right) (\bar{p}(n, c) - x^k \bar{u}_{\theta}(n, c) + x \bar{u}_{\theta}(n, c + k - 1)), \end{aligned}$$

where  $x = n/N$ . Summing over  $c$ ,

$$\bar{p}_{\theta}(n) \simeq \left(1 - \frac{x}{N+1} \frac{d}{dx}\right) (\bar{p}(n) + (x - x^k) \bar{u}_{\theta}(n)). \quad (8)$$

### Toy model

Despite its simplicity, the toy model in Fig. 1 captures many of our insights, and it is worth working through it explicitly. From eq. (6), the expected number of makeable words when we randomly draw  $1 \leq n \leq 4$  letters from a, b, c and d is

$$\begin{aligned} \bar{p}(n) &= 1 \binom{n}{1} / \binom{4}{1} + 5 \binom{n}{2} / \binom{4}{2} + 4 \binom{n}{3} / \binom{4}{3} \\ &= (2n^3 - n^2 + 2n) / 12, \end{aligned}$$

where the coefficients 1, 5 and 4 are the numbers of words of size 1, 2 and 3.

Now let us replace the letters a and d—which together we call  $\theta$ —with a new single component  $\theta$  whenever a and d show up in the same word. The expected number of makeable words when we randomly draw  $n \in [1, 5]$  letters from a, b, c and d and  $\theta$  is

$$\begin{aligned} \bar{p}_{\theta}(n) &= 4 \binom{n}{1} / \binom{5}{1} + 5 \binom{n}{2} / \binom{5}{2} + 1 \binom{n}{3} / \binom{5}{3} \\ &= (n^3 + 12n^2 + 35n) / 60, \end{aligned}$$

where the coefficients 4, 5 and 1 are the numbers of words of size 1, 2 and 3 after the technological recursion  $\theta \rightarrow \theta$ . For the sake of visualization, in Fig. 1 we interpolate between integer values of  $n$  by replacing the factorials in the binomial coefficients with the gamma function, even though such non-integer values are unphysical.

We can calculate  $\bar{p}_{\theta}(n)$  without explicitly making the substitutions by using eq. (8). We find

$$\begin{aligned} \bar{p}_{\theta}(n) &= \frac{4}{5} (1 + 3 - 0) \binom{n}{1} / \binom{4}{1} + \frac{3}{5} (5 + 3 - 3) \binom{n}{2} / \binom{4}{2} + \\ &\frac{2}{5} (4 + 0 - 3) \binom{n}{3} / \binom{4}{3} \\ &= (n^3 + 12n^2 + 35n) / 60, \end{aligned}$$

which matches the result of explicit substitution above.

### Approximating the number of modules

We calculate the number of modules by counting the subsets of size  $k$  of each technology (pairs, triples, and so on) and

getting rid of duplicates. For data sets that have technologies with many components, such as gastronomy where the largest recipes has 31 ingredients, this is computationally demanding. In this case we approximate the number of modules by assuming that the likelihood a subset of one technology belongs to another technology is negligible, the number of modules is then given by

$$\sum_{c \geq k} p(\mathbf{N}, c) \binom{c}{k}. \quad (9)$$

This is the dashed orange line in Fig. 3, which matches the true number well for  $c \geq 6$ .

### Acknowledgements

The authors thank Martin Reeves and Andriy Fedosyeyev for helpful discussions and Roman Rybiansky for assisting with the figures. A. Teimouri's postdoc position is funded by the BCG Henderson Institute.

\*tf@lims.ac.uk, †it@lims.ac.uk

- [1] D. H. Erwin, D. C. Krakauer, Insights into innovation, *Science* **304**, 1117 (2004).
- [2] T. Felin, S. Kauffman, R. Koppl, G. Longo, Economic opportunity and evolution: Beyond landscapes and bounded rationality, *Strateg Entrep J* **8**, 269 (2014).
- [3] F. Tria et al., The dynamics of correlated novelties, *Sci Rep* **4**, 5890 (2014).
- [4] V. Loreto, V. Servedio, S. Strogatz, F. Tria, Dynamics on expanding spaces: Modeling the emergence of novelties, *Creativity and Universality in Language*, (2016).
- [5] H. Youn, D. Strumsky, L. Bettencourt, J. Lobo, Invention as a combinatorial process: evidence from US patents, *J Roy Soc Interface* **44**, 0272 (2015).
- [6] T. Fink, M. Reeves, R. Palma, R. Farr, Serendipity and strategy in rapid innovation, *Nat Commun* **8**, 2002 (2017).
- [7] T. M. A. Fink, M. Reeves, How much can we influence the rate of innovation?, *Science Advances* **5**, eaat6107 (2019).
- [8] T. E. Stuart, J. M. Podolny, Local search and the evolution of technological capabilities, *Strateg Manage J* **17**, 21 (1996).
- [9] L. Fleming, Recombinant uncertainty in technological search, *Management Sci* **47**, 117 (2001).
- [10] W. B. Arthur, The structure of invention, *Res Policy* **36**, 274 (2007).
- [11] W. B. Arthur, *The Nature of Technology: What It Is and How It Evolves* (Penguin, London, 2010).
- [12] R. N. Langlois, Modularity in technology and organization *J Econ Behav Organ* **49** 1937 (2002).
- [13] K. B. Clark, C. Baldwin *Design Rules: The Power of Modularity* (MIT Press, Cambridge, 2000).
- [14] J. McNerney et al., Role of design complexity in technology improvement, *Proc Natl Acad Sci* **108**, 9008 (2011).
- [15] J. D. Farmer, F. Lafond, How predictable is technological progress?, *Res Policy* **45**, 647 (2016).
- [16] V. Sood et al., Interacting branching process as a simple model of innovation, *Phys Rev Lett*, **105**, 178701 (2010).
- [17] I. Iacopini, S. Milojević, V. Latora, Network dynamics of innovation processes, *Phys Rev Lett* **120**, 048301 (2018).
- [18] D. Rotolo, D. Hicks, B. R. Martin, What is an emerging technology?, *Res Policy* **44**, 1827 (2015).
- [19] C. Hidalgo, R. Hausmann, The buildings blocks of economic complexity, *Proc Natl Acad Sci* **106**, 10570 (2009).
- [20] M. Christelli et al., A metrics for the economic complexity of countries and products, *PLOS ONE* **8**, e70726 (2013).
- [21] R. Van Noorden, Physicists make 'weather forecasts' for economies, *Nature* **1038**, 16963 (2015).
- [22] M. Reeves, T. Fink, Harnessing the secret structure of innovation, *MIT Sloan Manag Rev* **37**, 59 (2017).
- [23] C. Hidalgo, Economic complexity: From useless to keystone, *Nat Phys* **14**, 9 (2018).
- [24] Y.-Y. Ahn, S. E. Ahnert, J. P. Bagrow, A.-L. Barabasi, Flavor network and the principles of food pairing, *Sci Rep* **1**, 196 (2011).
- [25] Y. Liu et al., DCDB 2.0: a major update of the drug combination database, *Database* **2014**, bau124 (2014).
- [26] W. B. Arthur, W. Polak, The evolution of technology within a simple computer model, *Complexity* **11**, (2006).
- [27] D. Verhoeven, J. Bakker, R. Veugelers, Measuring technological novelty with patent-based indicators, *Research Policy* **45**, 707 (2016).
- [28] D. Strumsky, J. Lobo, Identifying the sources of technological novelty in the process of invention, *Research Policy* **44**, 1445 (2015).
- [29] J. Hunter, S. Stephens, Is open innovation the way forward for big pharma?, *Nat Rev Drug Discov* **9**, 87 (2010).