Vocabulary sharing among subjects belonging to the hierarchy of sciences

John G. Benjafield

Received: 12 October 2019 / Published online: 17 August 2020 © Akadémiai Kiadó, Budapest, Hungary 2020

Abstract

To what extent do the vocabularies of mathematics, computing, astronomy, physics, chemistry, biology, psychology, sociology, economics, political science, philosophy, and linguistics overlap? To explore this question, samples of the anglophone vocabularies of these subjects were created using the *Oxford English Dictionary* (Benjafield in Scientometrics 118:1051–1064, 2019. https://doi.org/10.1007/s11192-019-03021-2). The first part of this study compared the vocabularies of the five empirical members of Comte’s hierarchy of the sciences (HoS) plus psychology (i.e., astronomy, physics, chemistry, biology, psychology, sociology). The results were generally consistent with the existence of an empirical HoS. For example, each subject shared its vocabulary the most with another subject adjacent to it in the hierarchy (i.e., astronomy with physics, physics with chemistry, biology with chemistry, psychology with biology, sociology with psychology). The second part of this study examined patterns of sharing between mathematics, computing, economics, political science, philosophy, linguistics and the six members of the empirical HoS. Among the most interesting results was the high degree of vocabulary sharing between mathematics, philosophy, and linguistics. Indeed, it turns out that all subjects share their vocabularies with all other subjects, to varying degrees. It was suggested that, in addition to comparing subjects in terms of a linear HoS, similarities between subjects should be examined independently of their position on the HoS.

Keywords Hierarchy of the sciences · Polysemy · Principle of proximity · Vocabulary sharing

Introduction

Comte (1851/1875, p. 26) proposed that that there is a hierarchy of sciences (HoS) beginning with mathematics and followed by astronomy, physics, chemistry, biology, and sociology. This hierarchy was “determined by the…increasing complexity of the phenomena; the more complex being naturally dependent upon those that are more simple.” Of course, a lot hinges on what is meant by *complexity*. Fanelli and Glänzel (2013, p. 1) suggested that
complexity “generally increases with increasing levels of the organization of matter. From subatomic particles to human societies there is an overall increase in the possible number of elements, combinations, interactions, etc.” Viewed from this perspective, a hierarchy of sciences makes sense. Consequently, after languishing for some time, the concept of a hierarchy of sciences has been revived, thanks to the work of Simonton (2004, 2006, 2015, 2018) as well as Fanelli (2010) and Fanelli and Glänzel (2013).

Some readers may feel that the use of the word hierarchy unjustly implies that some sciences are and always will be superior to others. While it is true that hierarchy originally referred to a heavenly order, the word has acquired additional senses over the years. For example, the word hierarchy can be used in a way that does “justice to the diversity of the sciences without thereby losing sight of their unity” (Bourdeau 2018). In what follows we will focus on the HoS and some other well-established subjects, including mathematics, computing, economics, political science, philosophy, and linguistics. In future work, we hope to consider branches of these subjects [e.g., ecology (a branch of biology)] as well as interdisciplinary approaches [e.g., anthropology and archaeology are often considered to be parts of the same whole (Archaeology and Anthropology 2020)] in order to better represent the complexity of the relationships between subjects.

**Vocabulary sharing and the HoS**

We are concerned with the degree to which the members of the HoS share each other’s vocabularies. In the first part of this study, we examine the hypothesis that the extent to which sciences share their vocabularies depends on how close to one another they are in the HoS. In the second part of this study, we explore the degree to which the members of the HoS share their vocabularies with selected subjects belonging to the humanities and social sciences.

Comte omitted psychology from his hierarchy because, in his view, the psychology of his day was not an objective science but dependent upon subjective data obtained through introspection. In Comte’s view, psychology could only have a scientific subject matter as a part of biology. Psychology is included in our study of the HoS because its vocabulary has continued to survive as that of an independent subject. Moreover, other studies (e.g., Simonton 2015) have provided evidence for psychology’s position between biology and sociology. Whether or not psychology will continue to survive as an independent subject or be assimilated by other subjects is a question that will be considered below.

As Fanelli and Glänzel (2013, p. 2) point out, the subject matter of mathematics is different than that of the other members of the HoS in that it is not empirical so much as it is “purely intellectual.” Consequently we will first present the patterns of vocabulary sharing of the five empirical members of the HoS plus psychology. These patterns will then be compared with those of mathematics, computing, political science, economics, philosophy, and linguistics.

**Methods**

**The Oxford English Dictionary**

In what follows we will make extensive use of the *Oxford English Dictionary* (OED) *Online* (2020) in order to provide a quantitative understanding of the degree to which
the subjects mentioned above share their vocabularies. The OED began in 1857 as a “complete reexamination of the language from Anglo-Saxon times onward” (History of the OED 2020). Under the leadership of James H. Murray, the original OED took a keen interest in the vocabulary of science (Berg 1991, p. 1), an emphasis that has continued through successive editions. Scientific vocabulary is not only emphasized in the OED, but “just as the naturalist strives to identify and demarcate the nuances of different species, so was the lexicographer to act as a scientist engaged in classifying the phenomena of language” (Mugglestone 2005, p. 122). For Murray, “lexicography... was emphatically a domain where the scientist must prevail” (Mugglestone 2005, p. 116). Thus both the vocabulary and the methods of science shaped and continue to shape the OED. The OED is perpetually being updated (Rewriting the OED 2020) and provides excellent coverage of the history of the vocabulary of the sciences “as anyone will soon discover on trying to find fault with the editors by sampling current research periodicals. Their crop of untreated words will prove to be a meager gleaning over a very well-harvested field” (Hoare and Salmon 2000, p. 170). A brief description of the process of adding words to the dictionary is given in How are words added to the OED? (2020).

One key to the OED’s success as a research tool is its Reading Programme (2020) which began in 1879 and continues today. The program makes use of a large number of readers who contribute about 200,000 quotations a year illustrating the history and use of words. In 1990, lexicographers began to use computing as well as the Reading Programme to make the OED’s coverage even more complete. Among other sources, the OED currently uses a corpus that “contains over 8 billion words of web-based news content from 2017 to the present day” (Corpus analysis of the language of Covid-19 2020). Thus, coronavirus was added to the dictionary in 1968 and covid-19 in May, 2020. Some idea of the OED’s ongoing role in research can be gleaned from The OED and Research (2020).

