The taxonomic correction process: how often are taxonomic hypotheses accepted, rejected and then accepted again?

1. Abstract

2. Introduction

The goal of taxonomy is to provide a complete, accurate catalogue of planetary biodiversity. When taxonomists encounter a putative new species, they collect evidence to support the hypothesis that it is distinct enough from any known species to necessitate its own name. If so, this species is formally described and is given a new name under the appropriate codes of nomenclature [1,2]. Over 16,000 species have been described every year between 2000 and 2010 [3], and both the number of new descriptions and the number of authors involved in species description across multiple plant and animal groups have been rising since the 1750s, while the number of species described by each author has been falling [4,5]. Some authors have concluded that these observations indicate that more taxonomists are chasing fewer remaining species, suggesting that species description is approaching completion [6]. Even so, the taxonomic process is not yet complete: once those species have been described, an unknown proportion of them will be re-tested and, if falsified, may be rejected in favour of other hypotheses of conspecificity. The proportion of species hypotheses that will eventually be falsified may be expected to vary over time as techniques and species delimitation philosophies change, as later taxonomists have access to more evidence than those working previously, and as previously falsified hypotheses are retested in light of new evidence. Understanding how often these changes take place may allow us to estimate when all taxonomic work — both species description and redescription — will be completed. These changes can destabilise biodiversity databases that aggregate data from multiple sources on the basis of scientific name alone [7]. Corrections that change the circumscription of a name — its “taxon concept” *sensu* Franz *et al.* [8] — may occur without any corresponding change in its scientific name. Rates of taxonomic correction have previously been investigated at the nomenclature level [9], but changes in the accepted name are relatively easy for biodiversity databases to deal with. Estimating how many taxon concepts each name is associated may allow us estimate how much error is introduced to large analyses by ignoring these types of corrections, and which species are most likely to eventually need correction.

The question of which species will eventually need correction is particularly pertinent to a recent discussion of the causes of taxonomic inflation [10]. Taxonomic inflation is a sharp increase in the number of subspecies being raised to full species since the 1950s (often referred to as “splitting”) across a wide range of animal groups [11], including primates [10,12] , amphibians [13] , bovids [14] and birds [15]. One possible cause raised by Isaac *et al.* was the philosophical change from the biological species concept (BSC) to the phylogenetic species concept (PSC), but Sangster [15,16] has shown convincingly that there have been no wholesale change in the criteria used to delimit species boundaries in birds since the 1950s. Groves [12] reviewed primate taxonomy since the 1890s and found the PSC to have a positive effect on the understanding of taxonomic boundaries because of its insistence on testability and falsifiability. An alternate explanation for the current high rate of splitting was raised by Gill [17], who noted that taxonomists today are working on a “backlog of polytypic species awaiting attention”, which may have been created by wholesale merging of more finely circumscribed species into polytypic species “in the early days of the biological species concept [...] not based on strong biological evidence” ([18], quoted in [17]). Gill postulates that not all taxonomic corrections represent taxonomic progress; some proportion of them are reversions of previous corrections. If true, this process represents a wastage of scarce taxonomic resources and could increase the time necessary to reach an end-state of taxonomy, in which almost all species have been described and almost all circumscriptions have reached a final, stable circumscription. This corresponds to a special category of taxonomic corrections — the “recorrections” of existing corrections — that are currently understudied.

Studies of species description generally start with a dataset of all species described within a particular time period, as recorded by an index of species descriptions (e.g. [4]) or by taking lists of recognised species and identifying when and by whom they had been described (e.g. [5]). Neither of these approaches can be applied to taxonomic corrections: no widely used dataset of taxonomic changes exist, although datasets of nomenclatural changes are being built by individual publishers ($$$https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3088019/). Taxonomic changes are also widely distributed in both scientific literature and a variety of non-peer-reviewed publications, from scientific monographs to taxonomic checklists to general-interest identification guides. Previous analyses have surveyed a set of journals where taxonomic changes are likely to be published (e.g. [15,16]), but there is no easy way to determine if a particular proposal has gained widespread recognition within its taxonomic community. Conventional methods to gauge the impact of a publication, such as citations counts, cannot help us: a contentious proposal may be heavily cited by scientists disputing it, while a generally accepted proposal may only be cited a few times before being incorporated into field guides or taxonomic checklists, which may then be cited instead.

