Heliyon 5 (2019) e02219

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.heliyon.com

Heliyon

Quantification of knowledge content of a high impact innovation: recombinant DNA

Hung Tseng [a,*], Henry Small [b,**]

[a] National Institutes of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, 6701 Democracy Boulevard, Bethesda, MD 20892, USA
[b] SciTech Strategies, Inc., 105 Rolling Road, Bala Cynwyd, PA, 19004, USA

ARTICLE INFO

Keywords:
Recombinant DNA,
Scientometrics
Knowledge quantification
Science of science
Molecular biology
Information science

ABSTRACT

Quantitative analysis of knowledge content of a significant technological innovation is a novel approach to understand the scientific discovery process. Here we describe such an analysis applied to the invention of recombinant DNA technology in the early 1970's. Two focal papers are selected, i.e., Jackson et al., 1972 and Cohen et al., 1973. A knowledge framework called EApc is described to categorize knowledge types and their quantification. The focal papers, along with their reference lists, are used to determine the minimal scientific knowledge necessary for generating the notions central to each focal paper. Attempts are made to trace how each type of knowledge was generated by various research communities. The results are discussed in terms of their potential implications in measuring, evaluating, understanding and managing the scientific research process.

1. Introduction

Quantification of scientific progress has been a major concern in scientometrics, whose declared mission is the "quantitative evaluation and inter-comparison of scientific activity productivity and progress" (de Solla Price, 1978; Garfield, 1979; attributed to M.T. Beck). Because of its bibliographic roots and the early impact of the Science Citation Index, the field has been focused on analyzing the communication/informational process as a platform for research on the development of science. As pointed out in a recent review: "Whilst scientometrics can, and to some extent does, study many other aspects of the dynamics of science and technology, in practice it has developed around one core notion—that of the citation." (Mingers and Leydesdorff, 2015) It is therefore possible and may be advantageous to develop new frameworks and tools in scientometrics that are not entirely citation based.

In search for an alternative to the published papers as units of analyzing scientific process, it appears knowledge itself is a good substitute. Knowledge can be viewed as the basic currency for all scientific investigations, being their starting materials as well as end products. More importantly, knowledge is the content being communicated in the scientific papers, hence many scientometric methods can be applied to capturing and analyzing it. The challenge of this approach is that to most research scientists, knowledge is an amorphous and abstract object, not suitable for measurement and analysis by the methods familiar to most scientific disciplines. On the other hand, the nature of knowledge has been the subject of inquiry in philosophy and epistemology. Centuries of studies in these fields have generated a wealth of understanding of the nature and essence of knowledge accumulation, which creates a fertile ground for frameworks, concepts, and ideas for gauging and measuring scientific activities in research enterprises and laboratories. This resource, however, is largely ignored by scientists in research fields, from training new investigators to professional practice of established researchers, thus missing the potential benefit of this knowledge (Bosch, 2018).

The problem of introducing structure into knowledge has been contemplated for centuries, and the earliest example, still easily accessible now, is Aristotle's "Categoriae", in which he named ten categories of knowledge (Smith, 2018). Ontology, a branch of philosophy, also provides many guiding principles on how to approach this problem. Another example of categorization that has long lasting impact on the practice of science comes from biology, namely Linnaeus's *Systema Naturae,* which introduced structure and order to a large and complex collection of entities. These examples suggest clearly that the first step in developing methods of qualitative and quantitative analyses of knowledge should be a categorization framework that can be applied across the scientific disciplines and is rooted in the scientific inquiry process, so it can also be used for explaining the scientific discovery process, thus making the scientometric analysis meaningful.

In this study, we attempt to approach measuring science from the

* Corresponding author.
** Corresponding author.
E-mail addresses: htsengpe@gmail.com (H. Tseng), hsmall@mapofscience.com (H. Small).

https://doi.org/10.1016/j.heliyon.2019.e02219
Received 18 September 2018; Received in revised form 11 April 2019; Accepted 30 July 2019

building blocks of scientific investigation, namely scientific knowledge, exploring its categorization and quantification, its basic and hierarchical structures, and the process, by which various knowledge types are generated. We aim to produce a new framework that can serve as an alternative platform, providing new perspectives, where new insights can be gained about scientific research and new tools can be generated to gauge its progression. We begin by proposing a theoretical framework, in which knowledge is divided into two large classes, *basic* and *compound*. Basic knowledge in turn consists of four categories 1) Entities and 2) Properties associated with an entity and 3) Actions and 4) Conditions associated with an action. We term this classification system the EApc framework. We define compound knowledge as a synthesis of basic knowledge. We proceed to demonstrate the utility of this framework by applying it to analyzing the recombinant DNA invention, showing that knowledge can be quantified with a precision that has not been achieved before; and that the framework can also provide a structure to dissect and understand the knowledge discovery process, yielding new insights that are either novel, complementary to or consistent with the existing views. Lastly, we discuss our results in the context of current concepts and notions in the science of science research.

2. Theory

2.1. A framework of knowledge

The knowledge framework proposed here is based on many concepts and ideas arising from studies in the philosophy of science, which can be traced back to the Aristotle's time, and later works by classic empiricists, logic positivists and other schools. The framework is described below and the origin of the ideas is referenced where possible. We term this basic knowledge categorization scheme EApc (Entity, Action and their associated Properties and Conditions). We adopt the following convention in this work regarding usage of upper case for the first letter in entities, actions, properties and conditions. When the word refers to a knowledge category, the upper case letter is used; when the word refers to an individual object or process, the lower case is used, unless dictated otherwise by grammatical rules (e.g, at the beginning of a sentence). The EApc acronym is adopted, instead of EAPC, to depict that properties and conditions are subsidiary to entities and actions, respectively.

In this work, "knowledge" refers to information dealt with in science, i.e., scientific knowledge, but not other types of knowledge, such as, historical knowledge, political knowledge, artistic knowledge, etc. This restriction is meant to simplify the analysis presented here. Whether this framework is applicable to other types of knowledge requires further investigation. In our definition, knowledge is relatively stable information, whose revision requires considerable scientific evidence. In this sense, a hypothesis is initially not knowledge, but a theory may be (e.g. Darwinian evolution). Therefore, knowledge can be used repeatedly in a variety of scientific activities. The framework described here does not deal with the ultimate validity of knowledge and assumes that all knowledge is tentative and may be subject to future revision. A distinction between knowledge and observation has to be made here. Knowledge and observation may share the attributes of longevity (persistence) and utility, but an observation lacks the explanatory power of knowledge and hence will be replaced by knowledge when a deeper understanding of the underlying generalizable knowledge is obtained. In this paper, we use mainly biochemical and molecular biology examples to illustrate this knowledge framework, because of the nature of the recombinant DNA invention.

2.2. Knowledge categorization

We divide knowledge, based on relationship, hierarchy and complexity into two classes, basic and compound.

For basic knowledge, we propose two principle categories: Entity and Action and two associated categories: Property and Condition (Table 1) (Fig. 1). Entity by our definition is an independently existing object, e.g., a protein, an enzyme, a fragment of DNA, RNA, a cell, an organ, a species, etc (Fig. 1A). The notion of Entity is influenced by Aristotle's concept of "substance" (Aristotle, *Categoriae*) as the most fundamental category of knowledge ("If there were no substances, there would be nothing else". Robin Smith, 2018). Action, by our definition, is a process that is required to transform an entity to another or to alter its property (e.g., breakage or formation of chemical bonds, physical displacement, force, energy consumption, etc (Fig. 1B). Aristotle classified action (he termed it *motion*) as a type of knowledge (predicate) but did not consider it as basic as "substance". The basic nature of Entity and Action was hinted in Paul Thagard's analysis of the evolution of scientific concepts (Thagard, 1992, e.g., figures on pp 42–44). A distinction between entity and action can be found in other information systems, such as in computer languages, e.g., the differentiation between variables and operators, which might have been derived from other knowledge systems, like mathematics (i.e., variables and operations) and natural language (nouns and verbs). Since both mathematics and natural language are considered a priori knowledge, the analogy (entity-action vs variable/operation vs noun/verb) suggests that the division of Entity and Action might be a true cognitive separation. Entities and Actions can either be physical or conceptual (existing only in the mind), and often are known first as conceptual then later physical.

2.3. Hierarchy in Entity and Action

Entities are assemblies of other entities, e.g., proteins are assemblies of amino acids, which in turn are assembled from atoms and other molecular moieties. An entity assembly may possess new properties not present in the component entities. This definition is consistent with the concept of "emergence" as defined by Goldstein, referring to "the arising of novel and coherent structure, patterns and properties during the process of self-organization in complex system" (Goldstein, 1999, 2007). We also adopt the multi-scale notion in meso-science (Li et al., 2009, 2013; Service, 2012) that substances (entities) can be grouped by their "scales" (Fig. 1A). Similarly, in our framework, actions can also be collections of other actions, which possess characteristics that differs from the component actions (Fig. 1B). For example, in biology, DNA replication is a collection of actions (i.e., hydrogen bonding, phosphate bonding, ionic interactions, and many other chemical and physical actions) and Darwinian natural selection is the result of multi-scale interactions from molecular to social.

2.4. Property and Condition are knowledge associated respectively with Entity and Action

Each entity has its associated properties, such as composition, structure, molecular weight, color, temperature, physical location, relationship with other entities, etc. And similarly, each action has its associated conditions, such as speed/duration, range, substrate/product, optimal pH/temperature, and co-factor requirements, etc. (Table 1) (Fig. 1) A property can be considered collectively in the context that it is shared by many physical entities, e.g., heat-resistance, but in most cases, especially in understanding a given process and inventing a method, properties are almost always associated with an entity.

2.5. Synthesis of basic knowledge (compound knowledge)

At the next level, knowledge is of a synthetic nature (Fig. 1C), of

Table 1
Definitions of the four types of basic knowledge.

Category	Definition
Entity	An object, physical or mental.
Property	Attributes of an entity.
Action	A process, required to change an entity or its property.
Condition	Attributes of an action.

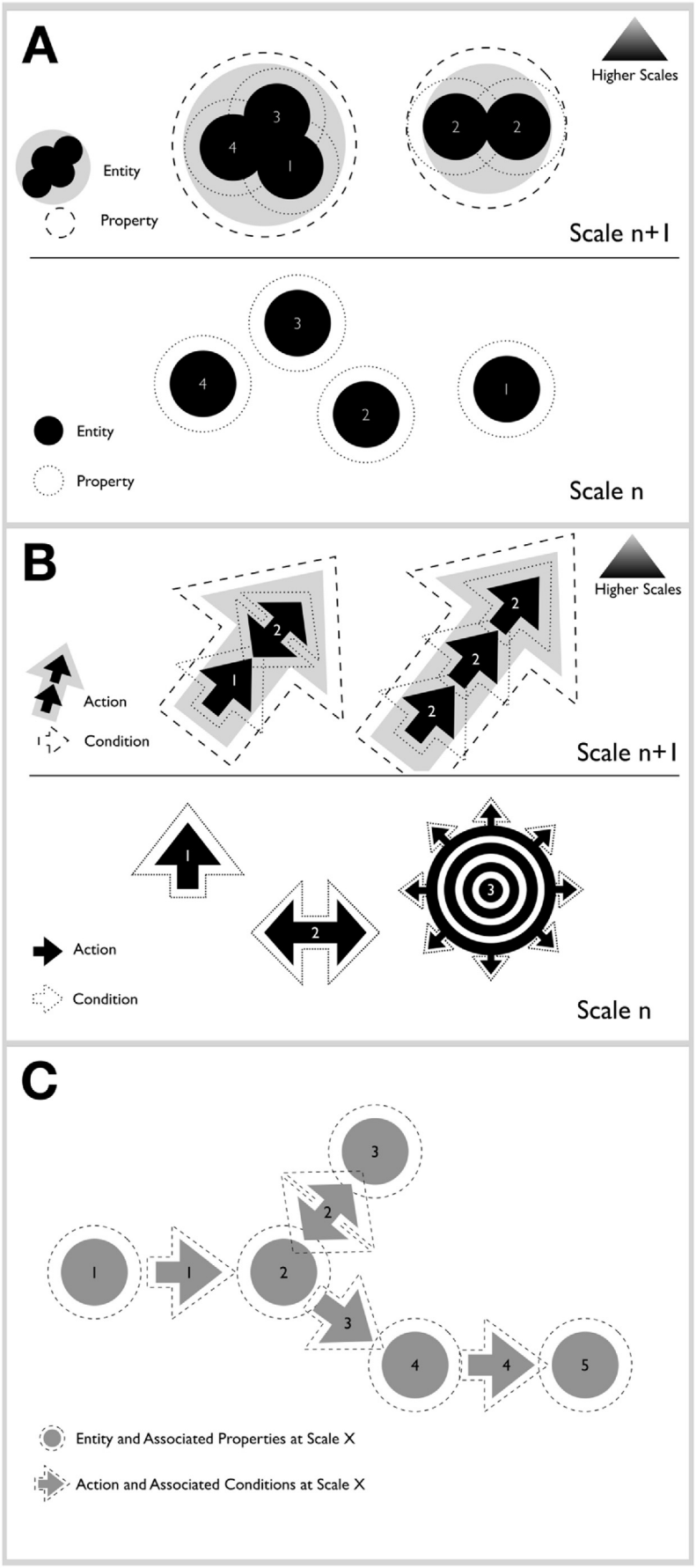

Fig. 1. A graphic depiction of the EApc framework. A, Entities and associated properties. Two scales of entities are shown, Scale n and n+1, to illustrate their relations. The scales can extend one way or the other to higher or lower scales. Although lower scale entities (black) are depicted in the higher scale (gray, in n+1), it should be clear that the higher scale entities behave as a unit and have properties that can be completely different from the lower scale entities. B, Actions and associated conditions. Similar to A, except that actions are depicted as directional, one-arrow head, or bidirectional, two-arrow head (2), or omnidirectional (3) as well. C, Synthesis of the basic knowledge (A and B). Entities and Actions are linked based on their associated properties and conditions. Note the scales of entities and actions are undetermined (X) in the figure because in the synthesis they can be across several orders of scales such as from molecules to bacteria (multi-scale).

