

1 The problem of defining life: a case study  
2 using family resemblance  
3

4 Jessica K. Abbott<sup>1,2</sup> & Erik Persson<sup>3</sup>

5 1. Corresponding author: [jessica.abbott@biol.lu.se](mailto:jessica.abbott@biol.lu.se)

6 2. Department of Biology

7 Lund University

8 221 00 Lund, Sweden

9 3. Department of Philosophy

10 Lund University

11 221 00 Lund, Sweden

12

13 Keywords: definition of life, Wittgenstein, de re definition, family resemblance, cluster analysis

14

15

## 16 Abstract

17 The question of how to define life has been an unresolved question in the philosophy of biology for  
18 many years, but developing a definition of life that is useful in both technical and everyday contexts  
19 has become more urgent as researchers around the world attempt to create fully synthetic cells in  
20 the laboratory, develop more and more intelligent and autonomous robots, and search for  
21 signatures of life elsewhere in the galaxy. Developments in these areas may end up overturning our  
22 current ideas about the distinction between life and non-life. It is therefore important to consider  
23 whether it is possible to develop a definition of life that encompasses currently known lifeforms,  
24 while at the same time having the potential to be applied to as-yet unknown lifeforms. Here, we  
25 discuss the pros and cons of some of the current approaches to defining life, then propose an  
26 alternative approach based on family resemblance. We also present preliminary data applying our  
27 new approach within a statistical modelling framework, and find that although living and non-living  
28 entities can be grouped according to overall similarity, it is difficult to find a single set of criteria  
29 which is sufficient for defining known forms of life while at the same time being inclusive enough to  
30 be useful in identifying or characterizing novel forms of life. We hope that the family resemblance  
31 approach will prove to be a fruitful alternative to traditional approaches to defining life.

32

## 33 Introduction

34 Imagine that you are reading a book or article and come across an unfamiliar word. If it is not  
35 possible to determine the meaning of the word from the context, chances are that you would look  
36 up the definition of the word in a dictionary. A typical dictionary definition includes a description of  
37 the meaning of the word, a list of possible alternative usages with examples of each usage, and  
38 probably a list of synonyms. Even an apparently simple word such as “chair” can be difficult to define  
39 if it has many possible uses depending on context. For example, the Oxford English Dictionary’s entry  
40 for the word “chair” is in fact over 5000 words long (Simpson and Weiner 1989).

41 This stands in contrast to how children learn new words when acquiring language. All humans learn  
42 how to speak their first language through their interactions with other individuals, rather than by  
43 looking words up in a dictionary (MacWhinney 1999). The child’s mother might say “sit down on the  
44 chair”, and use a gesture to help indicate her meaning. In this sort of situation, the child receives  
45 information not only about how a “chair” looks, but also what it’s used for. Over time, the child will  
46 see many examples of different types of chair and learn what they all have in common (MacWhinney  
47 1999). Eventually, the child becomes able to recognize even rather unusual examples of chairs as  
48 chairs (Figure 1).

49 This sort of learning process works well for words used in everyday situations, but what about more  
50 abstract concepts such as “life”? Technical and scientific definitions of the word “life” are necessary  
51 in various biological and chemical research fields, such as the origin of life (Pross 2016), but how well  
52 do these definitions align with a layperson’s definition of life that builds on personal experiences of  
53 life and death? Does it even matter if the technical and layperson definitions don’t align very well?

54 Although the question of how to define life has been an important issue in the philosophy of biology  
55 at least since Aristotle (Barnes 1984), it is becoming increasingly relevant as a result of current  
56 technological developments. Researchers around the world are now engaged in attempts to create  
57 fully synthetic cells in a laboratory setting (e.g. Gibson et al. 2010, Hutchison et al. 2016), and  
58 coordinated efforts are being made to try to detect signatures of life elsewhere in the galaxy (Seager  
59 2014). Developments in these areas may end up overturning our current ideas about the distinction  
60 between life and non-life. It is therefore important to consider whether it is possible to develop a  
61 definition of life that is useful in both technical and everyday contexts, while at the same time having  
62 the potential to be applied to as-yet unknown lifeforms. We will begin by discussing pros and cons of  
63 some current approaches to defining life, then propose an alternative approach, and finally present  
64 preliminary data applying our new approach.

## 65 Types of definitions

66 The two examples of the definition of the word “chair” discussed above – an adult who looks up an  
67 unknown word in the dictionary, and a child who learns new words through personal experience and  
68 interactions with other people – exemplify not only two different ways of acquiring information, but  
69 also two different approaches to defining objects and phenomena. In the first case, the dictionary  
70 definition, it is assumed that an exhaustive list of criteria can be made which must be fulfilled in  
71 order for an object to be considered a member of the category “chair”. This is typically called a *de re*  
72 *definition* (Føllesdal et al. 1988, Lübcke 1988, Bernadete 1993, Retana-Salazar and Retana-Salazar  
73 2004, Thompson 2008). “De re” is Latin and can be translated as “about the thing”; it is often  
74 contrasted with “lexical”, or “de dicto” (lat. “about what is said”) definitions (e.g. Gayon 2010). A *de*  
75 *re* definition therefore attempts to capture the essential properties of a phenomenon, while the *de*  
76 *dicto* reference aims to capture how a term is commonly used. *De re* definitions are common in  
77 dictionary definitions of everyday objects, as well as in technical definitions of objects and  
78 phenomena within the natural sciences (it is, for example, possible to unambiguously define an atom  
79 of gold based on the number of protons contained in the nucleus of the atom).

80 In the second case discussed above, similarities in form and function can be used to create an  
81 internal list of properties that a chair *usually* fulfils. A list of properties or criteria of this type is more  
82 fluid than in a *de re* definition. Not all criteria need always be fulfilled, the list of criteria can be  
83 dynamically updated, and criteria can be weighted differently depending on their relative  
84 importance. This more fluid type of definition makes it possible for us to recognize non-standard  
85 examples of chairs as long as they have enough in common with other types of chair we have  
86 previously encountered (Figure 1). This is an example of a *family resemblance* approach, where  
87 objects are understood based on their overall similarity rather than a list of necessary criteria  
88 (Wittgenstein 2001). Studies of language acquisition have shown that we generally use this type of  
89 approach when learning our native language (Medin and Schaffer 1978). The idea of family  
90 resemblance as a linguistic phenomenon was first developed by the philosopher Ludwig  
91 Wittgenstein (1889-1951) while working at the University of Cambridge in the early 1900’s. He  
92 wanted to investigate how we describe and understand complex phenomena such as art or games.

93 A game is a deceptively simple concept. Everyone knows what a game is, but to create a list of  
94 criteria shared by all games, while still excluding non-game phenomena (i.e. a typical *de re*  
95 definition) is likely impossible. The diversity of different types of games is simply too large; there are  
96 board games, ball games, card games, video games, role playing games, games that are played  
97 alone, games that are played as a group, cooperative games and competitive games, etc. The only

98 things that all these different types of games could perhaps be said to have in common is that they  
99 are entertaining and include some sort of element of competition (it must be possible to “win” the  
100 game, either by defeating the other team, the other players, or the game itself). But the same could  
101 be said of other types of phenomena, such as a structured debate, and it is not clear whether these  
102 properties apply to all sorts of games. This makes it difficult and perhaps even impossible to define  
103 games as a group in the same way as we define an atom of gold.