One source of taxonomic changes are both representative within a taxonomic group and are widely recognised are taxonomic checklists: expert-curated authoritative lists of recognized species within a taxonomic group in a particular geographical area. They are widely used by both scientists and non-scientists to distinguish species in field, museum and lab settings, and represent one of the most important outputs of taxonomy. In this study, we focussed on a single checklist that has been maintained over the last 127 years by the North American Classification Committee of the American Ornithologists' Union (AOU): the Check-List of North American birds, hereafter referred to as the "AOU Checklist". This checklist was first published in 1886, and has been maintained by the AOU through six major and fifty-seven minor updates over the subsequent 127 years [19]. Proposed changes are vetted by the Classification Committee; a two-thirds majority can accept proposals, which are published as a series of editions and supplements (http://checklist.aou.org/nacc/proposal\_guidelines.html). The checklist therefore provides a community review process for corrections and continues to be widely used as an authoritative source for taxonomic names among professional and amateur birders, either directly or indirectly through birding organisations and field guides track the AOU Checklist, such as the National Audubon Society ($$$http://www.audubon.org/news/here-are-biggest-changes-aou-checklist-north-american-birds), The Cornell Lab of Ornithology ($$$https://www.allaboutbirds.org/the-names-have-been-changed-inside-the-aou-check-list-committees-deliberations/), Sibley Guides ($$$http://www.sibleyguides.com/2011/07/new-bird-names-in-the-52nd-aou-checklist-supplement/), the American Birding Association ($$$http://blog.aba.org/2016/01/2016-aou-check-list-proposals-part-1.html) and National Geographic Field Guide to the Birds of Western North America ($$$https://books.google.com/books?id=5kVen0Hqgx0C&pg=PA6&lpg=PA6&dq=us+bird+field+guides+aou&source=bl&ots=gjTaOafU6M&sig=O2nRV5qNy8igzLnnysA0qVDaBEY&hl=en&sa=X&ved=0ahUKEwiW8suZkYTTAhUELSYKHTs8AaU4ChDoAQgZMAA#v=onepage&q=us%20bird%20field%20guides%20aou&f=false). Birds are generally believed to be a group that has been largely described (e.g. [20]), and North American birds have been studied for over 250 years [21]. Since this group was central to the development of both the biological species concept [22], the phylogenetic species concept [23] and Gill’s observations of recorrection patterns, they are a particularly apt group to begin studies of taxonomic change.

Our analysis focussed on changes that alter the circumscription of a scientific name without altering the name itself. These are of two kinds: the division of putative species into multiple species (“splits”), which usually occurs through the raising of a subspecies to a full species, and the union of putative species into polytypic species (“lumps”). A polytypic species as used here is simply one that contains several subspecies; we do not distinguish between polytypic species that are created through the description of new subspecies and those formed by lumping species. In order to better understand how post-description taxonomic circumscriptions change, we quantified several rates: we defined the “correction rate” as the proportion of recognized species that have ever been corrected, the “recorrection rate” as the proportion of recognized species that have been corrected more than once, and the “reversion rate” as the proportion of all corrections that partially reverted an earlier correction. In order to test whether post-1980s splitting is the direct result of pre-1980s lumping, we tested the null hypothesis that species lumped before 1980 are no more likely to be subsequently split than all species recognized in 1980: if Gill is right, we would expect the former proportion to be significantly higher than the latter. We also tested to see how many post-1980 splits reverted previous lumps, supporting Gill's supposition that increasing numbers of recognized species are in part being driven by the resplitting of previously lumped species.

These measurements will help us understand how often corrections, recorrections and reversions take place in this one localised group, but will have limited utility until this study can be replicated on other groups and the variability in these measurements can be assessed. However, it provides an essential first step to the broader examination of changing rates of lumping and splitting in this checklist over the 20th century, and the determination of factors for why some taxonomic groups may be more or less likely to be corrected than any other. Quantifying these patterns provides a more empirical basis to understanding trends in taxonomic work that have so far only been discussed qualitatively, and help to elucidate what potential end-states of taxonomic effort may look like.