An extremely useful feature of the OED is that the entry for each word “is organized into a hierarchy of senses, which include definitions, labels and cited quotations. Subject labels distinguish the subject classification of a sense, for example they signal how a word may be used in Anthropology, Music or Computing” (Langari and Tompa 2001). Thus, the OED is not a stereotypical dictionary that simply lists one or a few meanings for a word. To begin with, an OED entry provides the first known date of entry (DOE) of the word into the language. There then follow specific senses for the word with their dates of entry. For example, gauge entered the language about 1450 meaning a fixed or standard measure. It acquired a meaning in physics (DOE 1926) of any function introduced as an additional term into the equations of the potentials of a field such that the derived equations of observable physical quantities are unaltered by the introduction. There are quotations illustrating the use of each word. For example, for gauge in physics we have the following quotation from Pauli (1940, p. 718).

“By ‘gauge-transformation of the first kind’ we understand a transformation $U \rightarrow U e^{i\alpha}$ $U^* \rightarrow U^* e^{-i\alpha}$ with an arbitrary space and time function $\alpha$. By ‘gauge-transformation of the second kind’ we understand a transformation of the type $\phi_k \rightarrow \phi_k - i(\partial \alpha/\partial x_k)/\varepsilon$ as for those of the electromagnetic potentials.” The foregoing is a simplified description of the entry for a word. The best way to gain an understanding of the complexity of the OED is to log on to OED.com and browse the dictionary.

In what follows, an illustrative quotation will be given in the discussion of a specific word sense if that quotation appears to be particularly informative.
Compilation of vocabularies

For purposes of the present study, earlier versions of the vocabularies of biology, chemistry, computing, linguistics, mathematics, philosophy, physics, and psychology from Benjafield (2012, 2013, 2014, 2016, 2019a) were updated using the OED Online (2020). The vocabularies of astronomy, economics, political science, and sociology were newly assembled for this study, also using the OED Online (2020). All vocabularies were either updated or created in 2019. In addition, our sample includes multiword sequences (Ellis and Ogden 2017, p. 605; Benjafield 2019a, p. 1053) as well as single words. For example, condensed matter is a multiword in both astronomy and physics. In what follows, word refers to both single words and multiwords. Our study applies only to the anglophone vocabularies of subjects, an issue that will be considered later in the paper.

The vocabularies of six empirical subjects belonging to the HoS

The number of words the OED provides for each of the six empirical subjects belonging to the HoS (including psychology) are given in Table 1. Examination of these data suggests that the frequency with which words occur in each subject’s vocabulary can be summarized by two tendencies. The first tendency is due to a subject’s complexity (Fanelli and Glänzel 2013, p. 1). As the complexity of a subject increases, more words become necessary in order to capture its subject matter. Thus, the number of words in a subject increases as we go from astronomy to physics and then to chemistry. The second tendency is historical. The number of words a subject has will be partly determined by how long it has been in existence: the newer a subject is, the fewer words it tends have. Even though they are in principle more complex, psychology and sociology have markedly smaller vocabularies than the other subjects. Thus, there are two opposing tendencies, with biology being caught in the middle. As a result, biology has more words than psychology but fewer than chemistry. This result is consistent with Fanelli and Glänzel’s (2013, p. 332) finding that most of their measurements designed to distinguish between subjects showed that “the biological sciences had values intermediate between those of the physical and the social sciences.”

The size of the vocabularies of chemistry and sociology

The large number of words in chemistry relative to other subjects in the HoS may puzzle some readers. However, as Benjafield (2014, pp. 23–24) observed, to a greater extent than in other subjects the chemistry vocabulary consists of names for material substances
(e.g., benzol, buckytube, copernicium, crystallin, darmstadtium, eucalin, fustin, heptane, hydrazinium, melamine, olefin, picein, roentgenium, xanthophyll) or pieces of apparatus (e.g., absorptiometer, chromatograph, retort, sonicator).

By contrast with the size of chemistry’s vocabulary, the vocabulary of sociology is extremely small. As noted in the previous section, the size of sociology’s vocabulary may at least in part be determined by its relative recency. Nevertheless, the smallness of sociology’s vocabulary seems inconsistent with its role as the most complex subject-matter in the HoS. One possible reason for the size of sociology’s vocabulary may be that it is not necessary to create new words in order to deal with increasing complexity. Rather, the words that characterize the subject matter of a discipline can be given multiple meanings in order to convey multiple aspects of that subject matter (Piantadosi et al. 2012, p. 280; Gibson et al. 2019, p. 383). As shown in the previous paragraph, chemistry needs to create names for particular substances, and such names need to be unambiguous. It would defeat the purpose of naming a substance in chemistry if one gave a word like buckytube multiple meanings.

However, the problem that sociology faces is different. For example a word like society requires a variety of meanings in order to even begin to convey the complexity of society. A subset of the meanings of society given by the OED (abbreviated and without their associated quotations, DOE’s, etc.) is as follows: a political alliance, league, or contract; a system of sharing within a group; partnership in business or some commercial transaction; association or friendly interaction with other people; the system of customs and organization adopted by a group of people for harmonious coexistence or mutual benefit; the social organization of a group of animals of the same species; a particular section of society; the aggregate of fashionable, wealthy, or otherwise prominent people; to appear regularly at private or public entertainments, parties, etc.; an association or body of people united by a common aim, interest, or profession; a corporate body of people having a definite place of residence, as a college, monastery, or similar institution; a group associating for the purpose of worship and religious practice or service; a mutual or cooperative company; a group or organization formed for the purpose of promoting some branch of study or research by means of meetings, publications, etc. There is no single meaning of society.

The foregoing only scratches the surface of the complexity of the phenomena with which sociology must deal. Sociology can have the smallest vocabulary and still be the most complex subject in the HoS, since many of the words that are central to its subject matter have multiple meanings that express the richness of its subject matter. A subject like sociology does not use technical vocabulary to the extent that a subject that makes more extensive use of mathematics, such as physics, does. Rather, sociology relies relatively more on ordinary language, as demonstrated above with respect to society. This may well be a permanent shortcoming from a scientific perspective. Indeed, there are those who argue that complex subjects, such as sociology, can never achieve the degree of precision found in less complex subjects (e.g., Sanbonamatsu and Johnston 2019). This is a topic we will return to in the “Mathematics and computing” section.