which the recombinant DNA technology is a good example. This is where knowledge gains most of its explanatory and predictive power, as well as its utility. Most knowledge we deal with in research is a combination of entities/actions based on their properties/conditions. Basic Knowledge, as is here defined, i.e., knowledge of entities and actions, does not (or rarely) directly benefit human society. Instead, they form the foundation for another type of knowledge that has impact on us in our everyday life. We term this type of knowledge *Compound Knowledge* because it is the result of logical or intuitive synthesis of the basic knowledge or basic and other compound knowledge. Compound knowledge encompasses all of our understanding of processes and our invention of methods. In biomedical research, knowledge of processes and methods includes necessary/sufficient relationships of agents and physiological process and these agents' interventional usage in fighting diseases. It should be pointed out here that Hypotheses are synthesis of basic and compound knowledge but do not constitute knowledge themselves because they are guessworks to be tested and rejected or supported.

3. Study area

In this study, we select two papers for analysis, i.e., Jackson et al., 1972 and Cohen et al., 1973, which are original for the recombinant DNA invention. Recombinant DNA technology refers to a set of laboratory methods used for joining DNA fragments from various sources and propagating the resulting recombinants in live organisms, e.g., bacteria or animal cells. At the core of the technology are four entities, 1) a vector DNA, 2) a cutting enzyme, 3) a joining enzyme, 4) a host organism. The vector DNA (plasmid or viral DNA) is a circular DNA that is capable of living in a host organism and propagates (replicates) with the host. The cutting enzyme (i.e., restriction endonucleases) is used to open the vector DNA (from circular to linear) to allow the insertion of DNA from another source. The joining enzyme (i.e., DNA ligase) is used to link the vector DNA and the inserted DNA, creating a stable new circular recombinant molecule, which is to be introduced into a host organism for propagation via DNA replication. It has been generally acknowledged that two publications were at the center in the development of the recombinant DNA technology: Jackson et al. (1972) and Cohen et al. (1973), the former from the laboratory of Paul Berg at Stanford University and the latter, from that of Stanley Cohen and Herbert Boyer at Stanford and University of California, San Francisco, respectively. While the core concept of artificially combining DNA fragments stayed the same for both papers, they differed substantially in implementing this concept into practice. The two recombinant DNA methods used different vector DNAs (virus vs plasmid) and different host systems (animal cells vs bacteria); but employed the same cutting enzyme (restriction enzyme Eco RI) and the same joining enzyme (DNA ligase). Another intricate difference is how the two methods approached the DNA ends to be joined. The focal papers thus provide an interesting case to test a method of precision knowledge quantification, which cannot be achieved without delving into the content of the research papers.

The recombinant DNA technology is one of the most significant innovations of the 20th century and its impact is still being felt today. It played key roles in manufacturing pharmaceutical products such as insulin and vaccines, in improving agriculture and husbandry productions and in making gene therapy possible. Furthermore, in addition to its impact on the healthcare and economy, the technology in its original form and in combination with other technologies have served as foundations for other high impact research methodology innovations, such that it is hard to imagine what the biomedical science knowledge base would look like today without this technology. Therefore, few would question the qualification of this innovation as an "exemplar" of the scientific discovery and technological innovation process. Because of this status, the invention of the recombinant DNA technique has been the subject of numerous studies from a variety of disciplines, e.g., history of science, bibliography of science, scientometrics etc., which is supplemented by memoirs and retrospective accounts by participants, central and peripheral to the invention, as well as by contemporaries, creating a wealth of documents one could use to reconstruct many of the events.

The choice of the focal papers was partially influenced by the early works of ours (Small and Greenlee, 1980; Tseng, 1999). Small and Greenlee combined co-citation analysis with citation context analysis to probe the cognitive structure of the invention of recombinant DNA. The two focal papers, Cohen et al. (1973) and Jackson et al. (1972), emerged prominently in this analysis, which is also in agreement with later studies that attributed these two papers as key conceptual and technological contributors to the invention. Tseng analyzed the essence of Cohen et al. technique and proposed a cloning method without using restriction enzyme and DNA ligase.

4. Analyses

4.1. Principle of knowledge quantification with the EApc framework

At the basic level, knowledge can be quantified by counting entities and actions. Properties can also be quantified in this way, e.g., plasmid is a circular DNA; DNA polymerase I has a molecular weight of X Dalton etc.; each attribute counting as one property. Actions can be treated similarly, e.g., DNA ligation is one action, restriction digestion another, etc., and the conditions, e.g., substrates of the ligase, such as DNA or RNA, or product of restriction digestion, cohesive ends or blunt ends, etc. each of these conditions, or attributes of action can be counted as well.

4.2. Knowledge content of the focal papers

In assessing quantitatively the knowledge content of the recombinant DNA invention, the notion that a reference/citation is a "molecular unit of thought" (Garfield, 1955) was employed to help reveal the knowledge structure of the focal papers. Thus, the knowledge content of each focal paper is approached first from its reference list, which is considered to reflect the authors' own understanding of the key ingredients of their invention and their desire to acknowledge its sources. The reference lists of the focal papers were examined, looking for clues to what knowledge was essential for the recombinant DNA invention. To identify the relevant bibliographic references, a distinction was made between "scientific object" and "technological object(s)" as defined by Hans-Jörg Rheinberger (1992a, b). Using this definition, knowledge (entities, actions and compound knowledge) essential for the formation of the invention was the scientific object, whereas methodology for verification of the concept and demonstration of the feasibility were considered "technological objects". In general, references in a scientific publication can be classified as 1) for introduction/background, 2) for scientific premise, 3) for methodology, and 4) for speculation (discussion), following the IMRD structure. We consider the knowledge content of the subject matter (recombinant DNA) resides primarily in 2) and 3), and hence concentrate on these references as an initial approximation of the knowledge content of the focal papers (Table 2).

Analyzing the "scientific objects" of the referenced papers in each focal paper revealed several biochemical entities and actions. Both focal papers referenced publications on biochemical entities: 1) restriction endonuclease (Eco RI), 2) DNA ligase (E. coli DNA ligase), 3) cloning vectors (SV40 and plasmids, e.g., pSC101); and biochemical actions: a) breaking of phosphodiester bond in DNA (nuclease digestion), b) forming phosphodiester bond in DNA (ligation), c) adding nucleotides to DNA (DNA synthesis/replication and terminal transfer). This analysis led to the realization that entities and actions, along with their associated properties and conditions, which we refer here collectively as EApc, are at the foundation of this invention. And if so, EApc, which are *countable*, could be used to assess quantitatively the knowledge content of an invention and other scientific works.

To tabulate the EApc of each recombinant procedure, the processes were broken down into individual steps, with the adjacent steps separated by one or more action(s) (Arrowheads in Fig. 2).

This was done for Jackson et al. by using Fig. 1 in their paper, which is a stepwise flow-chart of the process. Because Cohen et al. did not provide such a flow-chart, the steps constituting the procedure were deduced from the text in the paper, primarily the description in the second paragraph on page 3243. Fig. 2 provides a graphic representation of the stepwise recombinant processes of Jackson et al. (Fig. 2A), and Cohen et al (Fig. 2B). The EApc for each step were identified and shown in Tables 3 and 4 for Jackson et al. and Cohen et al. respectively. Table 5 through 12 show the descriptions in these papers of the entities (Tables 5 and 9), properties (Tables 6 and 10), actions (Tables 7 and 11), and conditions (Tables 8 and 12). Table 13 lists the counts of EApc of the two recombinant procedures. The Cohen et al. procedure consisted of a branch structure, i.e., there were two alternatives for the second step, IIa and IIb, the former relying on DNA ligation within the bacteria (in vivo) and the latter on the action of DNA ligase in the test tube (in vitro). Each alternative is listed separately in Table 13. We also grouped the conditions required for enzymatic reaction and bacterial transformation as one condition, to simply the analysis. The following shorthand notations are used: Entity in Jackson et al. Ej, Entity in Cohen et al. Ec, and similarly, Aj and Ac, Pj and Pc, and Cj and Cc, for Action, Property and Condition in each paper, respectively.

Table 13 makes it apparent that the technique proposed by Jackson et al. requires twice the number of entities (Ej = 12 vs Ec = 5 or 6) and 2 to 3 times actions/conditions (Aj = 7 vs Ac = 2 or 3; Cj = 7 vs Cc = 2 or 3), relative to that of Cohen et al. Comparing the entities and actions listed in Tables 3 and 4 shows that Jackson et al. used a number of DNA modification enzymes, i.e., lambda exonuclease, terminal transferase, DNA polymerase I and exonuclease III and their associated enzymatic actions, that were not used by Cohen et al. Apparently, these enzymes were used to create complementary cohesive ends of SV40 DNA and to repair the imperfection of the gaps and fraying of single-stranded DNA after annealing and formation of the circular dimer SV40 DNA for ligation. Although both Jackson et al. and Cohen et al. cloning procedures shared an entity, the Eco RI restriction endonuclease, to cleave the vector DNA (i.e., plasmid for Cohen et al. and SV40 for Jackson et al.), Jackson et al. were seemingly unaware of or choose to ignore the DNA end structure generated by Eco RI (complementary cohesive ends). And additional DNA modification enzymes were used to create, on SV40 linear DNA (vector), homo-polymer ends complementary only to the homo-polymer ends on the insert, preventing recircularization of the vector, and hence increasing the yield of the recombinant (Discussion in Jackson et al.). On the other hand, Cohen et al. took advantage of the property of the complementary cohesive end of DNA created by the Eco RI's cleavage, which significantly simplified the procedure, when the foreign DNA to be inserted was also cleaved by Eco RI. It seems that Cohen et al. realized this significant advantage and devoted 25% (4/15) of the references in their paper to the DNA recognition sequences and cutting sites of restriction enzymes. In Cohen et al. the drawback of recircularization of the vector was overcome by using antibiotic selection to enrich the desired recombinant.

This strategy of Cohen et al. was shown in the citation list of their paper, with one third of references (5/15) on antibiotic resistance. Interestingly, the counts for properties are about the same for each procedure (Pj = 22 vs Pc = 19 or 20), albeit these properties are associated with different entities. A prominent difference is that Cohen et al. used a number of drug sensitivity/resistance properties of plasmids (i.e., tetracycline, kanamycin, and chloramphenicol), which facilitated the analysis of the cloning results. The average properties per entity for Cohen et al. is 3–4, compared to 2 for Jackson et al. This knowledge quantification result raises an interesting question in regard to the Occam's razor (or law of economy), which can be expressed as "Entities are not to be multiplied beyond necessity" (Encyclopaedia Britannica, Occam's razor). While our analysis shows that neither method used unnecessary entities for its stated concept, can one conclude that Cohen et al. is more widely used because it employs fewer entities (and actions)?