104 Some writers have therefore suggested that we should instead use the family resemblance approach  
105 for constructing definitions (Neuman 2012, Pennock 2012), not simply considering it as an  
106 interesting linguistic phenomenon. Using a family resemblance approach, one could use the  
107 properties of the majority of games to create a more intuitive definition (although we will not  
108 attempt this here). Under such a definition, not all criteria would need to be fulfilled, as long as the  
109 overall similarity is sufficiently high. For example, solitaire could be recognized as a game using a  
110 family resemblance approach even though most games require multiple players, based on its overall  
111 similarities with other types of card game.

## 112 “Fuzzy” definitions in biology

113 Biology encompasses many phenomena that are difficult to define. What is a gene, or a species? In  
114 both cases different definitions are used in different contexts, because it is difficult to develop a  
115 definition that works in all contexts (Dupré 1995, de Queiroz 2005). When it comes to defining  
116 species, there are a number of different widely-applied species concepts, and introductory  
117 textbooks in evolutionary biology typically include several different variants, each with its own pros  
118 and cons (e.g. Ridley 2003). One of the most common is the so-called biological species concept,  
119 which states that individuals that can successfully mate and produce fertile offspring are members of  
120 the same species (Ridley 2003, de Queiroz 2005). However this definition is not useful for organisms  
121 that do not reproduce sexually, such as bacteria or parthenogenetic organisms (which reproduce  
122 clonally via unfertilized eggs; Dupré 1995). It also does not take into account the fact that there are  
123 many plants and animals which can hybridize with other closely-related individuals when given the  
124 opportunity. For example, tigers (*Panthera tigris*) and lions (*Panthera leo*) can hybridize to produce  
125 partially fertile offspring (female hybrids are sometimes fertile but males are not), but this does not  
126 occur under natural conditions since the two species are found in different parts of the world (Li et  
127 al. 2016). Although these two species could therefore potentially be classed as different populations  
128 of the same species according to the biological species concept, this solution is not embraced by  
129 biologists due to their divergent morphology, non-overlapping ranges (Asia versus Africa),  
130 adaptation to different habitats (mainly forest versus mainly savannah), and different social

131 structure (solitary versus social; Castelló 2020). This simple example illustrates why many biologists  
132 embrace a pluralistic approach to defining species, where multiple possible definitions of what  
133 constitutes a species are acceptable, with different definitions being used in different contexts  
134 (Stanford 1995, de Queiroz 2005).

135 One reason why finding a single definition of a species is so challenging is because life is so diverse,  
136 and different groups of organisms can have completely different ways of living and reproducing. But  
137 this is not the whole story. Speciation is also a gradual process, which means that differences in  
138 morphology and behaviour can sometimes arise long before two populations become so genetically  
139 different that they can no longer hybridize (Li et al. 2016). This makes it very difficult to pinpoint any  
140 specific time when organisms cross the boundary from being two populations of the same species,  
141 to becoming two different species. There are therefore considerable similarities between the  
142 problem of defining species, and the problem of defining life. If anything, defining life is even more  
143 challenging since the origin of life is so difficult to study, and there are many entities which may in  
144 some sense be considered borderline between living and non-living, including viruses, intracellular  
145 parasites, or semi-autonomous components of organisms such as sperm.

146 Although we will not discuss it in any further detail, it is worth mentioning here the issue of “life”  
147 and “living” as the opposite of “death” and “dead”. Defining “living” versus “dead” is a question of  
148 individual status as alive or not, and is outside the scope of this work. Here, we are interested in  
149 being able to distinguish “living” from “non-living”, i.e. all organisms that currently exist, have  
150 existed, or may exist in the future, compared to objects or entities that have never been alive and  
151 never will. All non-avian dinosaurs are currently extinct, but they were alive while they still existed  
152 on earth. In contrast, a stone is not alive and never will be, so it would be nonsensical to call it  
153 “dead”. Nevertheless, some of the criteria that are traditionally associated with defining life versus  
154 non-life are also associated with defining living versus dead, making it difficult to discuss these issues  
155 completely independently of each other. For now, we intend to leave an analysis of the overlap  
156 between them to future work.

## 157 [How to define life? Problems and potential solutions](#)

158 Properties that are typically associated with life include, for instance, energy use, growth,  
159 reproduction, the ability to sense and react to the surrounding environment, and the ability to adapt  
160 to this environment (e.g. Solomon et al. 1993, Audesirk and Audesirk 1999). There are a number of  
161 biological systems that have some but not all of these properties, of which viruses are likely the most  
162 familiar. However other examples include transposable elements (DNA sequences which can copy  
163 themselves between different parts of the genome) or prions (proteins which can transform other

164 proteins into the same configuration), both of which could be considered to carry out a form of  
165 reproduction. Apart from the difficulties arising from known cases such as this, the challenge of  
166 defining life is multiplied when trying to develop a definition which could extend beyond the known  
167 forms of life. It is only a matter of time before we will have to deal with practical problems  
168 associated with alternative lifeforms, such as autonomous robots, general artificial intelligences, or  
169 synthetic cells (Persson et al. 2019). More speculative, but still within the realm of possibility, is the  
170 discovery of extra-terrestrial lifeforms, for example microorganisms under the surface of Mars. We  
171 need to be able to recognize these alternative forms of life even if they may be very different from  
172 currently known lifeforms. This is unlikely to be possible using a de re approach. If we cannot agree  
173 on a set of criteria to define known lifeforms, how much more difficult will it be to find a unique set  
174 of criteria which unify robotic, synthetic, and extra-terrestrial life?

175 There are several potential ways to move forward given these difficulties. We might decide to simply  
176 try harder, under the assumption that at some point we will have gained enough knowledge to  
177 construct a correct and all-encompassing definition of life. A problem with this approach is that all  
178 currently known life has a single origin (Pross 2016), which means that it may not be possible to  
179 determine which criteria are universal to all life and which are a contingent result of a common  
180 evolutionary history (Gould and Lewontin 1979). A second approach could therefore be to wait and  
181 see, i.e. put off trying to develop a single all-encompassing definition until we know what alternative  
182 forms of life it must include. Finally, we might decide to treat the definition of life in the same way as  
183 the species concept, and simply accept that life is too diverse for a single definition to be adequate.  
184 In this case different technical definitions of life could be developed depending on the context, such  
185 as the origin of life. The earliest biological systems probably lacked some of the criteria typically  
186 associated with modern lifeforms, such as the ability to sense the external environment (Pross 2016,  
187 West et al. 2017). Definitions of life that are used when researching the origin of life therefore tend  
188 to include broad criteria such as “energy use” and “dynamic equilibrium” (Pross 2016), rather than  
189 narrow ones such as “composed of cells” (Bedau 2010) or “contains information encoded in DNA”  
190 (Koshland 2002, Lazcano 2008, Bedau 2010). Given the ongoing discussion in the literature of the  
191 problem of defining life (Persson et al. 2019), none of these approaches seem to be particularly  
192 satisfying. We discuss why in more detail below.