Varieties of vocabulary sharing

There are different ways in which a word may be shared across subjects. The simplest is for a word to be introduced simultaneously in more than one subject. For example, quagma, meaning a plasma consisting of free quarks and gluons, has a date-of-entry (DOE) of 1985 in both astronomy and physics. Alternatively, a word may be created for a specific subject and later used by another subject. For example, the first meaning of asexuality was
biological and meant *propagating without sexual reproduction* (DOE 1845). Subsequently, the same word was introduced into psychology and meant *the state of having no sexual feelings or desires* (DOE 1899). Finally, a word may start out with an ordinary language meaning and then acquire new meanings for specific subjects. For example, *poison* (DOE circa 1225) was given a meaning in chemistry of *a substance which reduces or blocks the activity of a catalyst* (DOE 1904). Subsequently, *poison* acquired a sense in physics as *an impurity in a nuclear reactor which absorbs neutrons and thus slows the intended reaction* (DOE 1952). As we shall see, in the majority of cases, shared words are given one sense in one subject and subsequently given a different sense in another subject.

**Vocabulary sharing between subjects**

Figure 1 shows the percentage distributions of the words shared by each member of the empirical HoS with all other members. To begin with, consider the percentage distribution of words shared with astronomy. Physics has the largest percentage of the words that astronomy shares with other subjects, followed by chemistry, biology, psychology and sociology. This order is consistent with a hierarchy of sciences in which the proximity of subjects to one another will determine the extent to which they share each other’s
vocabularies. This principle of proximity also holds reasonably well for the percentage distribution of physics. Chemistry and astronomy are first and second in terms of sharing the vocabulary of physics, followed by biology, psychology, and sociology in that order. In the case of chemistry the principle of proximity is also quite consistent with the data. The two subjects adjacent to chemistry, physics and biology, share chemistry’s vocabulary to the greatest extent, followed by astronomy, psychology, and sociology. However, in the case of biology, the principle of proximity begins to break down. Biology shares its vocabulary firstly with chemistry and secondly with physics. Psychology is third followed by astronomy and finally sociology. Psychology is proximal to biology, and should be the first or second subject to share biology’s vocabulary if proximity was all that mattered. However, psychology comes third after physics. Here, once again, we see a historical tendency at work. This tendency can be stated as follows: the earlier the subject in the hierarchy, the more likely its vocabulary is to be shared with later subjects. Like biology, psychology’s pattern shows the influence of earlier subjects in that its vocabulary is shared the most with biology, chemistry, and physics, in that order, followed by sociology and finally astronomy. This is a pattern that Comte might have predicted for psychology, as will be discussed below. Finally, sociology has a distribution that is almost the mirror image of astronomy, which is at the opposite end of the hierarchy. Sociology shares the most words with psychology, followed by biology, physics, chemistry, and astronomy. The principle of proximity accounts for sociology’s pattern of sharing in much the same way as it does for astronomy’s. Notably, almost half of the words that sociology shares with other subjects belongs to psychology, a greater percentage than that between any other two subjects. The relationship between psychology and sociology will be explored further below.

Pairs of subjects that share the most words

Examination of Fig. 1 shows that every subject shares the most with a subject that is one of its nearest neighbors. Astronomy shares more of its vocabulary with physics than with any other subject. Physics shares the most with chemistry as chemistry does with physics. Biology shares the most with chemistry, psychology shares the most with biology, and sociology shares the most with psychology.

Astronomy and physics

Typically, words shared by two different subjects have different meanings in each subject (i.e., are polysemous). For example, of the 178 words shared by astronomy and physics, only 34 or 19% (95% CI = 13%, 25%) have the same meaning in both subjects [e.g., blueward (DOE 1877)], meaning towards the blue end of the visible spectrum. Each of the remaining 144 words has a different sense in astronomy than it does in physics. An example is hot spot for which the sense in astronomy is a region in the sun’s corona that is at a higher temperature than normal (DOE 1936), while the sense in physics is a local concentration of radioactivity (DOE 1955). Of the 144 words that have different senses in each subject, 60% (95% CI = 52%, 68%) have an earlier date of entry in physics than in astronomy. This suggests that there is a slight tendency for astronomy to borrow more words from physics than vice versa. However, when borrowing a word from the other subject, rather than simply adopting the sense of that word from the other subject, each subject tends to tailor the meaning of the word to suit its own needs. This is a widely observed phenomenon in language generally. As a word acquires additional senses it becomes more
ambiguous. However, it also more useful in the sense of having more applications in different contexts. “Ambiguity allows for greater ease of processing by permitting efficient linguistic units to be re-used” (Piantadosi et al. 2012, p. 280; Gibson et al. 2019, p. 383).

Physics and chemistry

The greatest percentage of physics’ shared words is with chemistry and the greatest percentage of chemistry’s shared words is with physics. They are the only pair of subjects for which this symmetrical pattern of sharing is true. Of the 259 words shared between physics and chemistry, 99 or 38% (95% CI = 32%, 44%) have the same meaning (e.g., homonuclear (DOE 1930), a molecule composed of atoms whose nuclei are alike, i.e. atoms of the same element or (more strictly) the same isotope). Of the remaining 160 shared words, 61% (95% CI = 53%, 69%) have an earlier date of entry for chemistry than for physics. For example, in chemistry lepton means one of the ultimate units of matter in a crystalline material (DOE 1921). In physics lepton is defined as a subatomic particle having a mass less than that of a nucleon (DOE 1948). Quotation: “The most elegant and sparse theory hypothesizes that Higgs bosons are bound states of very heavy quarks or leptons” (Smolin 2006, p. 72). The physics/chemistry distribution of word sharing slightly favors chemistry as the donor, just as the astronomy/physics distribution favored physics. Why one subject of a pair is the dominant lender is a question we will take up below. In general, words shared between subjects are similar to what are called loan-words (DOE 1874), meaning words adopted or borrowed from another language.