It is probably premature to reach a generalizable conclusion from the limited cases analyzed here. However, this analysis does suggest a principle of knowledge synthesis that knowing the entity is not sufficient, and knowing the right property is critical. Other than the recombinant DNA examined here, a significant step in the invention of the polymerase chain reaction (PCR) was the switch from the original thermolabile DNA polymerase to a thermostable enzyme, which "greatly simplifies the procedure and, … significantly improves the specificity, yield, sensitivity, and length of products that can be amplified." (Saiki et al., 1988). In this case, using an entity that combines the polymerase activity (Property) with the thermostability (Property) made the PCR technique simpler yet more powerful. We postulate that a successful invention is probably a skillful selection of entities that maximizes the number of properties useful for achieving the goals of the invention (i.e., the average number of useful properties per entity is high).

There is a great deal of compound knowledge used in the focal papers, e.g., DNA replication, bacterial transformation, antibiotic resistance, SV40 transformation of mammalian cells, etc., which was necessary for the invention; as well as methods, such as electrophoresis, electron-microscopy, buoyant density centrifugal analysis of DNA, DNA

Table 2
Classification of the references in Jackson et al. and Cohen et al. by their appearance in the sections (IMRD), their roles as to scientific or technological objects (see text) and their knowledge types and topic areas.

Reference by Jackson et al., 1972						
Authors	Title	Journal	IMRaD	Sci/Tech	Knowledge Category	Topic Area
Sambrook+Westphal+Srinivasan +Dulbecco	The Integrated State of Viral DNA in SV40-Transformed Cells	PNAS	Introduction	Scientific	Property	SV40 Integration
Dulbecco	Cell Transformation by Viruses	Science	Introduction	Background	Theory (synthesis)	Viral Cell Transformation
Matsubara+Kaiser	Lambda dv: An Autonomously Replicating DNA Fragment	CSHSQB	Methods	Scientific	Entity	DNA replication
Radloff+Bauer+Vinograd	A Dye-buoyant-density Method for the Detection and Isolation of Closed Circular Duplex DNA: The Closed Circular DNA in Hela Cells	PNAS	Methods	Technological	Entity	Circular DNA in Eukaryotic Cell
Little+Lehman+Kaiser	An Exonuclease Induced by Bacteriophage Lambda: I. Preparation of the Crystalline Enzyme	J Biol Chem	Methods	Scientific	Entity	Exonuclease
Kato+Gonalves+Houts+Bollum	Deoxynucleotide-polymerizing Enzymes of Calf Thymus Gland: II. Properties of the Terminal Deoxynucleotidyltransferase	J Biol Chem	Methods	Scientific	Property	Terminal Transferase

(continued on next page)

Table 2 (*continued*)

Reference by Jackson et al., 1972						
Authors	Title	Journal	IMRaD	Sci/Tech	Knowledge Category	Topic Area
Jovin+Englund+Kornberg	Enzymatic Synthesis of Deoxyribonucleic Acid: XXVII. Chemical Modifications of Deoxyribonucleic Acid Polymerase	J Biol Chem	Methods/ Results	Background	Property	DNA Polymerase
Olivera+Hall+Anraku+ Chien+Lehman	On the Mechanism of the Polynucleotide Joining Reaction	CSHSQB	Methods	Scientific	Action	DNA Ligation
Richardson+Lehman+Kornberg	A Deoxyribonucleic Acid Phosphatase-Exonuclease from Escherichia coli: II. Characterization of the Exonucleases Activity	J Biol Chem	Methods	Scientific	Property	Exonuclease
Symons	Preparation of [alpha-32P]nucleoside and deoxynucleoside 5'triphosphates from 32Pi and protected and unprotected nucleosides	Biochim Biophys Acto	Methods	Technological	Method (synthesis)	DNA Labeling
Davis+Simon+Davidson	Electron microscope heteroduplex methods for mapping regions of base sequence homology in nucleic acids	Method in Enzymology	Methods	Technological	Method (synthesis)	DNA Visualization
Chang+Bollum	Enzymatic Synthesis of Oligodeoxynucleotides	Biochemistry	Results	Scientific	Method (synthesis)	DNA Synthesis*
Sheldon+Jurale+Kates	Detection of Polyadenylic Acid Sequences in Viral and Eukaryotic RNA	PNAS	Methods/ Results	Technological	Entity	Poly A Sequence
Richardson+Schildkraut +Kornberg	Studies on the Replication of DNA by DNA Polymerases	CSHSQB	Results	Scientific	Action	DNA Replication
Little	An Exonuclease Induced by Bacteriophage Lambda: II. Nature of the Enzymatic Reaction	J Biol Chem	Results	Scientific	Action	Exonuclease
Sgaramella+van de Sande+Khorana	Studies on Polynucleotides, C. A Novel Joining Reaction Catalyzed by the T4-Polynucleotide Ligase	PNAS	Discussion	Scientific	Action	DNA Ligation
Melgar+Goldthwait(+Ukstins)	Deoxyribonucleic Acid Nucleases: II. The Effects of Metals on the Mechanism of the Action of Deoxyribonucleases I	J Biol Chem	Discussion	Scientific	Action (Condition)	DNase
Morrow+Berg	Cleavage of Simian Virus 40 DNA at a Unique Site by a Bacterial Restriction Enzyme	PNAS	Results	Scientific	Action	Restriction Endonuclease
Sgaramella+Lobban	(not found)	(Nature)	Discussion			**
Reference by Cohen et al. 1973						
Authors	Title	Journal	IMRaD	Sci/Tech	Knowledge Category	Topic Area
Cohen+Chang	Recircularization and Autonomous Replication of a Sheared R-Factor DNA Segment in Escherichia coli Transformants	PNAS	Intro/Metho/ Results	Scientific	Action	Plasmid Recircularization
Hedgpeth+Goodman+Boyer	DNA Nucleotide Sequence Restricted by the RI Endonuclease	PNAS	Introduction	Scientific	Property	Restriction Enzyme
Bigger+Murray+Murray	Recognition Sequence of a Restriction Enzyme	Nature New Biology	Introduction	Scientific	Property	Restriction Enzyme
Boyer+Chow+Hedgpeth +Goodman	DNA Substrate Site for the EcoRII Restriction Endonuclease and Modification Methylase	Nature New Biology	Introduction	Scientific	Property	Restriction Enzyme
Greene+Betlach+Goodman +Boyer	DNA Replication and Biosynthesis	Methods in Mol Biol	Intro/ Methods(?)	Scientific	Property (?)	DNA Replication
Mertz+Davis	Cleavage of DNA by RI Restriction Endonuclease Generates Cohesive Ends	PNAS	Intro/Metho/ Results	Scientific	Property	Restriction Enzyme
Cohen+Chang+Hsu	Nonchromosomal Antibiotic Resistence in Bacteria: Genetic Transformation of Escherichia coli by R-Factor DNA	PNAS	Intro/Metho/ Results	Background	Method (synthesis)	Drug Resistence
Cohen+Miller	Non-chromosomal Antibiotic Resistance in Bacteria II. Molecular Nature of R-factor isolated from Proteus mirabilis and Escherichia coli	J Mol Biol	Methods/ Results	Background	Property	Drug Resistence
Sharp+Hsu+Ohtsubo+Davidson	Electron Microscope Heteroduplex Studies of Sequence Relations among Plasmids of Escherichia coli	J Mol Biol	?	Technological	Property	DNA Analysis
Sharp+Cohen+Davidson	Electron Microscope Heteroduplex Studies of Sequence Relations among Plasmids of Escherichia coli II. Structure of Drug Resistance (R) Factors and F Factors	J Mol Biol	Results	Technological	Property	DNA Analysis
Modrich+Lehman	Deoxyribonucleic Acid Ligase: A Steady State Kinetic Analysis of the Reaction Catalyzed by the Enzyme from Escherichia Coli	J Biol Chem	Methods	Scientific	Action	DNA Ligase
Jacob+Brenner+Cuzin	On the Regulation of DNA Replication in Bacteria	CSHSQB	Results	Background	Theory (synthesis)	Bacterial DNA Replication
Guerry+VanEmbden+Falkow	Molecular Nature of Two Nonconjugative Plasmids Carrying Drug Resistance Genes	J Bacteriology	Results	Scientific	Property	Drug Resistence
Anderson+Lewis	Characterization of a Transfer Factor Associated with Drug Resistance in Salmonella typhimurium	Nature	Results	Technological	Property	Drug Resistence
Davies+Brzezinska+Benveniste	R Factors: Biochemical Mechanisms of Resistance to Aminoglycoside Antibiotics	Ann NY Acad Sci	Results	Scientific	Action	Drug Resistence

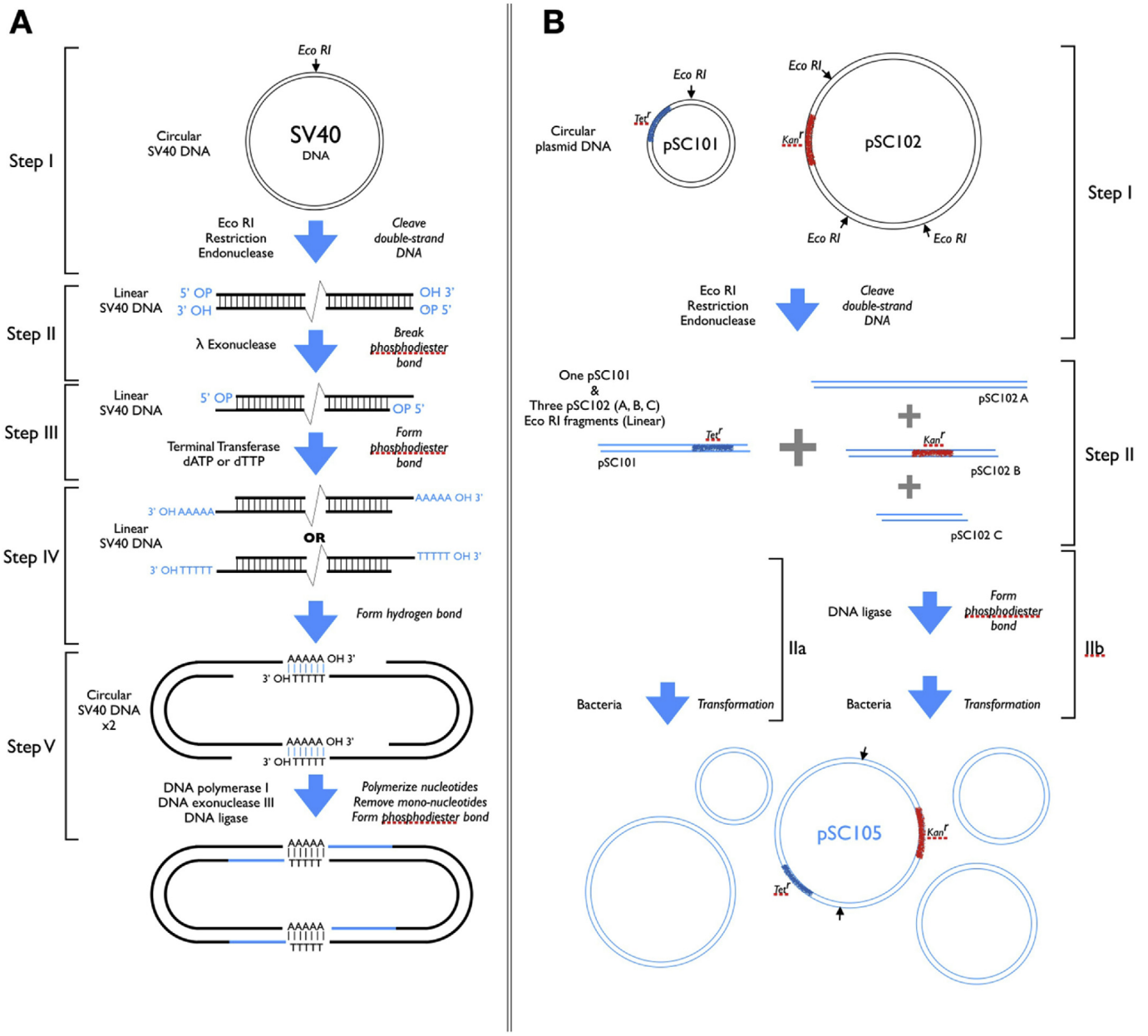

Fig. 2. A graphic representation of the recombinant DNA processes of Jackson et al. (**A**) and Cohen et al. (**B**). Panel A was adopted mainly from Fig. 1 in Jackson et al. (1972), with modifications (i.e., additional labels and colors) to link it to the content of this paper. Panel B was drawn based on the text in the Results section (mainly p.3243 second paragraph) and the Materials and Methods of Cohen et al. (1973) (see Tables 9, 10, 11, and 12 for additional information). The processes are broken-down into individual steps, which are labeled left to Panel A (Steps I to V) and right to Panel B (Steps I to IIa and IIb). The steps are divided by actions, depicted as light blue arrows. Changes between steps are also color-coded (light blue) for easy visualization of the progression of the process. Entities are written in regular font and shown to the left of the arrows, and Action in italic and to the right. DNA is drawn as two parallel solid lines; hydrogen bonds are shown in A as short parallel line arrays, but not in B because of lesser relevance to the concept. In B, antibiotic resistance are depicted as short curved lines (color coded) in the DNA. Both the location and the length of the lines were approximations because the exact location and size of the resistance genes were not known at the time. The circular SV40 dimer DNA and the plasmid pSC105 at the bottom of Panels A and B, respectively, are the desired products of the recombinant processes. In B, the byproducts generated by the process of Cohen et al. are depicted as circles of various sizes without label. Note the recombinant process brought the two antibiotic genes together into one plasmid, which could be selected by applying these antibiotics in the bacterial culture.

labeling, enzyme and DNA preparations, to name a few, which were analytical and preparative techniques used in the experimental demonstrations of the invention. For two reasons we decided not to analyze the knowledge contents of these methods and techniques: 1) the focal recombinant DNA papers do not provide enough detail for the analysis, and requires going back to the original publications describing these techniques; 2) inclusion of quantitative analysis of these compound knowledge here would require a much longer discussion. Thus we concentrate on illustrating the principle of knowledge quantification, and plan to describe additional analyses elsewhere.