### 193 [Try harder](#)

194 Although it may be valuable to continue searching for universal criteria for defining life, the question  
195 is how likely it is that this approach will succeed. Philosophers have attempted to define life at least  
196 since Aristotle (Barnes 1984), and if anything, the difficulties have increased over time rather than  
197 diminished as our knowledge has increased (Zimmer 2021). For example, whether viruses should be

198 considered living has been unclear since their discovery, and is still controversial among biologists  
199 today (Choi 2008, Hegde et al. 2009, López-Garcia and Moreira 2009, Ludmir and Enquist 2009,  
200 Moreira and López-Garcia 2009, Navas-Castillo 2009, Forterre 2010, Herrero-Urbe 2011), even  
201 though most introductory textbooks state that they are not living because they do not have a  
202 metabolism and cannot reproduce outside of a host cell (e.g. Solomon et al. 1993, Audesirk and  
203 Audesirk 1999). Proponents of the view that viruses should be considered living tend to consider the  
204 ability to reproduce and adapt via evolution, as well as evidence of their relatedness to the rest of  
205 the tree of life, to be more important criteria (Hegde et al. 2009, Ludmir and Enquist 2009). An  
206 informal survey of 40 professional biologists of all career stages at Lund University revealed an  
207 almost exact 50-50 split on this issue (Abbott, unpublished data). Recent research has not helped to  
208 clarify the issue.

209 There are three main hypotheses about the origin of viruses (Krupovic et al. 2019). One of the  
210 earliest hypotheses was that viruses are remnants of the most ancient lifeforms on earth, a  
211 transitional form between non-living organic molecules and living cells. Support for this hypothesis  
212 has decreased over time since it is hard to see how an entity that is completely dependent on living  
213 cells for reproduction could arise before the existence of such cells (Krupovic et al. 2019). A second  
214 hypothesis is that viruses arose from transposable elements that acquired the ability to transmit  
215 themselves horizontally (i.e. between unrelated individuals, rather than exclusively from parent to  
216 offspring). This hypothesis is fairly widely accepted since there are a number of structural and  
217 biochemical similarities between retroviruses and transposable elements (Biémont 2010, Krupovic et  
218 al. 2019). A third hypothesis is that viruses arose from bacteria or other intracellular parasites that  
219 lost more and more of their genetic material and independent functions as they became more  
220 dependent on their hosts (Krupovic et al. 2019). If this hypothesis is true, then this implies that  
221 viruses have evolved from living to non-living over time, which may be problematic for definitions of  
222 life that consider metabolism to be an essential criterion (Koshland 2002, Boden 2003, Bedau 2010).  
223 Recent work suggests that all three processes may have contributed to the origin of viruses  
224 (Krupovic et al. 2019), again making it problematic to definitively classify viruses as either living or  
225 non-living. Similar issues apply to mitochondria and chloroplasts, cell organelles with a bacterial  
226 origin – when does a symbiont transition from autonomous lifeform to cell component (George et al.  
227 2020)? We are therefore sceptical that the problem of defining life will be resolved if we simply  
228 collect more data.

229 [Wait and see](#)

230 As mentioned above, one of the reasons why it is difficult to define life is because we currently only  
231 have information about life on earth. We therefore cannot conclusively distinguish between



232 properties that are truly essential or universal from those which are specific to our kind of life. For  
233 example, some definitions of life include the criterion that individuals should encode information in  
234 DNA (Audesirk and Audesirk 1999), or be composed of cells (Solomon et al. 1993). It is relatively easy  
235 to imagine that another type of molecule could be used to encode information rather than DNA  
236 (Pinheiro and Holliger 2012), but less clear whether it is possible for lifeforms to exist that are not  
237 composed of cells. Experiments with synthetic lifeforms may be useful in helping to investigate these  
238 questions, but these synthetic lifeforms will likely be highly influenced by existing life, partly because  
239 it is most convenient if they can be kept under ordinary lab conditions, but also because existing  
240 lifeforms are often used as templates for synthetic organisms (Osbourn et al. 2012). For example,  
241 *Syn-3.0*, a synthetic bacterium created by Hutchinson et al. (2016) is a highly modified version of the  
242 bacterium *Mycoplasma mycoides*, an intracellular parasite of cattle and goats. Synthetic biology is a  
243 fast-developing field, yet it is unclear how long it will take to develop synthetic lifeforms that are  
244 radically different from known life.

245 Because of this, it would actually be much more valuable if extra-terrestrial life with a completely  
246 independent origin is eventually detected. But whether this is even feasible is currently unclear.  
247 Although it might be technologically feasible to bring back samples from Mars, either now or in near  
248 future, there is a history of traffic between the earth and Mars both through natural (meteorites)  
249 and artificial (probes) means, which means that should life be detected there then there is no  
250 guarantee that it will have an independent origin (Sullivan and Baross 2007). It is also conceivable  
251 that lifeforms based on different chemistries could exist in other parts of the solar system, and be  
252 accessible for direct study at some point in the future (Sullivan and Baross 2007, Petrowski et al.  
253 2020). But here we end up in a catch-22 of sorts. Without a definition of life that goes beyond  
254 currently known types of life, how are we supposed to recognize these novel lifeforms? We can  
255 therefore conclude that although we will surely learn much if and when we create or discover  
256 completely new types of life, we cannot wait until then to develop a better definition of life.

### 257 [Accept that there cannot be a single definition](#)

258 Another option is to abandon the search for a single all-encompassing definition of life (Jeuken 1975,  
259 Oliver and Perry 2006, Mix 2015), similar to the plurality of species concepts that was discussed  
260 above. Such an approach would result in the development or refinement of multiple technical  
261 definitions of life, each of which would be most useful within a specific context. Different criteria  
262 could be included depending on whether the definition was to be relevant for the transition from  
263 pre-biotic chemistry to living organism during the origin of life, when an autonomous robot or  
264 artificial intelligence could be considered alive, or what minimal properties a potential extra-  
265 terrestrial lifeform might need in order to qualify as such. The advantage of this approach is that it

266 reflects the fact that life is a dynamic process which is difficult to capture using a limited set of  
267 criteria.

268 However there are also disadvantages with this approach. For one thing, a single common definition  
269 of life would be very useful when discussing the challenges associated with novel forms of life. A  
270 definition of life which is broadly applicable and understandable for laypersons and policymakers as  
271 well as experts, would be an advantage during public debate of issues related to novel forms of life.  
272 This does not mean that a broad definition would necessarily replace the various technical  
273 definitions within a given field, but it might make interdisciplinary communication more successful.  
274 Another potential problem with existing technical de re definitions of life is that many of these  
275 definitions include criteria at different hierarchical levels of organization. For example, a typical  
276 textbook definition of life might include the criteria “energy use”, “growth and development”,  
277 “reproduction”, “homeostasis” (the ability to maintain a consistent internal state), and “evolutionary  
278 adaptation” (Campbell et al. 1987). Of these criteria, the first four can be applied to the individual  
279 organism, but the last one can only be applied to a population of organisms (Persson 2013). A  
280 population cannot undergo development from embryo to adult (other than via its component  
281 organisms), while a single individual cannot undergo evolutionary adaptation. This leads to an odd  
282 situation where this definition cannot sensibly be applied in its entirety to *either* a specific individual  
283 *or* to a specific population of individuals, calling its general utility into question.

284 Some authors have attempted to resolve this problem by developing definitions of life in terms of  
285 systems rather than individuals, circumventing the issue of criteria which are applicable at different  
286 hierarchical levels (Ruiz-Mirazo et al. 2004, Weber 2010, Pross 2016). Since evolutionary adaptation  
287 is usually considered one of the most important properties of life compared to other dynamically  
288 stable non-living systems (such as a self-driving car, or the earth’s system of ocean currents), this  
289 means in practice that the system must be defined at the population level or higher. The advantage  
290 of such a definition is that it is more internally consistent, but this usually comes at the expense of  
291 being more complicated, imprecise, or unintuitive (Ruiz-Mirazo et al. 2004, Weber 2010, Pross  
292 2016). The question is whether the problems discussed above are because defining life is impossible,  
293 or are they a result of the traditional fixation on a de re definition?