Biology and chemistry

Biology shares 234 words with chemistry. Twenty-two percent (95% CI = 17%, 27%) of these shared words have the same meaning [e.g., bilayer (DOE 1962), a layer of film two molecules thick]. The remaining 182 shared words are polysemous. Of these, 55% (95% CI = 48%, 62%) are words for which the chemistry sense entered earliest. For example, in chemistry entrainment means the separation of one or more components of an azeotropic mixture by the addition of a substance with which they form a new azeotrope (DOE 1936). In biology, entrainment means the conforming of a biological rhythm to the frequency of a periodic stimulus (DOE 1957).

Psychology and biology

Psychology shares 100 words with biology, of which 8 have the same meaning (e.g., stressor (DOE 1950), a single condition or agent that constitutes a stress for an organism). Of the remaining 92 words, 66% (95% CI = 56%, 76%) have an earlier meaning in biology than in psychology. A good example is convergent (DOE 1728), which originally meant inclining toward each other, or toward a common point of meeting; tending to meet in a point or focus. It is defined in biology as of similar structure but of different origin Quotation: “When man selects the offspring of two distinct species, he sometimes induces a considerable amount of convergence. This is the case with the improved breeds of pigs, which are descended from two distinct species. In the case of the convergent pigs above referred to, evidence of their descent from two primitive stocks is retained” (Darwin 1871, p. 231). The sense of convergent in psychology (DOE 1956) is thinking…that tends towards only one answer to a problem. Quotation: “Several studies
suggest that science students perform well with reasoning or numerical problems which require convergent thinking towards a single correct solution” (Rump and Dunn 1971).

Convergent is a good example of the fact that the OED does not merely give the most common meaning for a word, but also specifies the various senses of a word as it is used in different contexts. The OED does not pretend that every word has one essential meaning.

In general, psychology depends more on biology for its vocabulary than vice versa. This result is consistent with Simonton’s (2015, p. 339) finding that psychology’s position in the HoS is “most proximate to biology.”

**Sociology and psychology**

Sociology shares 102 words with other subjects, 48% (95% CI = 34%, 62%) of these words with psychology. This is the largest shared percentage between any two subjects. Moreover, 59% (95% CI = 45%, 73%) of the shared words between psychology and sociology have the same meaning. This is the highest percentage of shared words with the same meaning of any of the pairs of subjects we have considered. An example of a shared word between psychology and sociology is *Guttman scale* (DOE 1950) meaning a type of scale used to measure and assess mental attitudes and properties. Quotation: “A cumulative scale designed…to determine whether the items can be arranged along a single continuum of increasing intensity. When a series of items (attitudes) forms a Guttman scale, acceptance of any item on the scale indicates acceptance of all the items below it” (Theodorson and Theodorson 1970, p. 181).

Only 20 words have different dates of entry, of which 70% (95% CI = 49%, 91%) have an earlier date of entry in psychology. An example is *diffusion* (DOE 1859) which has a sense in psychology as the dissemination of impulses or excitation through the nervous system; the arousal of a widespread response by a stimulus. Its meaning in sociology is the spread of new concepts and products within an economy or society; the process whereby innovations are accepted by the market (DOE 1890).

Since sociology’s vocabulary is relatively small compared with that of psychology, the words shared between the two subjects make up a greater percentage of sociology’s vocabulary than they do of psychology’s vocabulary, in which they only constitute 16% (95% CI = 12%, 20%). Thus, their shared vocabularies may well have played a larger role in sociology than in psychology. “In the area of social psychology, sociologists have drawn more frequently from psychologists than the reverse” (Thoits 1995, p. 1231).

**A common pattern**

The pairs of subjects we have just considered have a common pattern. It is always the subject with the largest vocabulary, as shown in Table 1, that loans more words to the other subject than vice versa. Which subject is the most frequent lender is determined by which subject has the greatest number of word senses that entered the language before the other subject’s senses for the same words. Thus physics loans more to astronomy than vice versa, chemistry more to physics, chemistry more to biology, biology more to psychology, and psychology more to sociology. In every case, a subject borrows most heavily from another subject adjacent to it in the HoS.
Is psychology necessary?

For the most part our sharing data are consistent with the hypothesis of a HoS. The psychology pattern of sharing is the most inconsistent. Recall that Comte argued that psychology was not an independent subject properly so-called, but a part of biology (Heyd 1989). As it happens, the so-called “neuroscience revolution” (Benjafield 2019b; Tracy et al. 2004) has aligned psychology more and more closely with biology (David 2008). As we have seen, psychology shares its vocabulary with biology more than with any other subject, a result that is consistent with Simonton’s (2015, p. 339) finding that psychology’s position in the HoS is “most proximate to biology.” A significant part of psychology may be transitioning toward becoming a part of biology. Such a tendency is encouraged by the fact that major anglophone granting agencies favor subjects perceived as sciences as opposed to social sciences or humanities (Sonnert 2018, pp. 185–189). This provides an incentive for anglophone psychologists to adopt biological research methods, thus leaving behind those who are content to continue to use methods belonging to the social sciences or humanities. The outcome may be that those who work in psychology departments will increasingly be seen as belonging to a fragmented discipline (Balietti et al. 2015; Benjafield 2017) that occupies the space between biology and sociology, but not as a coherent subject. Interestingly, the pressure for psychology to become a biological science may not be as pronounced in, for example, the German version of psychology, because the “institutional landscape of research funding is quite different in Germany” (Sonnert 2018, p. 189) than it is in anglophone countries. The implication is that psychology’s position in the HoS may vary depending on the linguistic and cultural context.

Patterns of vocabulary sharing between mathematics, computing, economics, political science, philosophy, linguistics, and the HoS

This section explores the relationships between mathematics, computing, economics, political science, philosophy and linguistics, both with each other and with the empirical HoS. Mathematics was part of Comte’s original conception of the HoS. Computing has had a profound effect on a very wide range of subjects (Haigh 2013). Philosophy is connected to every member of the HoS both historically and through the philosophy of science (McLevey et al. 2018). Linguistics and economics are often represented as being sciences (e.g., The science of linguistics 2019; Shiller 2013), and the ambition behind naming a subject political science speaks for itself.
Table 2 shows the number of words the OED provides for each of mathematics, computing, political science, economics, philosophy, and linguistics. The two oldest subjects, mathematics and philosophy, have comparatively large vocabularies, coming first and third respectively. However, it is important to note that computing has the second largest vocabulary, a measure of its enormous impact over a relatively short time.