Table 3
Analysis of the Entity/Property and Action/Condition of the Jackson et al. recombinant procedure. The steps are based on Fig. 1 in Jackson et al. (page 2905).

Process	Entity	Property	Action	Conditions
Step I	Circular SV40 DNA	Can be introduced into host cells Circular DNA Cut by Eco RI at one site		
	Eco RI endonuclease	Cut DNA at specific site	Cleave double-stranded DNA	Enzymatic reaction Substrate Enzymatic activity Reaction conditions
		Generate complementary cohesive ends		
Step II	Linear SV40 DNA	Linear DNA End structure (thought to be blunt)		
	Lambda exonuclease	Remove mono-nucleotide from 5′ end Processive 5′ to 3′	Break phosphodiester bond	Enzymatic reaction Substrate Enzymatic activity Reaction conditions
Step III	Linear SV40 DNA	Linear DNA End structure 3′ overhang		
	Terminal transferase	Polynucleotide synthesis at 3′ OH Template independent	Form of phosphodiester bond	Enzymatic reaction Substrate
	dATP or dTTP	Complementary on DNA		Enzymatic activity Reaction conditions
Step IV	Linear SV40 DNA	Poly A or Poly T 3′ overhang end structure	Form hydrogen bond	Annealing temperature
Step V	SV40 DNA	circular dsDNA of m.w Dimer (two copies) Hydrogen-bonded Open or (frayed) single-stranded region at ends		
	E.coli DNA polymerase I	DNA-dependent DNA synthesis	Polymerize nucleotides	Enzymatic reaction Substrate Enzymatic activity Reaction conditions
	E.coli DNA exonuclease III	3′ to 5′ exonuclease	Remove mono-nucleotides from 3′ OH end of dsDNA	Enzymatic reaction Substrate Enzymatic activity Reaction conditions
	E.coli DNA ligase	Joining nicks in dsDNA	Form phosphodiester bond	Enzymatic reaction Substrate Enzymatic activity Reaction conditions

4.3. Discovery Process of Each Type of Knowledge

Categorization can help analyzing knowledge qualitatively and quantitatively, and it can also provide an intellectual framework for understanding and analyzing the discover process; for each type of knowledge may have a distinctive discovery path and specific roles in knowledge synthesis. Testing the EApc framework in this manner is critical because it endows meaning to the qualification and quantification of knowledge in understanding scientific progress. Below, we describe our observations from studying the knowledge accumulation process leading to the invention of the recombinant DNA technique based on the references in the two focal papers. We highlight the EApc categories by italicizing/bolding the words to help readers follow the use of the framework in the analysis.

4.4. Discovery of DNA ligase

The ***action*** carried out by DNA ligase (the reunion of double-stranded DNA ends) was first realized when genetic recombination was observed between bacteriophages, and that such recombination requires double-strand DNA breakage and DNA ends rejoining (Meselson and Weigle, 1961). Martin Gellert (1967) first reported the DNA end joining chemical reaction in E.coli cell extract, characterized the reaction and identified several necessary ***conditions*** (e.g., presence of ATP and magnesium) for the reaction to proceed. This study led to the isolation of a DNA ligase from E. coli. In parallel to Gellert's line of investigation, four other groups also reported isolation of DNA ligase from E. coli and T4 bacteriophage, thus firmly establishing the physical ***entity*** DNA ligase (Weiss and Richardson, 1967; Gefter et al., 1967; Cozzarelli et al., 1967; and Olivera and Lehman, 1967).

This line of research can be described in the framework we propose here: the ***action*** (the rejoining of double-stranded DNA ends) was first discovered as a conceptual chemical reaction required for genetic recombination. This ***action*** in bacteria was studied in detail to reveal the ***conditions*** of the ***action***, which hinted the existence of an ***entity*** (conceptual), an enzyme, and eventually led to the biochemical identification of the DNA ligase (physical ***entity***).

4.5. Discovery of restriction endonucleases

The discovery of restriction endonuclease can be traced back to observations made in the early 1950's when strains of bacterial virus (phage) were found to change their host-specificity during infection of host bacteria (Luria and Human, 1952; Bertani and Weigle, 1952). Werner Arber's group carried out extensive characterization of this phenomenon and proposed the DNA Modification and Restriction Hypothesis to explain the observation. The hypothesis stated that bacteria have the ability to destroy foreign DNA (e.g., DNA from different bacteria strains and bacteriophages) by enzymatic attacks, while protecting their own DNA with a molecular modification, thus portraying a bacterial immune function similar to that of the recently discovered CRISPR system. The DNA Modification and Restriction Hypothesis entailed two enzymatic systems (conceptual ***entities*** and ***actions***), modifying DNA by methylation and digesting unmodified DNA by endonucleolysis (and other conceptual entities less critical for recombinant DNA technology, which will not be discussed further here), as the underlying mechanisms of this bacterial defense (Arber, 1965; Arber and Linn, 1969).

Several restriction endonucleases were isolated from various bacterial strains in the ensuing years before the recombinant DNA invention. Meselson and Yuan assumed restriction was carried out by a nuclease

Table 4
Analysis of the Entity/Property and Action/Condition of the Cohen et al. recombinant procedure. The steps are deduced from the description in the second paragraph on page 3243 of Cohen et al. Cohen et al. described two alternative second steps, which are indicated as Step IIa and IIb.

Process	Entities	Properties	Actions	Conditions
Step I	pSC101 plasmid	double-strand DNA of m.w. Can be introduced into bacterial cells (E. coli) Contains replicator Cut by Eco RI at one site Tetracycline resistant Kanamycin sensitive		
	pSC 102 plasmid	double-strand DNA of m.w. Contains replicator Can be introduced into bacterial cells (E. coli) Cut by Eco RI at three sites Tetracycline sensitive Kanamycin resistant		
	Eco RI endonuclease	Cut double-strand DNA at specific site Generate complementary cohesive ends	Cleave double-strand DNA	Enzymatic action Substrate Enzymatic activity Reaction condition
Step IIa	plasmid Eco RI fragments	One pSC101 fragment Three pSC102 fragments m.w. of each fragment m.w. of each fall fragment carry Eco RI cohesive endsragment		
	E.coli bacteria	Plasmid DNA can be introduced into E.coli and propagate in it.	Transform bacteria	Number of bacteria Temperature Incubation duration
Step IIb	plasmid Eco RI fragments	One pSC101 fragment Three pSC102 fragments m.w. of each fragment All fragments carry Eco RI cohesive ends		
	E. coli DNA ligase	Joining nicks in dsDNA	Form phosphodiester bond	Enzymatic action Substrate Enzymatic activity Reaction condition
	E.coli bacteria	Plasmid DNA can be introduced into E.coli and propagate in it.	Transform bacteria	Number of bacteria Temperature Incubation duration

(conceptual ***entity***) and set up an assay system accordingly (known ***properties*** and ***conditions*** of nucleases) to isolate it (Meselson and Yuan, 1968). They called the purified enzyme Restriction Endonuclease K* because it was isolated from E. coli strain K. Pursuing a similar line of investigation under the framework of Arber's hypothesis, Boyer's group isolated E. coli B and RI restriction enzymes (physical ***entity***) (Boyer and Rouland-Dussoix, 1969), the latter turning out to be critical for the invention of recombinant DNA, because it was used by both focal papers (Jackson et al., 1972; Cohen et al., 1973). The third group, which isolated restriction enzymes from *Hemophilus influenzae,* was interestingly not initially guided by Arber's hypothesis, and instead, was looking for enzymatic systems (conceptual ***entities***) engaged in genetic recombination, a "chance discovery" as described by the authors (Smith and Welcox, 1970; Roberts, 2005). Although the DNA Modification and Restriction Hypothesis did not direct the isolation of endonuclease R (the *Hemophilus* enzyme), as it was called at the time, Arber's hypothesis did

Table 5
Entities and their descriptions in Jackson et al. recombinant process. The Step and Entity listings are identical to that in Table 3, thus this table serves as an extension and explanation of that table. Not all relevant descriptions are listed because of repetition and to conserve space.

Entity	Description in Jackson et al. (not all relevant descriptions are listed)
Circular SV40 DNA	Fig. 1 on p.2905; Materials and Methods section, under DNA, Enzyme, Conversion of SV40 (I) DNA to Unit Length Linear DNA SV40(LRI) with RI Endonuclease.
Eco RI endonuclease	Fig. 1 on p.2905; Materials and Methods section, under Enzyme: "The circular SV40 and λ-dvgal DNA molecules were cleaved with the bacterial restriction endonuclease RI (Yoshimori and Boyer, Unpublished…)"
Linear SV40 DNA	Fig. 1 on p.2905; Results section (p.2905), paragraph 1, "the first step requires conversion of the circular structures to linear duplex."
Lambda exonuclease	Fig. 1 on p.2905; Materials and Methods section, under Enzyme and Removal of 5′-Terminal Regions from SV40 (LRI). Also Results section (p. 2906) under General Approach "prior to removal of a short sequence (30–50 nucleotides) from the 5′-phosphorial termini by digestion with lambda exonuclease facilitates the terminal transferase reaction."
Linear SV40 DNA	Fig. 1 on p.2905; Results section (p.2905) under General Approach (see Step II under Linear DNA)
Terminal transferase	Fig. 1 on p.2905; Results section (p.2905) under General Approach "Relatively short (50–100 nucleotides) poly (dA) or poly (dT) extensions are added on the 3′-hydroxyl termini of the linear duplexes with terminal transferase."
dATP or dTTP	Fig. 1 on p.2905; Materials and Methods (p.2905) under Addition of Homopolymeric Extentions to SV40(LRIexo) with Terminal Transferase.
Linear SV40 DNA	Fig. 1 on p.2905; Results (p.2907), under Hydrogen-Bonded Circular Molecules Are Formed by Annealing SV40 (LRIexo)-(dA)80 and SV40 (LRIexo)-(dT)80 Together.
SV40 DNA	Fig. 1 on p.2905; Results (p.2907), under Covalently Closed-Circular DNA Molecules Are Formed by Incubation of Hydrogen-Bonded Complexes with DNA Polymerase, Ligase and Exonuclease III.
E.coli DNA Polymerase I	Fig. 1 on p.2905; Materials and Methods (p.2094) under Enzyme; and Results (p.2907), under Covalently Closed-Circular DNA Molecules Are Formed by Incubation of Hydrogen-Bonded Complexes with DNA Polymerase, Ligase and Exonuclease III.
E.coli DNA exonuclease III	Fig. 1 on p.2905; Materials and Methods (p.2094) under Enzyme; and Results (p.2907), under Covalently Closed-Circular DNA Molecules Are Formed by Incubation of Hydrogen-Bonded Complexes with DNA Polymerase, Ligase and Exonuclease III.
E.coli DNA ligase	Fig. 1 on p.2905; Materials and Methods (p.2094) under Enzyme; and Results (p.2907), under Covalently Closed-Circular DNA Molecules Are Formed by Incubation of Hydrogen-Bonded Complexes with DNA Polymerase, Ligase and Exonuclease III.