## 294 [A case study of a family resemblance approach](#)

295 We would like to suggest that continuing to search for a broadly applicable de re definition of life is  
296 unlikely to be fruitful, and that it would be better to try a new type of approach which can  
297 accommodate the complexity of life as we know it. In our view, a family resemblance-based  
298 approach has better potential to achieve the aim of producing a broadly applicable and intuitive

299 definition of life, since such a definition would be based on overall similarities across lifeforms and  
300 allow for occasional exceptions from specific criteria. A family resemblance approach could also  
301 allow for different weighting of criteria, for example if metabolism is considered to be a more  
302 important property of life than evolutionary adaptation (or vice versa).

303 A potential weakness of the family resemblance approach to definitions is that it is not sufficiently  
304 objective. What if we cannot all agree on what or how much two entities need have in common, in  
305 order for both to be classified as alive? This is of course a problem, but not, we would argue, an  
306 insurmountable one. We have attempted to use statistical modelling to determine which criteria are  
307 most useful for describing life as we know it, while excluding criteria that are uninformative. Using  
308 this information, it may be possible to develop a definition of life that builds on specific  
309 combinations of criteria, rather than a single exhaustive list. Here we present preliminary results  
310 from a case study illustrating how this type of approach might work.

## 311 [Methods](#)

312 First, we compiled a list of suggested criteria for defining life from the primary literature, as well as  
313 from introductory textbooks in biology (see table 1). Descriptions of a phenomenon were considered  
314 to be equivalent to specific terms with the same meaning, such that “maintenance of their  
315 functional systems” (Allaby 1991) was considered equivalent to “homeostasis” (Campbell et al.  
316 1987). This resulted in approximately 30 individual criteria, some of which were variants on a similar  
317 theme; e.g. “evolves”, “adapt to environment”, and “natural selection” were considered separate  
318 criteria since evolution can occur through purely neutral processes and does not necessarily require  
319 adaptation to the environment or natural selection (Ridley 2003). We then selected a number of  
320 living organisms and other entities (including objects, processes, and organismal components) which  
321 were intended to span the boundary between living and non-living (see table 2). Next, we attempted  
322 to determine whether these entities fulfilled the criteria in table 1 or not. This resulted in the  
323 production of a data matrix where 1 indicated that the criterion was fulfilled for that entity, and zero  
324 indicated that it was not. Note that it was not always obvious how a given criterion should be  
325 applied to a specific entity (e.g. does a bacterium have purposiveness?), but we discussed these  
326 cases thoroughly and attempted to use our best judgement. We also scored each entity according to  
327 whether it is usually considered to be living or not (see table 2). This classification as “living” or “non-  
328 living” is of course also open to interpretation, so in the cases where the answer was ambiguous  
329 (e.g. plastids or viruses), we carried out two sets of analyses – one more inclusive analysis where all  
330 ambiguous cases were classified as “living”, and one more conservative analysis where all ambiguous  
331 cases were classified as “non-living”.

332 Once the data matrix was complete, we analysed it in various ways. All analyses were carried out in  
333 the R Statistical Programming environment (R Core Team 2020). First, we checked whether our  
334 evaluation of all criteria produced a useful descriptive result by carrying out a cluster analysis  
335 (Ward's method using the `hclust()` function; Legendre and Legendre 2012, R Core Team 2020). We  
336 expected that successful scoring would result in living organisms clustering separately from other  
337 entities. Next, we carried out linear discriminant analysis (using the `lda()` function in the MASS  
338 package; Venables and Ripley 2002) to determine if it was possible to accurately assign entities to  
339 the classes "living" and "non-living" using the full set of criteria. If so, this might suggest that a de re  
340 definition is feasible after all. Finally, we checked which criteria were most highly correlated with  
341 classification as "living" using Spearman rank correlation (using the `cor.test()` function; Keough and  
342 Quinn 2002, R Core Team 2020), in order to determine which criteria may be most useful in  
343 constructing future definitions of life.

344 Several of the criteria were found to produce exactly the same result when evaluated across all  
345 entities. For example, entities that fulfilled the criterion "metabolism" invariably also fulfilled the  
346 criteria "growth" and "stimulus response" (at least within this particular dataset). We therefore  
347 collapsed perfectly correlated criteria into single variables in order to reduce the dimensionality of  
348 the dataset for the linear discriminant and correlation analyses. The full (unreduced) dataset was  
349 used for the cluster analysis, in order to preserve complete information about relative similarity.  
350 However results were qualitatively similar when carried out on the reduced dataset (data not  
351 shown).

## 352 Results

353 The cluster analysis revealed that non-living entities generally did not cluster with living entities (see  
354 figure 2). However both main clusters included some ambiguous cases; for example, red blood cells  
355 and sperm clustered together with living intracellular parasite species, and viruses clustered  
356 together with non-living entities. This means that even using a more conservative classification of  
357 viruses as non-living, it is still difficult to clearly separate (more or less) autonomous living organisms  
358 from their component parts. The linear discriminant analysis confirmed this. For the inclusive  
359 classification dataset, entities were correctly classified as "living" only 44% of the time (14/32 cases –  
360 4 correctly predicted as "non-living" and 10 correctly predicted as "living"). Entities which were  
361 classified as "living" in this dataset tended to be misclassified as "non-living" in the analysis more  
362 often than the reverse (14 living organisms incorrectly predicted as "non-living" compared to 4 non-  
363 living entities incorrectly predicted as "living"). For the conservative classification dataset, the  
364 criteria "feed" and "homeostasis" correctly predicted all cases.

365 The criteria that were most and least highly correlated with our classification as “living” are  
366 presented in table 3. Results were somewhat different for the inclusive and conservative  
367 classification datasets. The only criteria that were among the top 5 best predictors in both datasets  
368 were “autocatalytic cycles” and “enzymes”. Poorly-performing criteria across both datasets were  
369 more consistent, and included “mutation”, “reproduction by self or non-self”, “order”, “adapt to  
370 environment”, and “natural selection”. These results suggest (1) that exactly which criteria are best  
371 for defining life depend on the specific set of lifeforms we wish to define, and (2) that criteria which  
372 are likely to be most useful for identifying novel forms of life (e.g. “reproduction by self or non-self”,  
373 “order” or “adapt to environment”) are unlikely to be sufficient for classifying known forms of life.

## 374 Conclusions

375 Consistent with our expectations, living organisms tended to cluster separately from non-living  
376 objects and entities in our proof-of-concept analysis. However the distinction between cell  
377 components and intracellular parasites was not particularly clear, consistent with our expectation  
378 that defining life based on a single set of criteria is not easy (figure 2). In addition, the poor  
379 classification results in the linear discriminant analysis of the inclusive dataset suggest, as discussed  
380 above, that simply adding more information does not necessarily help to resolve problems with  
381 separating life from non-life. Interestingly, there seems to be a trade-off between accuracy and  
382 broad applicability when attempting to define life. Very general criteria such as “natural selection”  
383 or “order”, which are expected to be most useful in the context of the origin of life or extra-  
384 terrestrial life, were poorly correlated with classification as “living” or “non-living” in this dataset  
385 (table 3). This suggests that although it might be possible to develop a de re definition of life that  
386 encompasses all currently living organisms on earth, such a definition is unlikely to be useful in the  
387 context of understanding and characterizing novel lifeforms.