Figure 2 shows the percentage distributions of the words shared by mathematics, computing, economics, political science, philosophy, and linguistics with each other and with the members of the empirical HoS.

Mathematics and computing

The percentage distribution for mathematics shows that 56% (95% CI = 54%, 59%) of the vocabulary that mathematics shares with the subjects in Fig. 2 is shared with astronomy, physics, chemistry, and biology combined. This illustrates the centrality of mathematics to the core of the HoS. Physics is the subject with which mathematics shares the most, although only 55 or 24% of their shared words have the same sense (e.g., *Penrose process* (DOE 1970), meaning *a mechanism postulated by Penrose whereby energy can under certain
circumstances escape from a black hole). The remaining 176 words have different meanings in each subject. For example, In mathematics, gamma (DOE 1834) refers to a function (symbol \( \Gamma \)) which extends the notion of factorial \( n \) (written \( n! \)) from positive whole numbers to real and complex variables, given by \( \Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt \). In physics gamma (DOE 1903) means a unit of magnetic field strength, equal to \( 10^{-5} \) oersted (approx. \( 7.96 \times 10^{-4} \) A per metre).

At the other end of the scale are psychology, sociology, economics and political science, together having only 13% (95% CI = 11%, 15%) of mathematics’ shared vocabulary. These subjects are often regarded as relatively “soft,” meaning that their subject matter is more complex and their findings less reliable than “hard” HoS subjects (Fanelli 2010, 2019, p. 27). As was suggested above in the section on the size of the vocabularies of chemistry and sociology, such subjects may only be able to apply the precision afforded by mathematics to a part of their subject matter.

The percentage distribution for mathematics also shows that philosophy and linguistics occupy intermediate positions between these two groups of subjects. The relationship between mathematics, philosophy, and linguistics will be considered further below.

The percentage distribution for computing shows that 65% (95% CI = 62%, 69%) of the vocabulary that computing shares with the subjects in Fig. 2 is shared with mathematics and the core members of the HoS. However, it is worth noting that economics shares more with computing than does astronomy and that psychology is tied with astronomy. This range of influence demonstrates yet again how rapidly computing has become important to a wide variety of subjects. The rapid rise of computing and its high degree of sharing shows how labile disciplinary affinities can be.

Mathematics is the subject with which computing shares the most. Only 33 (22%) of the words shared by computing and mathematics have the same meaning. A well-known example of the vocabularies of both computing and mathematics is Church-Turing (DOE 1961) which designates the proposition that any function that can be computed using an algorithm can be computed by a Turing machine. The remaining 129 (78%) of words shared between computing and mathematics have different meanings in each subject. Thus, in mathematics kernel (DOE 1946) is The set of all the elements that are mapped by a given homomorphism into the identity element (for the group operation in the case of groups, for addition in the case of rings. In computing, kernel (DOE 1972) is The most basic level or core of an operating system, which allocates system resources to the other parts of the operating system and to the programs that run under it, may manage files and memory, and is responsible for low-level hardware interfaces.

**Economics and political science**

Economics shares much of its vocabulary with the “hard” end of the HoS. The first four subjects in the HoS plus mathematics account for 51% (95% CI = 45%, 57%) of economics shared vocabulary. Economics shares more words with mathematics than with any other subject. However, only 4 of the 40 shared words have the same meaning in both mathematics and economics. One of these is Nash (DOE 1952), a term from game theory that designates “a non-cooperative set of strategies in which each player adopts a strategy to maximize his or her own advantage, given the strategies adopted by the other players.”

One of the words that does not have the same meaning in both mathematics and economics is rational (DOE 1398), which initially meant having the faculty of reasoning; endowed with reason. The earliest mathematical sense of rational (DOE 1570) is a number able to be expressed as a ratio of two whole numbers. In economics, the sense of
rational (DOE 1915) is any of various methods of analysis or planning based on the calculation of a projected result. Some readers may wonder if the various methods of economists are really rational, in the original sense of the term (James et al. 2001).

It is worth noting that psychology and political science have the same percentage share of the vocabulary of economics as does chemistry, and more than astronomy. Thus, to some extent at least, economics shares its vocabulary quite widely.

Like economics, political science also shares much of its vocabulary with the “hard” end of the HoS. Indeed, mathematics, astronomy, physics, chemistry, and biology together account for 54% (95% CI = 49%, 60%) of political science’s shared vocabulary. Also like economics, political science shares the most with mathematics, but none of the 43 shared words have the same meaning. For example, the mathematical sense of triangulation (DOE 1940) means the process of dividing up a [topological] space into pieces which are homeomorphic with the interior of a triangle, while the sense of triangulation (DOE 1995) in political science is a process of positioning oneself politically between traditional left-wing and right-wing standpoints.

In the case of political science, none of the remaining subjects score as highly as any of the members of the HoS, with the exception of computing. Political scientists take a strong interest in statistical analyses of their subject matter (e.g., King 1986; Riba 1996), and thus would inevitably be drawn to computing. “Political scientists were unique among the emergent social disciplines in using ‘science’ in their chosen name [and they] looked to natural science as a model—either to emulate or against which to pattern their own methodologies and cultural authority” (Farr 2003, pp. 306–307).

**Philosophy and linguistics**

Philosophy has a complex distribution of sharing, as befits a discipline with such a long history. Psychology, mathematics, and linguistics are the three vocabularies that have the most in common with philosophy. Together they have 50% (95% CI = 46%, 54%) of philosophy’s shared vocabulary. However, psychology and philosophy share more than any other two subjects in this distribution. The similarity of the vocabularies of psychology and philosophy is unsurprising since psychology was considered by many to be a part of philosophy until the latter part of the nineteenth century (Green et al. 2001; Vidal 2011). Fifty-one of 109 shared words have the same meaning. A good example is consciousness (DOE 1678), one definition of which is the faculty or capacity from which awareness of thought, feeling, and volition and of the external world arises. However, the concept of consciousness has acquired many different meanings over the years, and continues to be problematic for both subjects (e.g., Klein 2020). The remaining 58 words have different meanings. Animus has meant the rational soul; mind, will [or] spirit in philosophy (DOE Old English), with a more contemporary psychological sense being C. G. Jung’s term for an archetype present in a man or woman from which the male aspects of the personality are derived (DOE 1923).