Table 6
Entities, associated properties and their descriptions in Jackson et al. recombinant process. The Step, Entity, and Property listings are identical to that in Table 3, thus this table serves as an extension and explanation of that table. Not all relevant descriptions are listed because of repetition and to conserve space.

Step	Entity	Property	Description in Jackson et al.
Step I	Circular SV40 DNA	Can propagate in host cells	Introduction (p.2904): "It is known that the DNA of the transforming virus SV40 can enter into a stable, heritable, and presumably covalent association with the genomes of various mammalian cells (1, 2)"
		Circular DNA	Fig. 1 (p.2905) and Results (p2905): "the first step requires conversion of the circular structures to linear duplexes."
		Cut by Eco RI at one site	See above "Circular DNA" and "This could be achieved by a double-strand scission at random locations (see Discussion) or, as we describe in this paper, at a unique site with RI restriction endonuclease."
	Eco RI endonuclease	Cut DNA at specific site	Results (p.2906): "Digestion of SV40(I) DNA with excess RI endonuclease yields a product ... and appears as a linear duplex with the same contour length as SV40(II) DNA…. The point of cleavage is at a unique site on the SV40 DNA,…"
		Generate complementary cohesive ends	Fig. 1 (p.2905) depicted the ends as blunt end. The authors appeared to be aware of the end structure because in the Results (p.2906): "the termini at each end are 5′-phosphoryl, 3′-hydroxyl (Mertz, J., Davis, R., in preparation)", Mertz and Davis (1972) reports the complementary cohesive ends of Eco RI cut.
Step II	Linear SV40 DNA	Linear DNA	Fig. 1 on p.2905; Results section (p.2905), paragraph 1, "the first step requires conversion of the circular structures to linear duplex."
		Substrate of lambda exonuclease	Results (p.2906):" Lobban and Kaiser (unpublished) found that P22 phage DNA became a better primer for homopolymer synthesis after incubation of the DNA with λ exonuclease….We confirmed their finding with SV40(LRI) DNA; after removal of 30–50 (p.2907) nucleotides per 5′-end (see Methods), the number of SV40(LRI) molecules that could be bound to poly(U) filters after incubation with terminal transferase and dATP increased 5- to 6-fold."
	Lambda exonuclease	Remove mono-nucleotide from 5′ end	Results (p.2906):"This enzyme removes, successively, deoxymononucleotides from 5′-phosphoryl termini of double-stranded DNA (15), thereby rendering the 3′-hydroxyl termini single-stranded."
		Processive 5′ to 3′	See above "...removes, successively…"
Step III	Linear SV40 DNA	Linear DNA	Fig. 1 on p.2905; Results section (p.2905), paragraph 1, "the first step requires conversion of the circular structures to linear duplex."
		End structure 3′ overhand	Fig. 1 on p.2905; Results section (p.2906), "... thereby rendering the 3′-hydroxyl termini single-stranded."
	Terminal transferase	Polynucleotide synthesis at 3′ OH	Fig. 1 on p.2905; Results (p.2906): "Terminal transferase has been used to generate deoxyhomopolymeric extensions on the 3′- hydroxyl termini of DNA (7); …"
		Template independent	Implied in Results (p.2906): "Lobban and Kaiser (unpublished) found that P22 phage DNA became a better primer for homopolymer synthesis after incubation of the DNA with λ exonuclease…" Also in Little et al. (1967), #5 reference in Jackson et al.
	dATP or dTTP	Complementary on DNA	Fig. 1 on p.2905; Results (p.2906): "Linear duplexes containing (dA), extensions are annealed to the DNA to be joined containing (dT)n extensions at relatively low concentrations."
Step IV	Linear SV40 DNA	Poly A or Poly T 3′ overhang end structure	Fig. 1 on p.2905; Results (p.2906): "Linear duplexes containing (dA), extensions are annealed to the DNA to be joined containing (dT)n extensions at relatively low concentrations."
Step V	SV40 DNA	Circular DNA of m.w.	Fig. 1 on p.2905; Results (p.2907) section Covalently Closed-Circular DNA Molecule Are Formed by Incubation of Hydrogen-Bonded Complexes with DNA Polymerase, Ligase and Exonuclease III.
		Dimer (two copies)	Fig. 1 on p.2905; Table 1 on p.2907; Results (p.2907): "DNA isolated from the heavy band of the Cs–Cl-ethidium bromide gradient contains primarily circular molecules, with a contour length twice that of SV40 (II) DNA (Table 1) when viewed by electron microscopy."
		Hydrogen-bonded	Fig. 1 on p.2905; Results (p.2907): "The hydrogen-bonded complexes described above can be sealed by incubation with the E.coli enzymes DNA polymerase I, ligase and exonuclease III."
		Open or (frayed) single-stranded region at ends	Fig. 1 on p.2905; Results (p.2906): "Repair of the four gaps is mediated by E. coli DNA polymerase with the four deoxynucleoside-triphosphates…" Other possible DNA structures after hydrogen-bond formation were not mentioned in the paper (e.g., frayed single-stranded DNA).
	E.coli DNA polymerase I	DNA-dependent DNA synthesis	Implied. See above. Also in Jovin et al. (1969), #7 reference in Jackson et al.
	E.coli DNA exonuclease III	3′ to 5′ exonuclease	Results (p.2907): "Exonuclease III is probably needed to remove 3′-phosphate groups from 3′-phosphoryl, 5-hydroxyl nicks introduced by the endonuclease contaminating the terminal transferase preparation."
	E.coli DNA ligase	Joining nicks in dsDNA	Implied. Also in Olivera and Lehman, 1967, #8 reference in Jackson et al.

provide the proper framework to understand the discovery. Endonuclease R played a key role in another revolutionary invention of analyzing genomic DNA (Danna and Nathans, 1971), and because this line of innovation is not directly related to the topic of recombinant DNA, it will not be pursued further here.

The discovery processes of restriction enzymes outlined above suggest that transition from conceptual ***Entities*** to physical ***Entities*** may take at least two paths; by direct guidance of what properties are known of the conceptual entity or, by chance, when seemingly unrelated research goals were pursued.

It should be pointed out that only one of the above-mentioned restriction enzymes, namely the RI restriction endonuclease, was used in the focal papers of the recombinant DNA invention. This is the result of extensive characterization of these enzymes (physical ***entities***) after their isolation and purification by their discoverers and others. The ***properties*** of these enzymes, e.g., molecular weight, peptidyl composition, host strain, etc., and the ***conditions*** of endonucleolytic ***action***, e.g., the nature of the substrate, recognition site on DNA, products and their ***properties***, and co-factor requirements of the enzymatic ***action***, were examined and characterized in detail. These characterizations led to the classification of the known (at the time) restriction enzymes into two types. Type I enzymes (e.g., Restriction Endonuclease K, Meselson and Yuan, 1968) recognize specific DNA sequence but their endonucleolytic cut is random, and thus unsuitable for the recombinant DNA technique as it was formulated in the focal papers. On the other hand, Type II enzymes (e.g., Restriction Endonuclease R, Smith and Welcox, 1970) recognize and cut DNA at the same site, making predicting and engineering DNA joining more manageable.

Two clear distinctions separate the recombinant DNA procedures proposed by Jackson et al. and Cohen et al. namely the vector (SV40 vs. plasmid) and the way RI restriction enzyme was used. As first reported by Gellert (1967), DNA-ligase-catalyzed DNA-end joining required hydrogen-bonded ends (a ***condition*** for DNA ligation ***action***). In the procedure proposed by Jackson et al. generating the hydrogen-bonded DNA ends was based on a method pioneered by Peter Lobban and Dale Kaiser (Lobban PhD thesis 1968, Lobban and Kaiser, 1973), in which, SV40 DNA was first cut by RI restriction enzyme (Eco RI) and then treated with a number of DNA modifying enzymes (i.e., lambda

Table 7
Action and their descriptions in Jackson et al. recombinant process. The Step and Action listings are identical to that in Table 3, thus this table serves as an extension and explanation of that table. Not all relevant descriptions are listed because of repetition and to conserve space.

Step	Action	Description in Jackson et al.
Step I	Cleave double-stranded DNA	Fig. 1 (p.2905). Results (p.2905): "This could be achieved by a double-stranded scission at random locations or, as we described in this paper, at a unique site with RI restriction endonuclease."
Step II	Break phosphodiester bond	Results (p.2906): "…incubation of DNA with λ exonuclease. This enzyme removes, successively, deoxynomonucleotides from 5′-phosphoryl termini of double-stranded DNA (15)."
Step III	Form phosphodiester bond	Implied in Results: "Terminal transferase has been used to generate deoxyhomopolymeric extensions on the 3′-hydroxyl termini of DNA (7)."
Step IV	From hydrogen bond	Materials and Methods (p.2905), section title: Formation of Hydrogen-Bonded Circular DNA Molecules.
Step V	Template-dependently polymerize nucleotides	Implied in Fig. 1 (p.2905) and Results (p.2907): "Covalent closure of the hydrogen-bonded SV40 DNA dimers is dependent on Mg2+, all four deoxynucleoside triphosphates, E. coli DNA polymerase I, and ligase", and reference #7 Jovin et al. (1969).
	Remove mono-nucleotides from 3′ OH end of double-stranded DNA	Results (p.2907): "Exonuclease III is probably needed to remove 3′-phosphate groups from 3′-phosphoryl, 5′-hydroxyl nicks introduced by the endonuclease contaminating the terminal transferase preparation."
	Form phosphodiester bond	Implied in Results (p.2906): "…covalent closure of the ring structure is effected by E. coli DNA ligase;…" and #8 reference (Olivera and Lehman, 1967).

Table 8
Actions, associated conditions and their descriptions in Jackson et al. recombinant process. The Step, Action, and Condition listings are identical to that in Table 3, thus this table serves as an extension and explanation of that table. Not all relevant descriptions are listed because of repetition and to conserve space.

Step	Action	Condition	Description in Jackson et al.
Step I	Cleave double-stranded DNA	Enzymatic reaction Substrate Enzymatic activity Reaction condition	Materials and Methods (p.2905): "[3H]SV40(I) DNA (18.7 nM) in 100 mM Tris•HCl buffer (pH 7.5)-10 mM MgCl2-2 mM 2-mercaptoethanol was incubated for 30 min at 37 °C with an amount of RI previously determined to convert 1.5 times this amount of SV40(I) to linear molecules [SV40(LRI)]"
Step II	Break phosphodiester bond	Enzymatic reaction Substrate Enzymatic activity Reaction condition	Materials and Methods (p.2905): "[3H]SV40(LRI) (15 nM) in 67 mM K-glycinate (pH 9.5), 4 mM MgCl2, 0.1 mM EDTA was incubated at 0 °C with λ-exonuclease (20 μg/ml) to yield [3H]SV40(LRIexo) DNA.
Step III	Form phosphodiester bond	Enzymatic reaction Substrate Enzymatic activity Reaction condition	Materials and Methods (p.2905): "[3H]SV40(LRIexo) (50 nM) in 100 mM K-cacodylate (pH 7.0), 8 miM MgCl2, 2 mM 2-mercaptoethanol, 150 μg/ml of bovine serum albumin, [α-32P]dNTP (0.2 mM for dATP, 0.4 mM for dTTP) was incubated with terminal transferase (30–60 μg/ml) at 37 °C."
Step IV	From hydrogen bond	Enzymatic reaction Substrate Enzymatic activity Reaction condition	Materials and Methods (p.2905): "[32P]dA and -dT DNAs were mixed at concentrations of 0.15 nM each in 0.1 M NaCl-10 mM Tris•HCl (pH 8.1)-1 mM EDTA. The mixture was kept at 51 °C for 30 min, then cooled slowly to room temperature."
Step V	Polymerize nucleotides	Enzymatic reaction Substrate Enzymatic activity Reaction condition	Materials and Methods (p.2905): "After annealing of the DNA, a mixture of the enzymes, substrates, and cofactors needed for closure was added to the DNA solution and the mixture was incubated at 20 °C for 3–5 h. The final concentrations in the reaction mixture were: 20 mM Tris•HCl (pH 8.1), 1 mM EDTA, 6 mM MgCl2, 50 μg/ml bovine-serum albumin, 10mM NH4Cl, 80 mM NaCl, 0.052 mM DPN, 0.08 mM (each), dATP, dGTP, dCTP, and dTTP, (0.4 μg/ml) E. coli DNA polymerase I, (15 units/ml) E. coli ligase, and (0.4 unit/ml) E. coli exonuclease III.
	Remove mono-nucleotides from 3′ OH end of dsDNA	Enzymatic reaction Substrate Enzymatic activity Reaction condition	
	Form phosphodiester bond	Enzymatic reaction Substrate Enzymatic activity Reaction condition	

Table 9
Entities and their descriptions in Cohen et al. recombinant process. The Step and Entity listings are identical to that in Table 4, thus this table serves as an extension and explanation of that table. Not all relevant descriptions are listed because of repetition and to conserve space.