3. Materials and Methods

**Source data.** The AOU Checklist consists of sixty-four checklists published between 1886 and 2016: seven major editions, which list every recognised species, and fifty-seven “supplements”, which list changes to the checklist since the previous supplement (S1 Table). We began with lists of additions, deletions and changes in scientific names to the AOU Checklist collected by one of the authors (DL) for checklists published between 1886 and 2012. These changes were collected as part of the online database Avibase [24], which also contains information on which taxon concepts are entirely contained within others [25]. Based on this information, we were able to identify additions and deletions that involve overlapping circumscriptions. We checked these changes against the AOU Checklists themselves to identify changes that were explicitly stated to be a lump or split in the publications; for instance, "...we divide *B[ranta] canadensis* by recognising a set of smaller-bodied forms as the species *B. hutchinsii*..." from [26]. This means that several putative lumps and splits were excluded from our analyses as we could not be certain that they were intended as such. For example, the 34th Supplement [27] includes the deletion of *Melanitta deglandi*, which Avibase records as having been lumped into *Melanitta fusca*. As this supplement consists solely of a list of recognized species, we have no evidence that this was indeed intended to be a lump and not a mere deletion; therefore, we conservatively assumed that the change had been made because of changes in species ranges and not taxonomic amendations. Our measures are therefore conservative counts that are likely smaller than the true values. For checklists between 2013 and 2016, we extracted the lumps, splits and name changes directly from the supplements themselves. We found 120 lumps and 185 splits.

**Geographic range changes in our data.** Our analysis was complicated by a dramatic change in the geographic range of the AOU Checklist in 1982 and 1983, expanding to include Mexico, the Hawaiian islands, the Caribbean Islands and Central America while removing species found only in Greenland. From approx. 789 species recognised in the 33rd Supplement (1976) [28], the number of recognized species rose to 933 species in the 34th Supplement (1982) [27] and to approx. 1,904 species in the 6th Edition (1983; all counts available in S1 Table). To obtain a consistent picture of taxonomic changes over as long a time period as possible, we eliminated all changes involving species first added to the checklist before 1982. This resulted in 114 unambiguous lumps and 89 unambiguous splits (S2 Table). The number of estimated recognised species after filtering changes varied from 771 (in 1886) to 977 (in 2015). We eliminated eleven checklists because no unambiguous lumps or splits took place in them (1894, 1899, 1909, 1912, 1920, 1953, 1957, 1982, 1983, 1991, 1998 and 2009). Large gaps in the data prevented us from using regression analysis, as we had no way of estimating what the counts would have been if it was spread more uniformly. Instead, we calculated the cumulative change in the number of lumps and splits to look at broad trends in the data (**Figure 1**). To identify the effect of the gaps, we also examined whether the size of a gap was correlated with the number of lumps or splits performed after it, which might suggest that taxonomic changes were “accumulating” over the period of the gap that were all accepted once the gap ended.

**Measuring taxonomic change rates.** To account for synonymy while measuring these rates, we assembled “name clusters” that link together species names that have been renamed. For example, *Acanthopneuste borealis*, *Phyllopseustes borealis* and *Phylloscopus borealis* all refer to the same species, which has been placed in several different genera since this species was first recognized by the AOU Checklist in 1886. Each name cluster refers to one or more taxon concepts, divided on the basis of lumps and splits. We built a list of these taxon concepts, along with the checklists in which they were first added, all the checklists that explicitly referred to this taxon concept, and (optionally) the checklist in which this concept was lumped or split.

We measured the correction rate as the proportion of currently recognised name clusters (after removing those added after 1981) that had more than one or more taxon concepts associated with it, and the recorrection rate as the proportion that have two or more taxon concepts associated with it. We measured the reversion rate as the proportional of all lumps and splits that partially undid a previous split or lump. For example, *Rallus obsoletus* was split from *Rallus longirostris* in 1890 (CITE), but was then lumped back into it in 1944 (CITE). We also measured a "perfect revisionary rate", in which a lump is paired with a split that perfectly reverts the change made earlier (S2 Table).

To test whether the increase in recognised species is caused by the resplitting of previously lumped species, we compared the probability of a split being relumped with the probability of a lump being resplit. We can compare this with the ratio of observed lumps to observed splits to determine if splits are more or less likely to be lumped than we would expect.