The relationships between mathematics, linguistics, and philosophy are more complex than that between psychology and philosophy. This pattern of sharing occurs because some of the most influential anglophone philosophers also concerned themselves with the philosophy of mathematics (e.g., Whitehead and Russell 1910, 1912, 1913; Wittgenstein 1956\(^1\)).

\(^1\) Wittgenstein wrote in German. However, translations of his work have been extraordinarily influential among anglophone philosophers.
For example, Russellian (DOE 1915) means characteristic of Bertrand Russell or his work in mathematics and philosophy. Moreover, language has also been a preoccupation of many influential anglophone philosophers (e.g., Austin 1962; Grice 1968; Rorty 1992).

The percentage distribution for linguistics shows that the majority (59%, 95% CI = 55%, 63%) of its shared vocabulary is with mathematics, biology, computing, and philosophy, with mathematics having the largest share. Of 125 shared words only 6 have the same meaning. For example, in both in mathematics and linguistics isomorphism (DOE 1892) means an identity of form and of operations between two or more groups or other sets; an exact correspondence as regards the number of constituent elements and the relations between them. Quotation: “The basic idea of Wittgenstein’s picture theory is the idea of an isomorphism obtaining between language and reality, a sort...” (Hintikka 1973, p. 42). Primitive is an example of the 119 words with different meanings. In mathematics, it means a geometrical figure or algebraic expression, equation, operation, etc.: from which another is in some way derived, or which is not itself derived from another (DOE 1728), while in linguistics the meaning of primitive is a parent language in an early, unrecorded, or reconstructed stage of its development into a group of dialects or languages (DOE 1878).

The presence of biology among the subjects that share the most with linguistics suggests that, like psychology, linguistics may be transitioning toward becoming a biological science (Boeckx and Piattelli-Palmarini 2005).

Why do subjects share their vocabularies?

The percentage distributions in Figs. 1 and 2 are notable partly because every subject shares its vocabulary with every other subject. No subject is ever completely left out of another subject’s sharing pattern. To some degree, this result is likely due to the universal tendency to reuse existing words in new contexts, if at all possible (Piantadosi et al. 2012, p. 280; Gibson et al. 2019, p. 383). This tendency was mentioned briefly earlier (e.g., in the “Astronomy and physics” section), but its importance should be obvious after examining all the sharing distributions in both Figs. 1 and 2.

As we have seen, some of the words shared between subjects have the same meaning in both subjects. However, in other cases the same word is given a different meaning in different subjects. The latter tendency may at first appear to create difficulties by making words more ambiguous. However, there is a growing body of evidence suggesting that such ambiguity need not make communication more difficult (Rodd 2020). Readers or listeners make use of the context in which a word is used in order to disambiguate it. Peloquin et al. (2020, p. 439) explore the possibility that “ambiguity is an efficient property of any communication system in which communication is contextualized. Simply put, it is useful to have a language that reuses low-cost material (has ambiguity) so long as the cost of disambiguating the material is low. In particular, context (or common ground) can provide useful information for disambiguation.” For example, one should not be surprised if the meaning of a word is different in a political science journal article than it is in a mathematics journal article. If (for whatever reason) one habitually reads both types of journal, one is probably able to disambiguate the way each subject uses the word conservative without much effort. That being said, it is of course the case that disambiguation will take more effort the less one knows about the two different contexts.

Finally, it is possible to give an information-theoretic (Shannon 1948; Fanelli 2019, pp. 2–4) account of why an efficient communication system can also have ambiguity as one
of its properties. See Fortuny and Corominas-Murtra (2013), Gibson et al. (2019), Piantadosi et al. (2012), and Solé and Seoane (2015) for details of possible information-theoretic explanations.

**Which subjects share the most?**

Although the vocabulary sharing distributions show that sharing is spread across all subjects, it is still the case that some subjects do more sharing than others. For example, there is a Spearman correlation of .93 between a subject’s vocabulary size and the number of times its words are shared with all other subjects. The more words a subject has, the more times its words are shared. Now consider the chart shown in Fig. 3, in which subjects are ordered from those which share the most with all other subjects to those which share the least. The four subjects that share the most with all other subjects (mathematics, physics, chemistry, and biology) do roughly half (51.5%) of the total amount of sharing. In other words, mathematics and the three other subjects that have the largest vocabularies in the HoS are also the subjects sharing their vocabularies the most with other subjects. These data support Fanelli and Glänzel’s (2013, p. 1) suggestion that the HoS provides an excellent context within which to explore the diversity of subjects.

**Summary and conclusion**

On the whole, our data are consistent with the existence of a hierarchy of sciences. The anchors of the hierarchy, astronomy and sociology, both have similar patterns of sharing to those that would be predicted on the basis of the HoS. The sharing patterns for physics and chemistry are also consistent with the HoS, in that they share the most with subjects that are adjacent to them in the HoS. Biology shares the most with one of its nearest neighbors, chemistry. Its other nearest neighbor, psychology, is third after physics. Here we see a preference on the part of biology for the vocabulary of an established member of the HoS over a subject that Comte did not regard as an autonomous subject in any case.

Our data suggest that much of what we now call psychology may end up being part of biology, in line with what Comte would have recommended. The remainder may coalesce around the study of aspects of the person that are not easily reducible to biology and consequently may develop stronger affiliations with the humanities (e.g., Smith 2019; Teo 2017; Vidal and Ortega 2017).

Political science and economics are partial to sharing the vocabularies of established sciences. However, unlike many psychologists, workers in these fields do not seem to be
in as much of a hurry to become biological scientists. Nevertheless, there are those who have argued that economics and political science would benefit by taking a more biological approach (e.g., Wiegele 1982/2018).

The sharing patterns for mathematics, philosophy, and linguistics show how intertwined these subjects are, at least in terms of their anglophone vocabulary. It is no doubt possible to assimilate these three subjects to the linear model represented by the HoS. However, it might also be interesting to further explore their interrelations independently of the HoS. Indeed, the same could be said for almost any combination of subjects, since they all share one another’s vocabulary to some extent. This last fact is notable because it shows that every subject is connected to every other subject, even if only to a small degree.

Acknowledgements

Jeremy Burman and two anonymous reviewers provided helpful suggestions concerning the manuscript.