Step	Entity	Description in Cohen et al. (1973)
Step I	pSC101 plasmid pSC102 plasmid Eco RI endonuclease	Results (p.3243): "A mixture of pSC101 and pSC102 plasmid DNA species,…., was treated with the Eco RI restriction endonuclease…"
Step IIa	plasmid Eco RI fragments E. coli bacteria	Results (p.3243): "A mixture of pSC101 and pSC102 plasmid DNA species,…., was treated with the Eco RI restriction endonuclease…" "and then was either used directly to transform E. coli…"
Step IIb	plasmid Eco RI fragments E.coli DNA ligase E. coli bacteria	Results (p.3243): "A mixture of pSC101 and pSC102 plasmid DNA species,…., was treated with the Eco RI restriction endonuclease…" "or was ligated prior to use in the transformation procedure."

exonuclease, terminal transferase) in sequence to generate the complementary single-stranded ends. After annealing (forming hydrogen-bond) of the ends, the gaps (remaining single-stranded region due to unmatching lengths of the complementary single-stranded ends) and fraying single-stranded DNA were repaired by E. coli DNA polymerase I and exonuclease III, followed by DNA ligation (see Fig. 1 in Jackson et al., 1972). It appears that these authors were not aware of (or chose to ignore) an important ***condition*** of the RI restriction enzymatic ***action***,

Table 10
Entities, associated properties and their descriptions in Cohen et al. recombinant process. The Step, Entity, and Property listings are identical to that in Table 4, thus this table serves as an extension and explanation of that table. Not all relevant descriptions are listed because of repetition and to conserve space.

Step	Entity	Property	Description in Cohen et al. (1973)
Step I	pSC 101 plasmid	dsDNA of m.w.	Fig. 1 (p.3241) legend: "(c) pSC101. The calculated molecular weight of the single fragment is 5.8 x 10⁴"
		contains replicator	Results (p.3241): "The ability of two plasmids derived from the same parental plasmid (i.e., R6-5) to exist stably as sparate replicons (12) in a single bacterial host cell suggests that the parent plasmid may contain at least two distinct replicator sites."
		can be introduced into bacterial cells (E. coli)	Materials and Methods (p.3240): "Other bacterial strains and R factors and procedures for DNA isolation, electron microscopy, and transformation of E. coli by plasmid DNA have been described (1, 7, 8)."
		cut by Eco RI at one site	Results (p.3240): "Only one band is observed after EcoRI endonucleolytic digestion of pSC101 DNA (Fig. 1c), suggesting that this plasmid has a single site susceptible to cleavage by the enzyme."
		tetracycline resistant kanamycin sensitive chloramphenicol sensitive	Introduction (p.3240): "This plasmid carries genetic information necessary for its own replication and for expression of resistance to tetracycline, but lacks the other drug resistance determinants and the fertility functions carried by R6-5 (1)." Also implied in Results (p.3242) Table 1: No pSC101 transformant seen in kanamycin (neomycin) and chloramphenicol selections.
	pSC102 plasmid	dsDNA of m.w.	Results (p.3241-2): "Closed circular DNA obtained from this isolate (plasmid designation pSC102) by CsCl-ethidium bromide gradient centrifugation has an S value of 39.5 in neutral sucrose gradients (Fig. 2A) and a contour length of 8.7 micron when nicked (Fig. 2B). These data indicate a molecular weight about 17 × 10⁶."
		contains replicator	Results (p.3241): "The ability of two plasmids derived from the same parental plasmid (i.e., R6-5) to exist stably as sparate replicons (12) in a single bacterial host cell suggests that the parent plasmid may contain at least two distinct replicator sites."
		can be introduced into bacterial cells (E. coli)	Materials and Methods (p.3240): "Other bacterial strains and R factors and procedures for DNA isolation, electron microscopy, and transformation of E. coli by plasmid DNA have been described (1, 7, 8)."
		cut by Eco RI at three (3) sites	Results (p.3242): "Treatment of pSC102 plasmid DNA with EcoRI restriction endonuclease results in formation of three fragments that are separable by electrophoresis is agarose gels (Fig. la); …"
		tetracycline sensitive kanamycin resistant chloramphenicol sensitive	Results (p.3241): "A single clone that had been selected for resistance to kanamycin and which was found also to carry resistance to neomycin and sulfonamide, but not to tetracycline, chloramphenicol, or streptomycin after transformation of E. coli by EcoRI generated DNA fragments of R6-5, was examined further. Closed circular DNA obtained from this isolate (plasmid designation pSC102)…"
	Eco RI endonuclease	Cut dsDNA at specific site	Introduction (p.3240): "The EcoRI endonuclease seemed especially useful for this purpose, because on a random basis the sequence cleaved is expected to occur only about once for every 4,000 to 16,000 nucleotide pairs (2)…"
		Generate complementary cohesive ends	Introduction (p.3240): (EcoRI) "The nucleotide sequences cleaved are unique and self-complementary (2–6) so that DNA fragments produced by one of these enzymes can associate by hydrogen-bonding with other fragments produced by the same enzyme."
Step IIa	plasmid Eco RI fragments	One pSC101 fragment Three pSC201 fragments	Fig. 1(p.3241) and Figure 4b (p.3242)
		m.w. of each fragment	Fig. 1 (p.3241) and Figure 4b (p.3242). Molecular weight of DNA can be inferred from their migration distance in the gel.
		All fragments carry EcoRI cohesive ends	Implied, because these plasmids were digested by EcoRI.
		Linear DNA with complementary cohesive ends can recircularize and be ligated in bacterial host	Results (p.3241): "The ability of cleaved pSC101 DNA to function in transformation suggests that plasmid DNA fragments with short cohesive endonuclease-generated termini can recircularize in E. coli and be ligated in vivo;…"
	E. coli bacteria	Plasmid DNA can be introduced into E. coli and propagate in it.	Abstract (p.3240): "Newly constructed plasmids that are inserted into Escherichia coli by transformation are shown to be biologically functional replicons…"
Step IIb	plasmid Eco RI fragments	One pSC101 fragment Three pSC201 fragments	Fig. 1(p.3241) and Figure 4b (p.3242)
		m.w. of each fragment	Fig. 1 (p.3241) and Figure 4b (p.3242). Molecular weight of DNA can be inferred from their migration distance in the gel.
		All fragments carry EcoRI cohesive ends	Implied, because these plasmids were digested by EcoRI.
	E.coli DNA ligase	joining nicks in dsDNA	Introduction (p.3240): "that DNA fragments produced by one of these enzymes can associate by hydrogen-bonding with other fragments produced by the same enzyme. After hydrogen-bonding, the 3′-hydroxyl and 5′-phosphate ends can be joined by DNA ligase (6)."
	E. coli bacteria	Plasmid DNA can be introduced into E. coli and propagate in it.	Abstract (p.3240): "Newly constructed plasmids that are inserted into Escherichia coli by transformation are shown to be biologically functional replicons"

which creates precisely matching complementary single-stranded ends, suitable for hydrogen-bond formation, rendering the elaborate procedure for creating the imprecise complementary ends (cohesive ends) unnecessary. By the time Cohen et al. formulated their procedure, this end structure of RI restriction action (***property*** of the RI-digested linear DNA) was known too, and taken advantage of by these authors (Sgaramella, 1972; Mertz and Davis, 1972). The result was a simpler method easily adaptable for wide use.

This example illustrates the importance of ***Properties*** and ***Conditions*** of ***Entities*** and ***Actions***, and may partially explain, along with curiosity, that after making the initial discoveries of entities, the discoverers spent years to characterize them, publishing follow-up papers describing the results (Arthur Kornberg's lab published 36 serial papers titled "Enzymatic Synthesis of Deoxyribonucleic Acid I-XXXVI" with individual subtitles from 1957 to 1971). On one hand, both focal papers used knowledge from these papers, probably more than they acknowledged in citations. On the other hand, the knowledge used is still a very small fraction of that contained in the publications of these vast and diverse research programs from the field of DNA modification enzymology.

The analyses thus far suggest that ***Properties*** (***Entities***) and

Table 11
Actions and their descriptions in Cohen et al. recombinant process. The Step and Action listings are identical to that in Table 4, thus this table serves as an extension and explanation of that table. Not all relevant descriptions are listed because of repetition and to conserve space.

Step	Action	Description
Step I	Cleave double-stranded DNA	Introcution (p.3240): "Two recently described restriction endonuclease, Eco RI and Eco RII, cleave double-stranded DNA so as to produce short overlapping single-stranded ends."
Step IIa	Transform bacteria	Results (p.3243): "…was treated with the Eco RI endonuclease, and the was either used directly to transform E. coli…"
Step IIb	Form phosphodiester bond	Implied in the statement: "After hydrogen-bonding, the 3′-hydroxyl and 5′-phosphate ends can be joined by DNA ligase (6)." in Introduction (p.3240).
	Transform bacteria	Results (p.3243): "…was treated with the Eco RI endonuclease, and the was either used directly to transform E. coli…"

Table 12
Actions, associated conditions and their descriptions in Cohen et al. recombinant process. The Step, Action and Condition listings are identical to that in Table 4 of the main text, thus this table serves as an extension and explanation of that table. Not all relevant descriptions are listed because of repetition and to conserve space.

Step	Action	condition	Description
Step I	Cleave double-stranded DNA	Enzymatic action Substrate Enzymatic actitivity Reaction condition	Materials and Methods (p.3240): "Purification and use of the Eco RI restriction endonuclease have been described (5)."
Step IIa	Transform bacteria	Number of bacteria Temperature Incubation duration	Materials and Methods (p.3240): "…and transformation of E. coli by plasmid DNA have been described (1, 7, 8). "
Step IIb	Form phosphodiester bond (by E.coli DNA ligase)	Enzymatic action Substrate Enzymatic actitivity Reaction condition	Materials and Methods (p.3240): "E. coli DNA ligase was a gift from P. Modrich and R. L. Lehman and was used as described (11)."
	Transform bacteria	Number of bacteria Temperature Incubation duration	Materials and Methods (p.3240): "…and transformation of E. coli by plasmid DNA have been described (1, 7, 8). "

Table 13
Quantification of knowledge content of the recombinant DNA procedures.

	Entity	properties	Action	conditions	Total
Jackson et al.	12	22	7	7	48
Cohen et al.‡	5	19	2	2	28
Cohen et al.*	6	20	3	3	32

‡ Based on Cohen et al. Step IIa, which relies on DNA circularization and ligation in bacteria.

* Based on Cohen et al. Step IIb, which uses DNA ligase in vitro to join Eco RI-digested DNA fragments.

Conditions (***Actions***) play two roles in the discovery process, they serve as hints or clues of the existence of an ***Entity*** or ***Action***, leading to their conceptualization, while initial ***Properties*** and ***Conditions*** guide identification, isolation and purification of the underlying physical ***Entities*** and ***Actions***.

4.6. Plasmid and antibiotic resistance

The discovery of plasmid as a physical ***entity*** spanned more than a decade (approximately from 1952 to 1967), involving multiple disciplines (e.g., bacterial genetics, bacterial conjugation, extra-chromosomal inheritance, bacterial drug resistance) with intertwining interests. The conceptual ***entity*** preceding that for plasmid, the F factor (fertility factor), was first proposed in 1952, after observations made by Esther M Lederberg of a pattern in the results of bacterial conjugation experiments. The essence of the discovery process was revealed in her own words "… when the attributes of F were known: namely its transfer to other cells by contact. I called it F without knowing what it was, but knowing that its presence in one parent was necessary for recombination." (Lederberg, 1993) Thus, similar to the way in which conceptual ***entities*** of enzymes were formed (see above), the conceptual ***entity*** F factor was proposed based on the observed ***properties*** of a presumed physical ***entity***. Joshua Lederberg coined the term "plasmid" in late 1952, in an attempt to clarify the confusion in the terminology referring to the conceptual ***entity*** in the cytoplasmic inheritance (e.g., pangenes, bioblasts, plasmagenes, plastogenes, chondriogenes, cytogenes and provirus) and its relationship to nucleus inheritance (Lederberg, 1952).