388 This case study only includes a limited number of entities and criteria, and this will of course affect  
389 the outcome of the analyses to some extent. For example, correlations between criteria and  
390 classification as living will likely change depending on exactly which combinations of criteria and  
391 entities are included in the analysis. The differences in outcome between the inclusive and  
392 conservative classification datasets reflect this. Similarly, how each criterion should be interpreted  
393 with respect to a given entity is not always straightforward. For example, does a red blood cell have  
394 genetic control of development? Yes, in the sense that our genes control the development of our red  
395 blood cells, and that the maturing cell itself must actively express these genes in order to develop  
396 normally (Moras et al. 2017). But it would also be reasonable to argue that the answer should be no,  
397 in the sense that the red blood cell does not itself contain any genes when it is mature, and

398 therefore cannot produce any new red blood cells via genetic control of development (Moras et al.  
399 2017). The specific results presented here should therefore be considered preliminary, and it might  
400 be necessary to reach some sort of consensus with respect to the evaluation of the criteria in order  
401 to obtain robust results.

402 As presented here, our approach is mainly descriptive. However we do not feel that this is a major  
403 drawback. For one thing, any useful definition must be able to reflect our intuitions about life, which  
404 means that describing these intuitions is an important first step towards being able to construct a  
405 broad definition of life. An advantage of this approach is that it can also be applied iteratively,  
406 dynamically updating our descriptions of life as new information is obtained. In addition, further  
407 work could build on these results to help us get closer to a broadly-applicable definition of life. For  
408 example, which combinations of criteria best describe each cluster within figure 2? What happens if  
409 we include hypothetical examples of novel forms of life? Some authors have suggested that life  
410 might be a matter of degree, rather than a binary property (Hazen 2009, Bedau 2010, Jager op  
411 Akkerhuis 2010, Tirard et al. 2010). Would it then be more useful to include additional levels of  
412 classification than simply “living” versus “non-living”? There are many outstanding questions and  
413 possible directions to explore. Nevertheless, we hope that this case study of a family resemblance  
414 approach to defining life shows the potential which this approach provides.

415

416

## References

417

- 418 Allaby, M. 1991. *The concise Oxford dictionary of zoology*, Oxford, UK.  
419 Audesirk, T., and G. Audesirk. 1999. *Biology: Life on Earth*, 5rd edition. Prentice Hall, Hoboken, USA.  
420 Barnes, J. 1984. *The complete works of Aristotle*. Princeton University Press, Princeton, USA.  
421 Bedau, M. A. 2010. An Aristotelian account of minimal chemical life. *Astrobiology* **10**:1011-1020.  
422 Bernadete, J. A. 1993. Real definitions: Quine and Aristotle. *Philosophical Studies* **72**:265-282.  
423 Biémont, C. 2010. A brief history of the status of transposable elements: from junk DNA to major  
424 players in evolution. *Genetics* **186**:1085-1093.  
425 Boden, M. A. 2003. Alien life: how would we know? *International Journal of Astrobiology* **2**:121-129.  
426 Campbell, N. A., L. G. Mitchell, and J. B. Reece. 1994. *Biology: Concepts and connections*.  
427 Benjamin/Cummings Publishing Company, San Francisco, USA.  
428 Campbell, N. A., J. B. Reece, and L. G. Mitchell. 1987. *Biology*, 5th edition. Pearson, London, UK.  
429 Castelló, J. R. 2020. *Felids and hyenas of the world*. Princeton University Press, Princeton, USA.  
430 Choi, C. Q. 2008. It's alive? *Scientific American* **299**:18.  
431 de Queiroz, K. 2005. Ernst Mayr and the modern concept of species. *Proc. Nat. Acad. Sci. USA*  
432 **102**:6600-6607.  
433 Dupré, J. 1995. *The disorder of things: metaphysical foundations of the disunity of science*. Harvard  
434 University Press, Cambridge MA.  
435 Fenchel, T. 2002. *Origin and early evolution of life*. Oxford University Press, Oxford, UK.  
436 Føllesdal, D., L. Walløe, and J. Elster. 1988. *Argumentasjonsteori, språk og vitenskapsfilosofi*.  
437 Universitetsforlaget, Oslo, Norway.

438 Forterre, P. 2010. Defining Life: The Virus Viewpoint. *Origins of Life and Evolution of Biospheres*  
439 **40**:151-160.

440 Gayon, J. 2010. Defining life: Synthesis and conclusions. *Origins of Life and Evolution of Biospheres*  
441 **40**:231-244.

442 George, E. E., F. Husnik, D. Tashyreva, G. Prokopchuk, A. Horák, W. K. Kwong, J. Lukeš, P. J. Keeling,  
443 and M. Saad. 2020. Highly reduced genomes of protist endosymbionts show evolutionary  
444 convergence. *Current Biology* **30**:925-933.

445 Gibson, D. G., J. I. Glass, C. Lartigue, V. N. Noskov, R.-Y. Chuang, M. A. Algire, G. A. Benders, M. G.  
446 Montague, L. Ma, M. M. Moodie, C. Merryman, S. Vashee, R. Krishnakumar, N. Assad-Garcia,  
447 C. Andrews-Pfannkoch, E. A. Denisova, L. Young, Z.-Q. Qi, T. H. Segall-Shapiro, C. H. Calvey, P.  
448 P. Parmar, C. A. Hutchinson, III, H. O. Smith, and J. C. Venter. 2010. Creation of a bacterial  
449 cell controlled by a chemically synthesized genome. *Science* **329**:52-56.

450 Gould, J. L., and W. T. Keeton. 1995. *Biological science*, 6th edition W. W. Norton & Company, New  
451 York, USA.

452 Gould, S. J., and R. C. Lewontin. 1979. The spandrels of San Marco and the Panglossian paradigm: a  
453 critique of the adaptationist programme. *Proceedings of the Royal Society of London B*  
454 **205**:581-598.

455 Greener, M. 2008. It's life, but just as we know it. *EMBO Reports* **9**:1067-1069.

456 Hansen, A. 2008. Unqualified interests, definitive definitions: *Washington v. Glucksberg* and the  
457 definition of life. *Hastings Constitutional Law Quarterly* **36**:163.

458 Hazen, R. M. 2009. Emergence and the experimental pursuit of the origin of life. Pages 21-46 *in* C. M.  
459 Bertka, editor. *Exploring the origin, extent, and future of life*. Cambridge University Press,  
460 Cambridge, UK.

461 Hegde, N. R., M. S. Maddur, S. V. Kaveri, and J. Bayry. 2009. Reasons to include viruses in the tree of  
462 life. *Nature Reviews Microbiology* **7**:615.

463 Herrero-Urbe, L. 2011. Viruses, definitions and reality. *Revista de biologia tropical* **59**:993-998.

464 Hutchison, C. A., III, R.-Y. Chuang, V. N. Noskov, N. Assad-Garcia, T. J. Deerinck, M. H. Ellisman, J. Gill,  
465 K. Kannan, B. J. Karas, L. Ma, J. F. Pelletier, Z.-Q. Qi, R. A. Richter, E. A. Strychalski, L. Sun, Y.  
466 Suzuki, B. Tsvetanova, K. S. Wise, H. O. Smith, J. I. Glass, C. Merryman, D. G. Gibson, and C.  
467 Venter. 2016. Design and synthesis of a minimal bacterial genome. *Science* **351**:aad6253.

468 Jager op Akkerhuis, G. A. J. M. 2010. Towards a hierarchical definition of life, the organism, and  
469 death. *Foundations of Science* **15**:245-262.

470 Jeuken, M. 1975. The biological and philosophical definitions of life. *Acta Biotheoretica* **24**:14-21.

471 Keough, M. J., and G. P. Quinn. 2002. *Experimental design and data analysis for biologists*.  
472 Cambridge University Press, Cambridge, UK.