References

Archaeology and Anthropology. (2020). Retrieved June 29, 2020, from http://www.ox.ac.uk/admissions/undergraduate/courses-listing/archaeology-and-anthropology.

Austin, J. L. (1962). How to do things with words. Oxford: Oxford University Press.

Balietti, S., Mäss, M., & Helbing, D. (2015). On disciplinary fragmentation and scientific progress. PLoS ONE, 10, e0118747. https://doi.org/10.1371/journal.pone.0118747.

Benjafield, J. G. (2012). The long past and short history of the vocabulary of Anglophone psychology. History of Psychology, 15, 50–71. https://doi.org/10.1037/a0023386.

Benjafield, J. G. (2013). The vocabulary of Anglophone psychology in the context of other subjects. History of Psychology, 16, 36–56. https://doi.org/10.1037/a0030532.

Benjafield, J. G. (2014). Patterns of similarity and difference between the vocabularies of psychology and other subjects. History of Psychology, 17, 19–35. https://doi.org/10.1037/a0035386.

Benjafield, J. G. (2016). The digital history of the anglophone vocabulary of psychology: An exploration using Zipfian methods. History of Psychology, 19, 125–140. https://doi.org/10.1037/hop0000026.

Benjafield, J. G. (2017). Between pink noise and white noise: A digital history of The American Journal of Psychology and Psychological Review. The American Journal of Psychology, 130, 505–519. https://doi.org/10.5406/amerjpsy.130.4.0505.

Benjafield, J. G. (2019a). Keyword frequencies in anglophone psychology. Scientometrics, 118, 1051–1064. https://doi.org/10.1007/s11192-019-03021-2.

Benjafield, J. G. (2019b). Major paradigms and approaches in psychology. In R. J. Sternberg & W. E. Pckren (Eds.), The Cambridge handbook of the intellectual history of psychology (pp. 4–28). Cambridge: Cambridge University Press.

Berg, D. L. (1991). The user’s guide to the Oxford English Dictionary. Oxford: Oxford University Press.

Boeckx, C., & Piattielli-Palmarini, M. (2005). Language as a natural object—linguistics as a natural science. The Linguistic Review, 22, 447–466.

Bourdeau, M. Auguste Comte. The Stanford encyclopedia of philosophy (E. N. Zalta (Ed.)). Retrieved June 22, 2020, from https://plato.stanford.edu/archives/sum2018/entries/comte/.

Comte, A. (1875). System of positive polity (J. H. Bridges, Trans.). London: Longmans, Green. (Original work published 1851)

Corpus analysis of the language of Covid-19. (2020). Retrieved June 28, 2020, from https://public.oed.com/blog/corpus-analysis-of-the-language-of-covid-19/.

Darwin, C. (1871). The descent of man, and selection in relation to sex. London: J. Murray.

David, G. (2008). Psychology departments are changing their behavior. Chronicle of Higher Education, 55(18), A1–A11.

Ellis, N. C., & Ogden, D. C. (2017). Thinking about multiword constructions: Usage-based approaches to acquisition and processing. Topics in Cognitive Science, 9, 604–620. https://doi.org/10.1111/tops.12256.

Fanelli, D. (2010). “Positive” results increase down the hierarchy of the sciences. PLoS ONE, 5, e10068. https://doi.org/10.1371/journal.pone.0010068.
Fanelli, D. (2019). A theory and methodology to quantify knowledge. *Royal Society Open Science, 6*, 181055. https://doi.org/10.1098/rsos.181055.

Fanelli, D., & Glänzel, W. (2013). Bibliometric evidence for a hierarchy of the sciences. *PLoS ONE, 8*, e66938. https://doi.org/10.1371/journal.pone.0066938.

Farr, J. (2003). Political science. In D. M. Porter & D. Ross (Eds.), *The Cambridge history of science: The modern social sciences* (pp. 306–328). Cambridge: Cambridge University Press.

Fortuny, J., & Corominas-Murtra, B. (2013). On the origin of ambiguity in efficient communication. *Journal of Logic, Language and Information, 22*, 249–267. https://doi.org/10.1007/s1084-013-9179-3.

Gibson, E., Futrell, R., Piantadosi, S. T., Dautriche, I., Mahowald, K., Bergen, L., et al. (2019). How efficiency shapes human language. *Trends in Cognitive Sciences, 23*, 389–407. https://doi.org/10.1016/j.tics.2019.02.003.

Green, C. D., Shore, M., & Teo, T. (Eds.). (2001). *The transformation of psychology: Influences of 19th-century philosophy, technology, and natural science* (pp. 107–131). Washington, DC: American Psychological Association.

Grice, H. P. (1968). Utterer’s meaning, sentence-meaning, and word-meaning. *Foundations of Language, 4*, 225–242.

Haigh, T. (2013). Historical reflections. *Communications of the ACM, 56*, 37–40. https://doi.org/10.1145/2398356.2398369.

Heyd, T. (1989). Mill and Comte on psychology. *Journal of the History of the Behavioral Sciences, 25*, 125–138. https://doi.org/10.1002/1520-6696(198904)25:2%3c125:AID-JHBS230025203%3e3.0.CO.

Hintikka, J. (1973). Logic, language-games, and information. Oxford: Clarendon Press.

History of the OED. (2020). Retrieved June 25, 2020, from https://public.oed.com/blog/page-tags/history-of-the-oed/.

Hoare, M. R., & Salmon, V. (2000). The vocabulary of science in the OED. In L. Mugglestone (Ed.), *Lexicography and the OED: Pioneers in the untrodden forest* (pp. 166–181). New York: Oxford University Press.

How are words added to the OED? Retrieved June 17, 2020, from https://public.oed.com/how-words-enter-the-OED/graphic/.

James, T., Soroka, L., & Benjafield, J. G. (2001). Are economists rational or just different? *Social Behavior and Personality, 29*, 359–364. https://doi.org/10.2224/sbp.2001.29.4.359.

King, G. (1986). How not to lie with statistics: Avoiding common mistakes in quantitative political science. *American Journal of Political Science, 30*, 666–687.

Klein, S. B. (2020). Thoughts on the scientific study of phenomenal consciousness. *Psychology of Consciousness: Theory, Research, and Practice*. https://doi.org/10.1037/cns0000231.