**Which species are most likely to be lumped or split?** To determine whether different taxonomic groups showed significantly different amendment rates, we modelled the number of taxonomic changes (lumps + splits) involving currently recognized name clusters (as of [19]) as a Poisson distribution, in which the rate at which new circumscriptions are added to existing names is assumed to be constant within a taxonomic group. Our model had three hierarchical levels of grouping: at the level of genus (*π*), family (*τ*) and order (*ρ*). Additionally, we included an offset to account for the different lengths of time that that different species have been in the checklist. Our hierachical model can be described as:

Each of these parameters were modeled as normally distributed random variables, with a mean of zero and with variable standard deviations (*σπ*, *στ* and *σρ* respectively). *ti* is the number of years that this species has been recognized in the AOU Checklist, to control for some species having had a longer time span within which to be lumped or split than others. This model failed to converge in rSTAN 2.12.1 [29], and so we used transformed parameters to define standard normal deviations that were multipled by the variable standard deviations (see S7 Code). This model converged successfully in rSTAN and gave us an estimate of the overall mean rate as well as the mean rate for every genus, family and order (S4-S6 Tables).

4. Results

**Overall trends in lumping and splitting.** We discovered 114 unambiguous lumps and 89 unambiguous splits on species clusters added before 1982. We could not use linear regression to extrapolate between our data points since there is no way of knowing how many lumps or splits would have taken place had a supplement been published in a particular year. Examining the lumping and splitting directly (**Figure 1**) shows a general pattern of many small changes, divided by gaps of different lengths. Particularly long gaps in lumps and splits can be seen in the 1910s and the 1960s. Individual checklists can large numbers of lumps, in particular the 30 lumps in the 19th supplement in 1931 ($$$), 19 lumps in the 4th edition in 1931 ($$$), and the 15 lumps in 32nd supplement in 1973 ($$$). While there are no specific spikes in the number of splits, most (68, or 80%) of the splits in our dataset took place after 1980.

To look at temporal trends in the data, we calculated the cumulative number of lumps and splits over time. These follow the same pattern we have already seen: most of the lumping took place between 1920 and 1980, while most of the splitting took place between 1980 and 2016. Lumping appears to have all but ceased since 1980. To summarise, we appear to see a period of lumping followed by a period of splitting. Based on this data, we cannot forecast whether lumping has permanently ceased (i.e. only splitting remains), or whether this is a repeating pattern (i.e. the current period of splitting will be followed by an additional period of lumping).

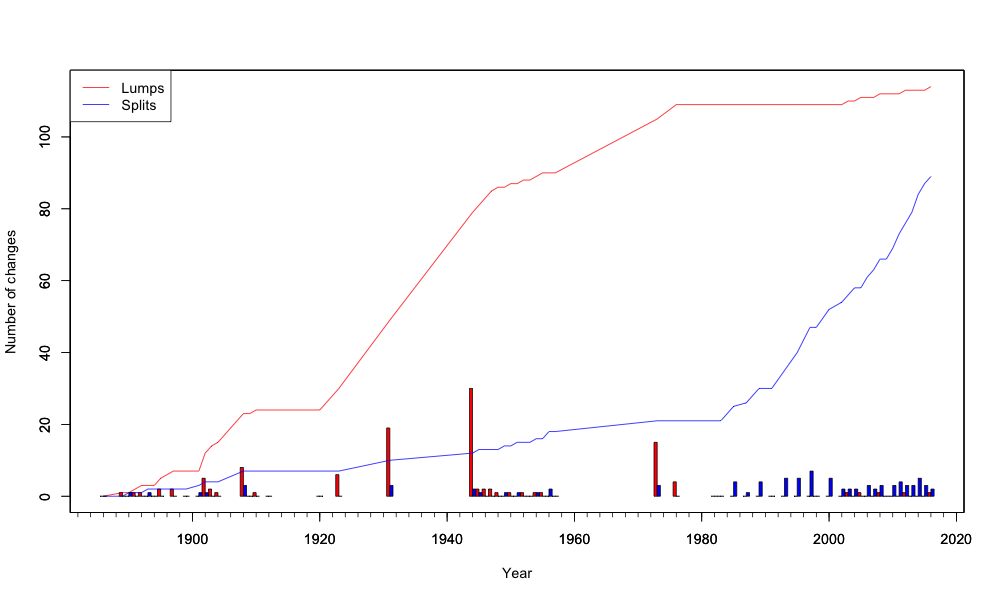


Figure 1 Cumulative lumps and splits, showing trends in taxonomic changes within the AOU Checklist between 1886 to 2016.