473 Korzeniewski, B. 2001. Cybernetic formulation of the definition of life. *Journal of Theoretical Biology*  
474 **209**:275-286.

475 Koshland, D. E., Jr. 2002. The seven pillars of life. *Science* **295**:2215-2216.

476 Krupovic, M., V. V. Dolja, and E. V. Koonin. 2019. Origin of viruses: primordial replicators recruiting  
477 capsids from hosts. *Nature Reviews Microbiology* **17**:449-458.

478 Lawrence, E. 2000. *Henderson's dictionary of biological terms*, 12th edition. Prentice Hall, New York,  
479 USA.

480 Lazcano, A. 2008. Towards a definition of life: the impossible quest? *Space Science Reviews* **135**:5-  
481 10.

482 Legendre, P., and L. Legendre. 2012. *Numerical Ecology*. Elsevier, Amsterdam, Netherlands.

483 Li, G., B. W. Davis, E. Eizirik, and W. J. Murphy. 2016. Phylogenomic evidence for ancient  
484 hybridization in the genomes of living cats (Felidae). *Genome Research* **26**:1-11.

485 López-Garcia, P., and D. Moreira. 2009. Yet viruses cannot be included in the tree of life. *Nature*  
486 *Reviews Microbiology* **7**:615-617.

487 Lübcke, P. 1988. *Filosoflexikonet*. Forum, Stockholm, Sweden.

488 Ludmir, E. B., and L. W. Enquist. 2009. Viral genomes are part of the phylogenetic tree of life. *Nature*  
489 *Reviews Microbiology* **7**:615.

490 Macklem, P. T., and A. Seely. 2010. Towards a definition of life. *Perspectives in Biology and Medicine*  
491 **53**:330-340.

492 MacWhinney, B. 1999. *The Emergence of Language*. Lawrence Erlbaum Associates, Mahwah, USA.

493 Mader, S. S. 2001. *Biology*, 7th edition. McGraw Hill, New York, USA.

494 Medin, D. L., and M. M. Schaffer. 1978. Context theory of classification learning. *Psychological*  
495 *Review* **85**:207-238.

496 Mix, L. J. 2015. Defending definitions of life. *Astrobiology* **15**:15-19.

497 Moore, A. 2012. Life defined. *BioEssays* **34**:253-254.

498 Moras, M., S. D. Lefevre, and M. A. Ostuni. 2017. From erythroblasts to mature red blood cells:  
499 organelle clearance in mammals. *Frontiers in Physiology* **8**:1076.

500 Moreira, D., and P. López-García. 2009. Ten reasons to exclude viruses from the tree of life. *Nature*  
501 *Reviews Microbiology* **7**:306-311.

502 Morowitz, H. 1992. *Beginnings of cellular life: Metabolism recapitulates biogenesis*. Yale University  
503 Press, New Haven, USA.

504 Navas-Castillo, J. 2009. Six comments on the ten reasons for the demotion of viruses. *Nature*  
505 *Reviews Microbiology* **7**:615.

506 Neuman, Y. 2012. The definition of life and the life of a definition. *Journal of Biomolecular Structure*  
507 *and Dynamics* **29**:643-646.

508 Oliver, J. D., and R. S. Perry. 2006. Definitely life but not definitively. *Origins of Life and Evolution of*  
509 *Biospheres* **36**:515-521.

510 Osbourn, A. E., P. E. O'Maille, S. J. Rosser, and K. Lindsey. 2012. Synthetic biology. *New Phytologist*  
511 **196**:671-677.

512 Pennock, R. T. 2012. Negotiating boundaries in the definition of life: Wittgensteinian and Darwinian  
513 insights on resolving conceptual border conflicts. *Synthese* **185**:5-20.

514 Persson, E. 2013. Philosophical aspects of astrobiology. Pages 29-48 *in* D. Dunér, J. Parthemore, E.  
515 Persson, and G. Holmberg, editors. *The History and Philosophy of Astrobiology*. Cambridge  
516 Scholars, Newcastle upon Tyne, UK.

517 Persson, E., J. K. Abbott, C. Balkenius, A. Cabak Redei, K. A. Čáková, D. Dravins, D. Dunér, M.  
518 Gunneflo, M. Hedlund, M. Johansson, A. Melin, and P. Persson. 2019. How will the emerging  
519 plurality of lives change how we conceive of and relate to life? *Challenges* **10**:32.

520 Petrowski, J. J., W. Bains, and S. Saegge. 2020. On the potential of silicon as a building block for life.  
521 *Life* **10**:84.

522 Pinheiro, V. B., and P. Holliger. 2012. The XNA world: progress towards replication and evolution of  
523 synthetic genetic polymers. *Current Opinion in Chemical Biology* **16**:245-252.

524 Pross, A. 2016. *What is life? How chemistry becomes biology*. Oxford University Press, Oxford, UK.

525 R Core Team. 2020. *R: A language and environment for statistical computing*. R Foundation for  
526 Statistical Computing, Vienna, Austria.

527 Raven, P. H., and G. B. Johnson. 1996. *Biology*, 4th edition. McGraw-Hill, New York, USA.

528 Retana-Salazar, A. P., and S. Retana-Salazar. 2004. Towards a simple logic in the determination of  
529 biological groups: the species and supraspecific groups. *Revista de biologia tropical* **52**:19-26.

530 Ridley, M. 2003. *Evolution*. Wiley-Blackwell, Hoboken, USA.

531 Ruiz-Mirazo, K., J. Peretó, and A. Moreno. 2004. A universal definition of life: Autonomy and open-  
532 ended evolution. *Origins of Life and Evolution of Biospheres* **34**:323-346.

533 Sadava, D. E., H. C. Heller, W. K. Orians, G. H. Purves, and D. M. Hillis. 2008. *Life: the science of*  
534 *biology*, 8th edition. W. H. Freeman, New York, USA.

535 Sapp, J. 2003. *Genesis: the evolution of biology*. Oxford University Press, Oxford, UK.

536 Seager, S. 2014. The future of spectroscopic life detection on exoplanets. *Proc. Nat. Acad. Sci. USA*  
537 **111**:12634-12640.

538 Simpson, J., and E. Weiner. 1989. *Oxford English Dictionary*. Oxford University Press, Oxford, UK.



539 Solomon, E. P., L. R. Berg, D. W. Martin, and C. Vilee. 1993. *Biology*, 3rd edition. Saunders,  
540 Philadelphia, USA.

541 Stanford, P. K. 1995. For pluralism and against realism about species. *Philosophy of Science* **62**:70-  
542 91.

543 Strickberger, M. W. 2000. *Evolution*, 3rd edition. Jones and Bartlett Publishers, Burlington, USA.

544 Sullivan, W., T., III, and J. Baross. 2007. *Planets and life*. Cambridge University Press, Cambridge, UK.

545 Thompson, M. 2008. *Life and action: Elementary structures of practice and practical thought*.  
546 Harvard University Press, Cambridge, USA.

547 Tirard, S., M. Morange, and A. Lazcano. 2010. The definition of life: a brief history of an elusive  
548 scientific endeavor. *Astrobiology* **10**:1003-1008.

549 Venables, W. N., and B. D. Ripley. 2002. *Modern Applied Statistics with S*, 4th edition. Springer, New  
550 York, USA.

551 Weber, B. H. 2010. What is life? Defining life in the context of emergent complexity. *Origins of Life*  
552 *and Evolution of Biospheres* **40**:221-229.