Langari, A., & Tompa, F. W. (2001). Subject classification in the Oxford English Dictionary. In N. Cercone, L. Tsau Young, & W. Xindong (Eds.), *Proceedings of the 2001 IEEE international conference on data mining* (pp. 329–336). San Jose, CA. Retrieved September 24, 2019, from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=989536.

McLevey, J., Graham, A. V., McIlroy-Young, R., Browne, P., & Plaisance, K. S. (2018). Interdisciplinarity and insularity in the diffusion of knowledge: An analysis of disciplinary boundaries between philosophy of science and the sciences. *Scientometrics, 117*, 331–349. https://doi.org/10.1007/s11192-018-2866-8.

Mugglestone, L. (2005). *Lost for words: The hidden history of the Oxford English Dictionary*. New Haven, CT: Yale University Press.

Oxford English Dictionary Online. (2020). Retrieved June 22, 2020, from www.oed.com.

Pauli, W. (1940). The connection between spin and statistics. *Physical Review, 58*, 716–722.

Peloquin, B. N., Goodman, N. D., & Frank, M. C. (2020). The interactions of rational, pragmatic agents lead to efficient language structure and use. *Topics in Cognitive Science, 12*, 433–445. https://doi.org/10.1111/tops.12489.

Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition, 122*, 280–291. https://doi.org/10.1016/j.cognition.2011.10.004.

Reading programme. (2020). Retrieved June 25, 2020, from https://public.oed.com/blog/reading-progr-ammes-the-art-of-reading-for-the-oed/.

Rewriting the OED. (2020). Retrieved June 3, 2020, from https://public.oed.com/history/rewriting-the-oed/.

Riba, C. (1996). The use of mathematics in political science: A survey of European and American journals. *European Journal of Political Research, 29*, 477–508. https://doi.org/10.1111/j.1475-6765.1996.tb00663.x.

Rodd, J. M. (2020). Settling into semantic space: An ambiguity-focused account of word-meaning access. *Perspectives on Psychological Science, 15*, 411–427. https://doi.org/10.1177/1745698885860.
Rorty, R. (Ed.). (1992). *The linguistic turn: Essays in philosophical method*. Chicago, IL: University of Chicago Press.

Rump, E., & Dunn, M. (1971). Extensions to the study of science students’ divergent thinking ability. *Nature*, 229, 349–350. https://doi.org/10.1038/229349b0.

Sanbonmatsu, D. M., & Johnston, W. A. (2019). Redefining science: The impact of complexity on theory development in social and behavioral research. *Perspectives on Psychological Science*, 14, 672–690. https://doi.org/10.1177/1745691619848688.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(379–423), 623–656.

Shiller, R. J. (2013). Is economics a science? *Shiller Feeds*. Retrieved September 3, 2019, from http://www.shillerfeeds.com/2013/11/is-economics-science.html.

Simonton, D. K. (2004). Psychology’s status as a scientific discipline: Its empirical placement within an implicit hierarchy of the sciences. *Review of General Psychology*, 8, 59–67. https://doi.org/10.1037/1089-2680.8.1.59.

Simonton, D. K. (2006). Scientific status of disciplines, individuals, and ideas: Empirical analyses of the potential impact of theory. *Review of General Psychology*, 10, 98–112. https://doi.org/10.1037/1089-2680.10.2.98.

Simonton, D. K. (2015). Psychology as a science within Comte’s hypothesized hierarchy: Empirical investigations and conceptual implications. *Review of General Psychology*, 19, 334–344. https://doi.org/10.1037/gpr0000039.

Simonton, D. K. (2018). Hard science, soft science, and pseudoscience: Implications of research on the hierarchy of the sciences. In A. B. Kaufman & J. C. Kaufman (Eds.), *Pseudoscience: The conspiracy against science* (pp. 77–99). Cambridge, MA: MIT Press.

Smith, R. (2019). Resisting neurosciences and sustaining history. *History of the Human Sciences*, 32, 9–22. https://doi.org/10.1177/0952695118810286.

Smolin, L. (2006). *The trouble with physics: The rise of string theory, the fall of a science, and what comes next*. Boston, MA: Houghton Mifflin.

Solé, R. V., & Seoane, L. F. (2015). Ambiguity in language networks. *The Linguistic Review*, 32, 5–35. https://doi.org/10.1515/tlr-2014-0014.

Sonnett, G. (2018). Social science and sozialwissenschaft: Categorical and institutional boundaries of knowledge. *Journal of the History of the Behavioral Sciences*, 54, 178–197. https://doi.org/10.1002/jhbs.21909.

Teo, T. (2017). From psychological science to the psychological humanities: Building a general theory of subjectivity. *Review of General Psychology*, 21, 281–291. https://doi.org/10.1037/gpr0000132.

*The OED and research*. (2020). Retrieved June 17, 2020, from https://public.oed.com/the-oed-and-research/.

*The science of linguistics*. (2019). Retrieved September 3, 2019, from https://www.linguisticsociety.org/resource/science-linguistics.

Theodorson, G. A., & Theodorson, A. G. (1970). *A modern dictionary of sociology*. London: Methuen.

Thoits, P. A. (1995). Social psychology: The interplay between sociology and psychology. *Social Forces*, 73, 1231–1243. https://doi.org/10.2307/2580444.

Tracy, J. L., Robins, R. W., & Gosling, S. D. (2004). Tracking trends in psychological science. In T. C. Dalton & R. B. Evans (Eds.), *The life cycle of psychological ideas*. New York: Kluwer Academic/Plenum.

Vidal, F. (2011). *The sciences of the soul: The early modern origins of psychology* (S. Brown, Trans.). Chicago, IL: University of Chicago Press.

Vidal, F., & Ortega, F. (2017). *Being brains: Making the cerebral subject*. New York, NY: Fordham University Press.

Whitehead, A. N., & Russell, B. (1910, 1912, 1913). *Principia mathematica* (Vols. 1–3). Cambridge: Cambridge University Press.

Wiegele, T. C. (2018). *Biology and the social sciences: An emerging revolution*. New York, NY: Routledge.

*Original work published 1982*

Wittgenstein, L. (1956). *Remarks on the foundations of mathematics* (G. E. M. Anscombe, Trans.). Oxford, UK: Blackwell.