Another conceptual ***entity*** for plasmid, the R Factor (for drug resistance), emerged at the beginning of the 1960's, during an effort to understand the rapid and wide spread of antibiotic resistance in post-war Japan (Watanabe, 1963). This conceptual ***entity*** (R factor) was proposed mainly based on the ***property*** of plasmid transmitting antibiotic resistance (and other antagonistic ***properties*** to other anti-bacterial substance). Watanabe and his colleagues were among the first to recognize that in the R factor, the drug resistance (the resistance factor) could be separated (physically) from the transfer action, which they termed "resistance transfer factor" or RTF. They observed and hypothesized that RTF was similar to the F factor, because the former competed with the latter in bacterial conjugation experiments. The convergence of the conceptual ***entities***, i.e. the F factor and the R factor, in drug resistance studies led to their biochemical purification (Cohen and Miller, 1969; Nisioka and Mitani, 1969; Silver and Falkow, 1970) and the eventual demonstration that both were extrachromosomal circular DNA species (Falkow et al., 1966; Roth and Helinski, 1967, among others).

The discovery of plasmid illustrates a challenge in the knowledge accumulation process that a physical ***entity*** can be represented by many conceptual ***entities***, each formed based on a limited number of ***properties*** of the physical ***entity***. And a conceptual ***entity*** may represent more than one physical ***entity*** with overlapping ***properties*** (e.g., the conceptual entity termed episome). This situation can cause considerable confusion in both understanding the biochemical nature of the underlying ***entity*** as well as in guiding the isolation and purification of the ***entity***. Donald R. Helinski, a veteran researcher of plasmid, commented insightfully of this situation "While these various classification schemes serve a useful purpose in relating distinguishable R factors to each other, the many different combinations of these properties observed for R factors to date preclude the application of any one or two criteria in the identification of a single R plasmid." (Helinski, 1973). This difficulty is amplified many times when ***Entities and Actions*** with high complexity are the subject matter of research, such as biological species and diseases.

4.7. Nescience in theory formation

Another lesson learned from the discovery of plasmid is the role of the ***Entity*** *and* ***Actions*** (conceptual or physical), the basic knowledge by our classification, in the synthesis of compound knowledge, which explains the bacterial drug resistance phenomenon. Bacterial antibiotic resistance was widely noted in the early post World War II era and attempts were made to explain the biological nature of this resistance (for an in-depth analysis, see Creager, 2007). Two opposing hypotheses were offered, the "adaptation hypothesis" and the "mutation and selection hypothesis". The adaptation hypothesis was mainly backed by bacteriologists who were influenced by the concurrent view in the inducibility of bacteria to utilize various carbon sources, whereas the mutation-selection hypothesis was supported by geneticists mainly focusing on nucleus-centered inheritance. Because neither side was aware of, nor paid attention to the mode of extra-chromosomal inheritance (e.g., plasmid, an ***entity*** and its ***properties***), the debate turned out to be a "false dichotomy" (Creager, 2007). This knowledge deficit issue, which we term "nescience in theory formation", appears to be a recurring theme in the knowledge synthesis process aiming either at explaining (understanding) a natural phenomenon or inventing a method. In a sense, the Jackson et al. recombinant DNA method may have suffered a minor nescience, i.e., not knowing the cohesive ends created by the RI restriction enzyme (a ***condition*** of the cutting ***action***), but fortunately a work-around was devised to demonstrate the powerful notion. The plasmid and restriction enzyme examples show that the EApc framework can pinpoint the category of missing knowledge (nescience) in the knowledge synthesis process. We will mention a few more well-known examples of the "nescience" problems in the Discussion.

5. Discussion

A framework of knowledge has been described based on integrating concepts and notions developed in the fields of philosophy of science, history of science and more recently, in scientometrics. In this paper, we have used this framework to test a method of quantifying knowledge essential for the invention of the recombinant DNA technique. We have also examined the knowledge creation and accumulation process, which supplied the basic knowledge to the above invention.

Can knowledge be quantified? This is a centuries old question but few attempts have been made to tackle it in contemporary science. One exception may be the work of Nicholas Rescher, who coined the term "epistemetrics" (knowledge measurement) in his book by the same name (Rescher, 2006). In his book, Rescher declared that "Epistemetrics is not yet a scholarly specialty", which can provide a perspective for the novelty of the study presented here. Rescher believed that while scientometrics, the discipline interested in measuring scientific information, is a centerpiece of epistemetrics, the latter covers knowledge beyond those generated by science. Rescher described a number of epistemetric principles derived from analysis of works in history of science, philosophy of science and cognitive science, providing insights into the quantitative attributes of knowledge regarding its structure, complexity and limitations. However, Rescher did not present a knowledge categorization scheme, nor a precise quantification method.

Knowledge can certainly be quantified in more than one way. The current prevailing method in scientometrics relies on counting research publications as an approximation of the quantity of knowledge output from a research field, program, project or an individual. While this approach has been effective, it is at best an approximation, and can hardly offer any precision measurement of the true size and nature of the knowledge content being created and accumulated. By the EApc type of reasoning, scientific publications are complex entities with highly variable properties. This imprecision is probably due to their heterogeneity in purpose, discipline, subject matter, tradition, content quality, journal style and format etc. As a result, any analysis based on published papers is inherently inexact. This problem is recognized in a recent review article (Fortunato et al., 2018)

> "…a fundamental challenge going forward is accounting for undeniable differences in culture, habits, and preferences between different fields and countries. This variation makes some cross-domain insights difficult to appreciate and associated science policies difficult to implement."

Categorization can be seen as a prerequisite for meaningful quantification. For quantification of knowledge, a challenge lies in the lack of a knowledge categorization system that can be applied across scientific disciplines. Without such a system, quantification is not meaningful. Our EApc framework is subject-matter neutral, therefore potentially applicable to a wide range of scientific disciplines, but this claim needs to be further tested. It should be noted that the EApc system can be applied to computational knowledge, in that a variable with a defined range (properties) can be considered as an entity and mathematical operations can be viewed as actions that transform variables into different entities. Furthermore, this category system is relatively simple, i.e., there are only four basic categories, which satisfies "cognitive economy", a principle of categorization (Rosch, 1999 p.190), and should facilitate its application and automation. We realize that the use of this framework requires some discipline-specific expertise, e.g., molecular biology expertise is required for quantifying knowledge in the recombinant DNA invention.

A salient feature of current scientometric methodology is its reliance on citation. A vast literature has demonstrated the power and utility of citation analysis in revealing intellectual structure and value, concept genealogy, collaborative relationship, and much more in the research landscape. An inherent requirement of citation analysis is its time dependency because it can take years for a paper to accumulate citations, which hinders the immediate assessment of the paper. Dependency on citation also excludes the use of this method for the evaluation of uncitable materials, such as grant applications, which also require rapid assessment (in weeks to month). The EApc framework is by nature independent of citation, and therefore can be applied to scientific publications and proposals as soon as they are available. Furthermore, as we showed in the examples mentioned in "Discovery Process of Each Type of Knowledge", the EApc analysis can potentially be applied to ongoing research, prior to publication, helping researchers to understand better the knowledge structure of the study they plan to or currently conduct.

Kuhn wrote: "… normal science is what produces the bricks that scientific research is forever adding to the growing stockpile of scientific knowledge." (Kuhn, 2000) On the other hand, Authur Kornberg, a Nobel Laureate for his contribution to understanding DNA replication, reportedly said at one of his lab meetings: "I am not interested in making bricks that somebody could use to build a castle; I want to build the castle!" (Burgers, 2007). Juxtaposition of the two remarks reveals a fact that scientific researchers contribute to knowledge accumulation in at least two ways, 1) in their stated goals (e.g., cure cancer or understanding DNA replication), and 2) as "brick" makers, unknowingly or reluctantly, for other castle builders; in other words, they are intentional and unintentional contributors.

Our knowledge classification differs from the current model, which is based on research intention/goal classification, i.e., basic (non-mission oriented), translational (mission-oriented) and applied (product development). The mission orientation classification was originally used by the TRACES study (The Illinois Institute of Technology, 1968), and its results showed that non-mission research is performed more in universities and research institutes and that product development research is done in industry. Mission-oriented investigations can be done in both types of research establishments. Our knowledge classification, *basic knowledge*, i.e., entity and action, and their associated properties and conditions (more brick-like), and *compound knowledge*, i.e., theories and methods (more castle-like), is research-goal or intention neutral. Our examination of the recombinant DNA invention suggests that basic knowledge can arise from all three types of research activities according

to the current goal-orientated classification. Tumor virus (SV40) research can be considered translational research in that it aimed at understanding cancer biology (Dulbecco, 1969), and bacterial drug-resistance investigations in hospitals were clinical research with a goal of combating infections, a serious and urgent medical/clinical issue then and now (Watanabe, 1963; Lushniak, 2014). Both lines of research have contributed basic knowledge (i.e., the properties of entities such as SV40 and plasmids) to the recombinant DNA invention. On the other hand, all three types of research activities generated compound knowledge, the bias being that for non-mission-oriented activities, compound knowledge is more of explanatory in nature, and for product development activities, more utilitarian in nature. All three types of research activities generate methods, also a kind of compound knowledge. Thus, the EApc knowledge categorization provides a new perspective for examining the ever-challenging question of how to balance the three types of research activities (i.e., basic, translational and applied, according to the intention/goal classification) in order to maximize the return of financial investment.

According to the EApc framework, basic knowledge is discovered and compound knowledge is synthesized. Niels Bohr said once that "It is wrong to think that the task of Physics is to find out how nature is. Physics concerns what we can say about Nature." (Petersen, 1963). In describing Nature, we may often not have the right vocabulary at hand, be it a noun, verb, adjective or adverb (roughly corresponding to Entity, Action, Property and Condition). We propose the term "nescience in theory" to refer to the missing knowledge in the theory formation and method invention process. It is of particular interest to look into the circumstances when the nescience is unknown to the research community at large, and consider its diverse consequences and its implication in understanding the discovery process and in managing the support of science. In the Analysis section above, an example was given in that the nescience in understanding the extrachromosomal inheritance (*entity* plasmid) led to futile debate on the mechanism of bacterial antibiotic resistance. Another example was mentioned in the TRACES analysis (The Illinois Institute of Technology, 1968) that a nescience in understanding of the chemical nature of reproductive hormones as steroids, a *property*, hindered the advance in contraception development (the Pill) for a decade. Judah Folkman's notion of using anti-angiogenesis as a treatment for cancer contained a nescience (it might not have been an entirely unknown nescience) that the factor (an *entity*) responsible for tumor angiogenesis was not known. This precluded its application in clinical practice for more than three decades (Folkman, 1971; Kim et al., 1993; Ferrara et al., 2004).

Unknown nescience may not only be a hindrance to knowledge synthesis in its formative stage, which can preclude correct theories from being contrived or translated into practice, but can also be embedded in the theories that are the prevailing paradigms of a research field at a time. A historical example is the "phlogiston" theory for combustion, which lacked the understanding of the *entity* oxygen, as analyzed by many science historians (e.g., Thagard, 1992, pp 40–50). A modern biomedical example (among many others) is the shifting of understanding the etiology of Marfan syndrome, which allegedly affected well-known figures such as Abraham Lincoln. Marfan syndrome is considered a monogenic disease because its inheritance follows the Mendelian Law. When the mutation for the Marfan syndrome was mapped to the gene fibrillin 1, encoding an extracellular protein (FBN1) known for its structural role in tissue mechanical strength, the disease was thought to be a structural defect. This was the prevailing paradigm of the disease for about a decade. However, this theory did not explain fully the clinical symptoms of the disease, and subsequent research showed that FBN1 has another non-structural role in connective tissue, namely, it regulates the activity (bioavailability) (a *property*) of transformation growth factor beta (TGF-beta). Marfan mutation impairs this function of the FBN1 (Neptune et al., 2003). This led to a paradigm shift in understanding the disease and its treatment (Brooke et al., 2008). And in this case, the nescience was a lack of understanding of the *property* of FBN1. These instances prompted the speculation that unknown nescience in the prevailing theories are seeds for the future Kuhnian style revolutions (paradigm shifts). In each case analyzed here, the nescience is invariably a type in the basic knowledge class (vs compound). In a broad sense, one could argue that Newtonian laws of classical mechanics contained a nescience, i.e., not knowing atomic and quantum physics, and as a result, classical mechanics could not be applied to movements at the subatomic level. If such a solid and widely applied theory contained a nescience, then an interesting philosophical question is "Do all prevailing theories contain nescience (missing knowledge)?"