553 West, T., V. Sojo, A. Pomiankowski, and N. Lane. 2017. The origin of heredity in protocells. *Phil.*  
554 *Trans. R. Soc. Lond. B Biol. Sci* **372**.

555 Wingo, W. 1963. Definition of life. *The Science News-letter* **84**:147.

556 Wittgenstein, L. 2001. *Philosophical Investigations*. Blackwell Publishing, Oxford.

557 Zimmer, C. 2021. *Life's edge: The search for what It means to be alive*. Dutton, New York, USA.

558

559

560  
561

Table 1: List of criteria defining for defining life obtained from the literature and from introductory textbooks in biology.

<b>Criterion</b>	<b>Source</b>
Adapt to environment	(Wingo 1963, Campbell et al. 1987, Solomon et al. 1993, Mader 2001)
Autocatalytic cycles	(Morowitz 1992, Fenchel 2002, Macklem and Seely 2010)
Boundary	(Campbell et al. 1994, Greener 2008, Macklem and Seely 2010, Pennock 2012)
Can die	(Hansen 2008)
Cells	(Solomon et al. 1993, Campbell et al. 1994, Fenchel 2002, Sadava et al. 2008)
Chromosomes	(Wingo 1963)
Decrease in entropy	(Wingo 1963, Morowitz 1992, Fenchel 2002, Moore 2012, Pennock 2012)
Development	(Campbell et al. 1987, Solomon et al. 1993, Gould and Keeton 1995, Raven and Johnson 1996, Mader 2001)
DNA	(Campbell et al. 1994, Audesirk and Audesirk 1999, Fenchel 2002)
Enzymes	(Lawrence 2000, Fenchel 2002, Sapp 2003)
Evolves	(Wingo 1963, Campbell et al. 1987, Solomon et al. 1993, Gould and Keeton 1995, Audesirk and Audesirk 1999, Fenchel 2002, Sapp 2003, Greener 2008, Sadava et al. 2008, Pennock 2012)
Feeding	(Greener 2008)
Genes	(Gould and Keeton 1995, Strickberger 2000, Sapp 2003, Sadava et al. 2008)
Genetic control of development	(Pennock 2012)
Genetic material isolated from environment	(Pennock 2012)
Growth	(Wingo 1963, Campbell et al. 1987, Allaby 1991, Solomon et al. 1993, Raven and Johnson 1996, Audesirk and Audesirk 1999, Lawrence 2000, Strickberger 2000, Fenchel 2002, Greener 2008, Pennock 2012)
Homeostasis	(Campbell et al. 1987, Allaby 1991, Solomon et al. 1993, Audesirk and Audesirk 1999, Sadava et al. 2008)
Metabolism	(Wingo 1963, Morowitz 1992, Solomon et al. 1993, Campbell et al. 1994, Gould and Keeton 1995, Lawrence 2000, Strickberger 2000, Greener 2008, Sadava et al. 2008, Pennock 2012)
Movement	(Solomon et al. 1993)
Mutation	(Pennock 2012)
Natural selection	(Greener 2008)
Nucleic acids	(Lawrence 2000)
Order	(Wingo 1963, Campbell et al. 1987, Gould and Keeton 1995, Raven and Johnson 1996, Mader 2001, Sapp 2003, Moore 2012)
Organic molecules	(Audesirk and Audesirk 1999, Fenchel 2002, Pennock 2012)

Protoplasm	(Wingo 1963, Sapp 2003)
Purposiveness	(Pennock 2012)
Regulatory mechanisms	(Raven and Johnson 1996, Korzeniewski 2001, Macklem and Seely 2010, Pennock 2012)
Related	(Sadava et al. 2008)
Replicate chemical information	(Fenchel 2002, Moore 2012, Pennock 2012)
Reproduction by self or non-self (not specified)	(Campbell et al. 1987, Allaby 1991, Solomon et al. 1993, Gould and Keeton 1995, Raven and Johnson 1996, Lawrence 2000, Strickberger 2000, Mader 2001, Fenchel 2002, Greener 2008, Moore 2012, Pennock 2012)
Reproduction by self-replication	(Wingo 1963, Morowitz 1992, Audesirk and Audesirk 1999, Sadava et al. 2008, Macklem and Seely 2010)
Stimulus response	(Wingo 1963, Campbell et al. 1987, Morowitz 1992, Solomon et al. 1993, Raven and Johnson 1996, Audesirk and Audesirk 1999, Lawrence 2000, Mader 2001, Greener 2008, Pennock 2012)
Use external substances	(Campbell et al. 1987, Allaby 1991, Audesirk and Audesirk 1999, Lawrence 2000, Mader 2001, Fenchel 2002, Greener 2008, Sadava et al. 2008)

562

563

564 Table 2: List of entities included in the analysis of life. Entities indicated as “Yes/No” were considered  
 565 ambiguous, and analysed separately as both living and non-living in order to see how this influenced  
 566 the outcome of the analysis.

Entities	Classified as living?
<i>Arabidopsis thaliana</i> (thale cress)	Yes
<i>Caenorhabditis elegans</i> (nematode)	Yes
<i>Chlamydomonas reinhardtii</i> (single-celled alga)	Yes
Clay crystals	No
Colloids	No
<i>Danio rerio</i> (zebrafish)	Yes
DNA virus (e.g. herpes simplex)	Yes/No
<i>Drosophila melanogaster</i> (common vinegar fly)	Yes
<i>Escherichia coli</i> (intestinal bacterium)	Yes
<i>Gallus gallus</i> (domestic chicken)	Yes
<i>Homo sapiens</i> (human)	Yes
<i>Hydra vulgaris</i> (freshwater polyp)	Yes
Liposome (phospholipid bilayer vesicle)	No
<i>Macrostomum lignano</i> (flatworm)	Yes
Memes	No
<i>Mus musculus</i> (house mouse)	Yes
Plastids (cell organelle, e.g. chloroplast)	Yes/No
<i>Poecilia formosa</i> (Amazon molly)	Yes
Prion (e.g. CJD-causing)	No
Red blood cell	Yes/No
Ribozyme (catalytic RNA)	No
<i>Rickettsia typhi</i> (intracellular parasitic bacterium)	Yes
RNA virus (e.g. HIV)	Yes/No
<i>Saccharomyces cerevisiae</i> (brewer’s yeast)	Yes
Snowflake	No
Sperm	Yes/No
<i>Synechocystis</i> sp. PCC 6803 (cyanobacterium)	Yes
Transposon	No
Viroid (e.g. tomato chlorotic dwarf viroid)	Yes/No
Virophage (viral parasite on other viruses)	Yes/No
<i>Wolbachia</i> (intracellular parasitic bacterium)	Yes
<i>Xenopus tropicalis</i> (western clawed frog)	Yes

567

568

569 Table 3: Criteria which are most (white rows) and least (grey rows) strongly associated with a priori  
 570 classification as living versus non-living. Criteria with equal correlation coefficients are presented on  
 571 the same row. A) Results when ambiguous cases are classified as "living". B) Results when  
 572 ambiguous cases are classified as "non-living". Significant correlations are indicated in bold.