In summary, we proposed a framework of knowledge, which we termed EApc, and tested its utility in precision knowledge quantification and in analyzing the knowledge discovery process. The EApc knowledge framework is developed as a tool for analyzing scientific progress and research status. In principle, it can provide a direct assessment of the intrinsic value of a scientific work based on its knowledge content, unlike the current prevailing method, which assesses the value via peers' opinions (citation), whose formation is influenced by many variables. The EApc based method is, therefore, a complement but not a replacement, of the current methodology.

We argue that this framework is:

- Subject matter neutral, in that it can potentially be applied to many research disciplines.
- Research goal neutral, in that it provides an alternative perspective to the current basic-translational-applied research activity classification.
- Citation independent, in that it can be applied to research work immediately after publication and research proposals (un-citable).

In using the EApc framework to analyze the invention and discovery processes of recombinant DNA, we made the following observations:

- Successful innovation may require a selection of entities that maximize the number of useful properties per entity.
- In exploratory research, property/condition discovery often precedes the discovery of entity/action possessing that property/condition.
- Property/Condition guides the transition from conceptual entity/action to physical entity/action.
- Assigning the observed property/condition to conceptual entity/action can be a source of confusion in discovering physical entity/action.
- Nescience in theory is often due to missing basic knowledge (Entity, Action, Property, Condition).

Many areas in the framework need further investigation, revision, refinement and expansion. It is recognized that the definition of *Entity* and *Action* need further clarification, and *Properties* and *Conditions* contain subcategories that need additional definition and description. The structure and hierarchy of Compound Knowledge has been barely touched upon in the present work. We plan to investigate this in a future work, because it may suggest a way to measure level of innovation.

Declarations

Author contribution statement

Hung Tseng: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Henry Small: Analyzed and interpreted the data; Wrote the paper.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing interest statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

Acknowledgements

We wish to thank the anonymous reviewers for their comments and suggestions. Hung Tseng's views expressed here are personal and do not represent those of the NIAMS/NIH. He would like to dedicate this work to the memory of Dr. Steve Katz.

References

Arber, W., 1965. Host-controlled modification of bacteriophage. Annu. Rev. Microbiol. 19, 365–378.
Arber, W., Linn, S., 1969. DNA modification and restriction. Annu. Rev. Biochem. 38, 467–500.
Bertani, G., Weigle, J.J., 1952. Host controlled variation in bacterial viruses '. J. Bacteriol. 65, 113–121.
Bosch, G., 2018. Train PhD students to be thinkers not just specialists. Nature.
Boyer, H.W., Roulland-dussoix, D., 1969. A complementation analysis of the restriction and modification of DNA in Escherichia coli. J. Mol. Biol. 41 (3), 459–472.
Brooke, B.S., Jabashi, J.P., Judge, D.P., Patel, N., Loeys, B., Dietz, Harry C., I., 2008. Angiotensin II blockade and aortic-root dilation in Marfan's syndrome. N. Engl. J. Med. 358 (26), 2787–2795.
Burgers, P., 2007. Arthur Kornberg (1918-2007). Mol. Cell 28 (November 30), 530–532.
Cohen, S.N., Chang, A.C.Y., Boyer, H.W., Helling, R.B., 1973. Construction of biologically functional bacterial plasmids in vitro. Proc. Natl. Acad. Sci. 70 (11), 3240–3244.
Cohen, S.N., Miller, C.A., 1969. Multiple molecular species of circular R-factor DNA isolated from Escherichia coli. Nature 224 (27 December), 1273–1277.
Cozzarelli, N.R., Melechen, N.E., Jovin, T.M., Kornberg, A., 1967. Polynucleotide cellulose as a substrate for a polynucleotied ligase induced by phage T4. Biochem. Biophys. Res. Commun. 28 (4), 578–586.
Creager, A.N.H., 2007. Adaptation or selection? Old issues and new stakes in the postwar debates over bacterial drug resistance. Stud. Hist. Philos. Sci. C Stud. Hist. Philos. Biol. Biomed. Sci. 38 (1), 159–190.
Danna, K., Nathans, D., 1971. Specific cleavage of simian virus 40 DNA by restriction endonuclease of Hemophilus influenzae. Proc. Natl. Acad. Sci. 68 (12), 2913–2917.
de Solla Price, D., 1978. Editorial statements. Scientometrics 1 (1), 3–8.
Dulbecco, R., 1969. Cell transformation by viruses. Science (New York, N.Y.) 166 (3908), 962–968.
Falkow, S., Citarella, R.V., Wohlhieter, J.A., Watanabe, T., 1966. The molecular nature of R factors. J. Mol. Biol. 17, 102–116.
Ferrara, N., Hillan, K.J., Gerber, H.-P., Novotny, W., 2004. Discovery and development of bevacizumab, an anti-VEGF antibody for treating cancer. Nat. Rev. Drug Discov. 3 (01 May), 391–400.
Folkman, J., 1971. Tomor angiogenesis: therapeutic implications. N. Engl. J. Med. 285 (18 November), 1182–1186.
Fortunato, S., Bergstrom, C.T., Barabasi, A.-L., 2018. Science of science. Science.
Garfield, E., 1955. Science citation Index. Library 122 (3159), 108–111.
Garfield, E., 1979. Scientometrics comes of age. Current Comments 46, 5–10.
Gefter, M.L., Becker, A., Hurwitz, J., 1967. The enzymatic repair of DNA. I. Formation of circular lambda-DNA. Proc. Natl. Acad. Sci. U. S. A 58 (1), 240–247.
Gellert, M., 1967. Formation of covalent circles of lambda DNA by E. coli extracts. Proc. Natl. Acad. Sci. U. S. A 57 (1), 148–155.
Goldstein, J., 1999. Emergence as a construct: history and issues. Emergence 1 (1), 49–72.
Goldstein, J.A., 2007. Emergence and identity. Classic Paper: Emergence and Identity 9 (3), 75–96.
Helinski, D.R., 1973. Plasmid determined resistance to antibiotics: molecular properties of R factors. Annu. Rev. Microbiol. 27, 437–470.
Jackson, D.A., Symons, R.H., Berg, P., 1972. Biochemical method for inserting new genetic information into DNA of simian virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of Escherichia coli. Proc. Natl. Acad. Sci. 69 (10), 2904–2909.
Jovin, Thomas M., Englund, Paul T., Kornberg, Authur, 1969. Enzymatic synthesis of deoxyribonucleic acid. J. Biol. Chem. 244 (11), 3009–3018.
Kim, K.J., Li, B., Winer, J., Armanini, M., Gillett, N., Phillips, H.S., Ferrara, N., 1993. Inhibition of vascular endothelial growth factor-indcued angiogenesis suppresses tumour growth in vivo. Nature 362 (29 April), 841–844.
Kuhn, T.S., 2000. What is scientific revolution? In: Conant, J., Haugeland, J. (Eds.), The Road since Structure: Philosophical Essays, 1970-1993, with an Autobiographical Interview (Paperback. The University of Chicago Press, Chicago, p. 13.
Lederberg, E., 1993. The True History of Fertility Factor F: Esther M. Zimmer Lederberg's Response to the Seven Questions. Retrieved from. http://www.esthermlederberg.com/Clark_MemorialVita/HistoryF EML Response5.html#HISTF.
Lederberg, J., 1952. Cell genetics and hereditary symbiosis. Physiol. Rev. 32 (4), 403–430.
Li, Jinghai, Ge, Wei, Kwauk, Mooson, 2009. Meso-scale phenomena from compromise. Arxiv.Org.
Li, Jinghai, Huang, Wenlai, Edwards, Peter P., Kwauk, Mooson, Houghton, John T., Slocombe, Daniel, 2013. On the universality of mesoscience: Science of 'the in-between. Arxiv.Org.
Lobban, P.E., Kaiser, A.D., 1973. Enzymatic end-to end joining of DNA molecules. J. Mol. Biol. 78 (3), 453–471.
Luria, S.E., Human, M.L., 1952. A nonhereditary, host-induced variation of bacterial viruses. J. Bacteriol. 64, 557–569.
Lushniak, B.D., 2014. Antibiotic resistance: a public health crisis. Public Health Rep. 129 (4), 314–316.
Mertz, J.E., Davis, R.W., 1972. Cleavage of DNA by R I restriction endonuclease generates cohesive ends. Proc. Natl. Acad. Sci. U. S. A 69 (11), 3370–3374.
Meselson, M., Weigle, J.J., 1961. Chromosome breakage accompanying genetic recombination in bacteriophage*†. Proc. Natl. Acad. Sci. U. S. A 47 (6), 857–868.
Meselson, M., Yuan, R., 1968. DNA restriction enzyme from E. Coli. Nature 217 (March 23), 1110–1114.
Mingers, J., Leydesdorff, L., 2015. A review of theory and practice in scientometrics. Eur. J. Oper. Res. 246 (1), 1–19.
Neptune, E.R., Frischmeyer, P.A., Arking, D.E., Loretha, M., Bunton, T., Gayraud, B., et al., 2003. Dysregulation of TGF-beta activation contributes to pathogenesis in Marfan syndrom. Nat. Genet. 33 (24 February), 407–411.
Nisioka, T., Mitani, M., 1969. Composite circular forms of R factor deoxyribonucleic acid composite circular forms of R factor. Deoxyribo- nucleic Acid Molecules 97 (I), 376–385.
Olivera, B.M., Lehman, I.R., 1967. Linkage of polynucleotides through phosphodiester bonds by an enzyme from Escherichia coli. Proc. Natl. Acad. Sci. U. S. A 57 (5), 1426–1433.
Petersen, A., 1963. The philosophy of niels bohr. Bull. At. Sci. 19 (7), 8–14.
Rescher, Nicholas, 2006. Epistemetrics. Cambridge University Press, Cambridge.
Rheinberger, H.J., 1992a. Experiment, difference, and writing: I. Tracing protein synthesis. Studies in History and Philosophy of Science 23 (2), 305–331.
Rheinberger, H.J., 1992b. Experiment, difference, and writing: II. The laboratory production of transfer RNA. Studies in History and Philosophy of Science 23 (3), 389–422.
Roberts, R.J., 2005. How restriction enzymes became the workhorses of molecular biology. Proc. Natl. Acad. Sci. 102 (17), 5905–5908.
Rosch, E., 1999. Principles of categorization. In: Margolis, E., Laurence, S. (Eds.), Concepts, Core Readings. The MIT Press, Cambridge, Massachusetts, London, England.
Roth, T.F., Helinski, D.R., 1967. Evidence for circular DNA forms of a bacterial plasmid. Proc. Natl. Acad. Sci. U. S. A 58 (2), 650–657.
Saiki, Randall K., David, H., Gelfand, Susanne Stoffel, Scharf, Stephen J., Russell, Higuchi, Glenn, T., Horn, Mullis, Kary B., Erlisch, Henry A., 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239, 487–491.
Service, R.F., 2012. The next big(ger) thing. Science 335, 1167.
Sgaramella, V., 1972. Enzymatic oligomerization of bacteriophage P22 DNA and of linear simian. Virus 40 DNA 69 (11), 3389–3393.
Silver, R.P., Falkow, S., 1970. Specific labeling and physical characterization of R-factor deoxyribonucleic acid in Escherichia coli. J. Bacteriol. 104 (1), 331–339.
Small, H., Greenlee, E., 1980. Citation context analysis of a co-citation cluster: recombinant-DNA. Scientometrics 2 (4), 277–301.
Smith, H.O., Welcox, K.W., 1970. A Restriction enzyme from Hemophilus influenzae: I. Purification and general properties. J. Mol. Biol. 51 (2), 379–391.
Smith, R., 2018. Aristotle's Logic. Retrieved from. https://plato.standord.edu/entries/aristotle-logic/#cat.
Thagard, Paul, 1992. Conceptual Revolution. Princeton University Press, Princeton, Oxford.
The Illinois Institute of Technology, 1968. Technology in Retrospect and Critical Events in Science (TRACES) (Report prepared for the National Science Foundation under contract NSF-C535.).
Tseng, H., 1999. DNA cloning without restriction enzyme and ligase. BioTechnique 27, 1240–1244.
Watanabe, T., 1963. Infective heredity of multiple drug resistance in bacteria. Bacteriology Reviews 27, 87.
Weiss, B., Richardson, C.C., 1967. Enzymatic breakage and joining of deoxyribonucleic acid, I. Repair of single-strand breaks in DNA by an enzyme system from Escherichia coli infected with T4 bacteriophage. Proc. Natl. Acad. Sci. U. S. A 57 (4), 1021–1028.