573 A)

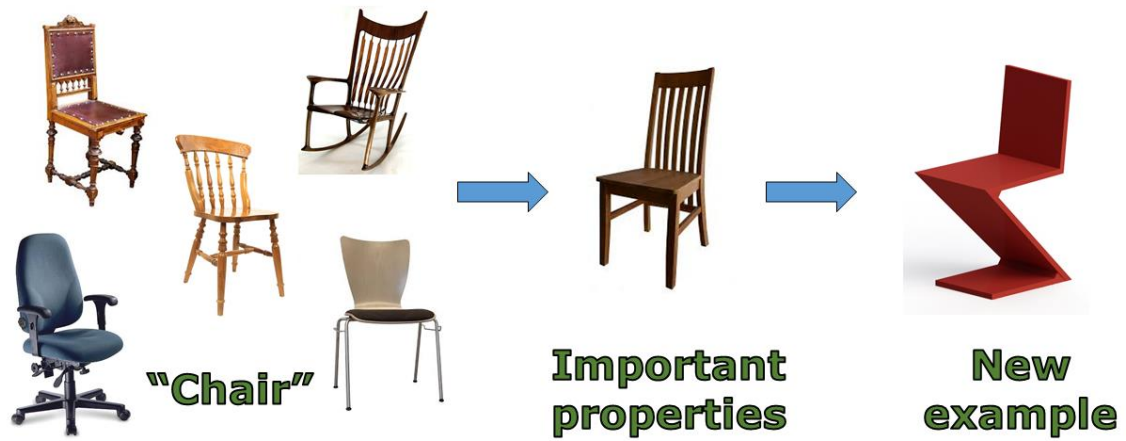
Criteria	Correlation coefficient	P-value
Autocatalytic cycles, Chromosomes, Genes	0.833	<b>3.23*10<sup>-9</sup></b>
Enzymes	0.832	<b>3.59*10<sup>-9</sup></b>
Movement	0.762	<b>3.95*10<sup>-7</sup></b>
Nucleic acids, DNA	0.745	<b>9.87*10<sup>-7</sup></b>
Reproduction by self-replication	0.698	<b>8.96*10<sup>-6</sup></b>
Mutation	0.494	<b>0.00417</b>
Reproduction by self or non-self (not specified)	0.462	<b>0.00773</b>
Purposiveness	0.361	<b>0.0423</b>
Order	0.149	0.415
Adapt to environment, Natural selection	-0.149	0.415

574

575 B)

Criteria	Correlation coefficient	P-value
Feed, Homeostasis	1.00	<b>0</b>
Movement, Protoplasm	0.939	<b>1.96*10<sup>-15</sup></b>
Cells	0.881	<b>3.03*10<sup>-11</sup></b>
Growth, Metabolism, Stimulus response, Regulatory mechanisms, Genetic material isolated from the environment	0.825	<b>6.52*10<sup>-9</sup></b>
Autocatalytic cycles, Enzymes	0.770	<b>2.50*10<sup>-7</sup></b>
Nucleic acids, Evolves, Replicate chemical information	0.511	<b>0.00278</b>
Mutation, Can die, Related	0.458	<b>0.00837</b>
Organic molecules, Adapt to environment, Natural selection	0.402	<b>0.0224</b>
Order	0.342	0.0551
Reproduction by self or non-self (not specified)	0.275	0.128

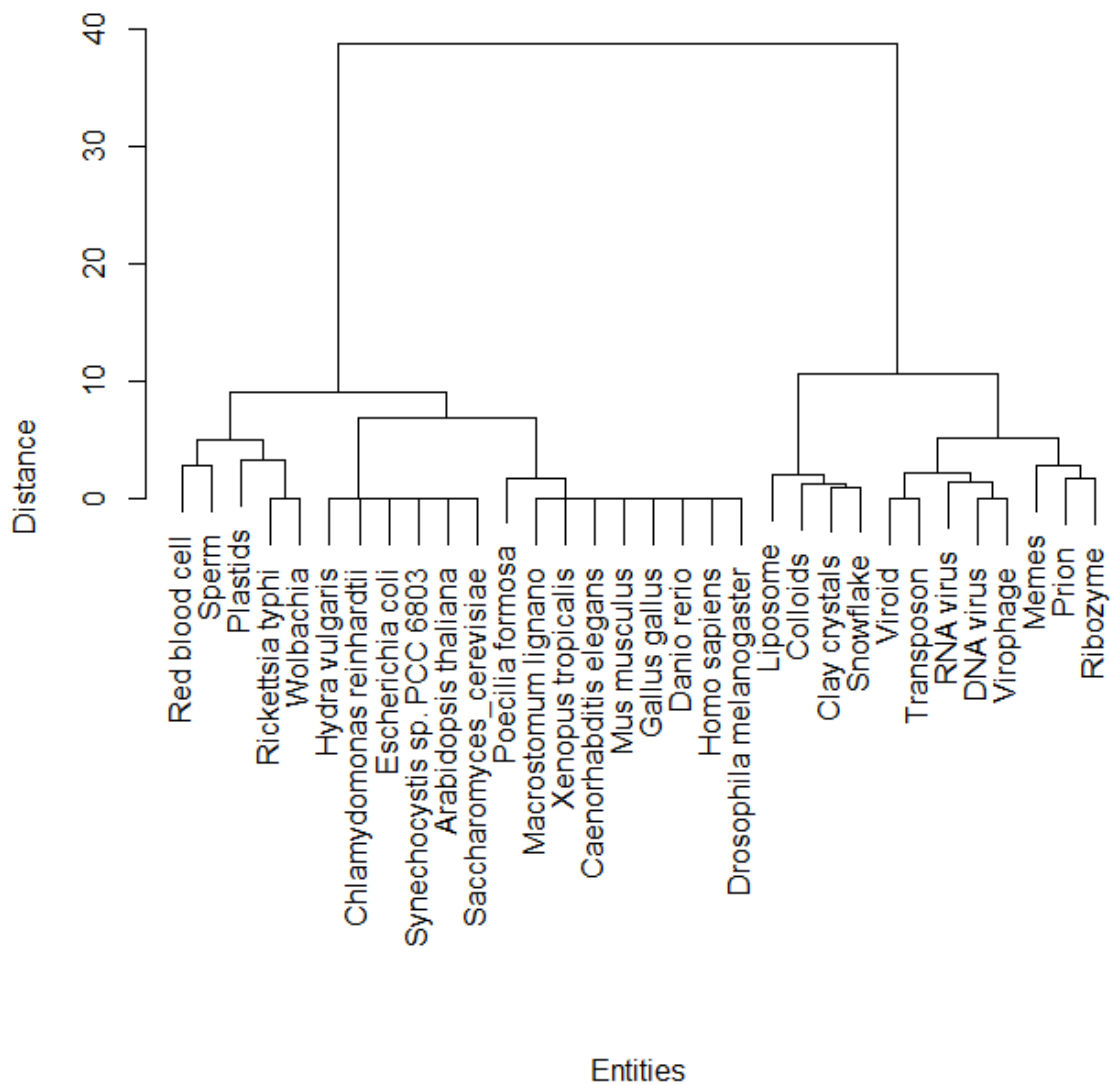
576



577

578 Figure 1: Spontaneous learning of the word "chair". After seeing multiple examples of a given type of  
579 object, we learn to identify the important properties associated with this type of object. This later  
580 allows us to recognize non-standard examples of chairs. (All images obtained from Wikimedia  
581 commons.)

582



583

584 Figure 2: Results of cluster analysis. Non-living entities generally cluster together at the right-hand  
 585 side of the plot, and living entities at the left-hand side of the plot. However both main clusters  
 586 include some ambiguous cases; red blood cells and sperm cluster together with intracellular parasite  
 587 species on the left-hand side of the plot, and viruses cluster together with non-living entities on the  
 588 right-hand side of the plot.

589