**Analysis of gaps.** We found that the length of the gap in publication before a checklist could predict the number of lumps made in that checklist () but not the number of splits (). This is likely an artifact of the pattern of lumping and splitting: the period of maximal lumping has many large gaps in them, while the period of maximal splitting has few gaps, most of 1-2 years. It does suggest that the spikes in lumping we observe between 1900 and 1950 might have been lower if checklists had been published more consistently, and so are a result of the gaps, not a cause of the gaps.

**Measuring taxonomic change rates.** Our dataset contained a total of 3,041 taxon concepts for 2,393 name clusters (S3 Table). Of these, 1,213 taxon concepts refer to the 974 name clusters that are currently recognised *($$$ This is a discrepancy of two species from 976 recognised species! Why?)*. 180 name clusters (18.48%) were corrected at least once (18.48%), while 49 name clusters were corrected two or more times (5.03%). To summarise, almost one in five currently recognised species are associated with at least two circumscriptions, while one in twenty currently recognised species are associated with more than two circumscriptions.

The correction processes can also be examined from the perspective of lumps and splits. Of the 114 lumps and 89 splits in our dataset, 21 lumps (18.42%) and 20 splits (22.47%) were perfectly reverted. In some cases, this occurred several times: for example, *Branta nigricans* was lumped into *Branta bernicla* in 1944, then was resplit in 1951, and then was relumped in 1976: we count two reverted lumps and one reverted split from this one series of changes. We identified 19 series of changes involving lumps and splits being reverted. We also counted partial reversions, in which one change (say, a lump) lumps two or more species that are later split in another change (in this case, a split), even if either the original or later change involves other species. We found 39 lumps (34.21%) and 36 splits (40.45%) that had at least one other change that partially reverts it, or in which the partial change was itself reverted.

We compared the different patterns of perfect reversions we observed. We eliminated one instance of a species being lumped twice (*Trogon ambiguus* being lumped into *Trogon elegans* in both $$$aou\_4\_19 and $$$aou\_4\_22), one instance of a species being lumped, split and then lumped again, and two instances of a species being split, lumped and then split again. Of the remaining 15 perfect reversions, 12 consisted of a lump followed by a split, while 3 consisted of a split followed by a lump. If our null hypothesis is that a lump should be as likely to be undone as a split, we would expect to see them in the same ratio as we see lumps and splits generally, i.e. 114 lumps to 89 splits. We do not see a significant difference in these ratios (Fisher’s Exact Test, ).

**Which species are most likely to be lumped or split?**

5. Discussion

6. Acknowledgments

7. Supplementary materials

1. S1 Table: List of supplements, counts of recognised species and changes
2. S3 Table: Species definitions

3. S4-S6 Table: Reserved for higher taxonomy calculation

4. S7 Code: Software code for this project

7.1.

8. Cover Letter

9. Notes

10. Text dump

11. Bibliography

1. Ride WDL, Cogger HG, Dupuis C, Kraus O, Minelli A, Thompson FC, et al., editors. International Code of Zoological Nomenclature [Internet]. International Trust for Zoological Nomenclature; 1999. Available: http://www.worldcat.org/isbn/9780853010067

2. McNeill J, Barrie FR, Buck WR, Demoulin V, Greuter W, Hawksworth DL, et al. International Code of Nomenclature [Internet]. 1 Jan 2012. Available: http://www.iapt-taxon.org/nomen/main.php

3. IISE. Retro SOS 2000-2009: A decade of species discovery in review [Internet]. 1 Jan 2011. Available: http://www.esf.edu/species/SOS.htm

4. Tancoigne E, Dubois A. Taxonomy: no decline, but inertia. Cladistics. 2013;29: 567–570.

5. Joppa LN, Roberts DL, Pimm SL. The population ecology and social behaviour of taxonomists. Trends in Ecology & Evolution. 2011;26: 551–553.

6. Costello MJ, Wilson S, Houlding B. More taxonomists describing significantly fewer species per unit effort may indicate that most species have been discovered. Systematic Biology. 2013;62: 616–624. doi:10.1093/sysbio/syt024

7. Remsen D. The use and limits of scientific names in biological informatics. ZooKeys. 2016;550: 207–223. doi:10.3897/zookeys.550.9546

8. Franz N, Peet R, Weakley A. On The Use Of Taxonomic Concepts In Support Of Biodiversity Research And Taxonomy. 2008.

9. Alroy J. How many named species are valid? Proceedings of the National Academy of Sciences of the United States of America. 2002;99: 3706–3711. doi:10.1073/pnas.062691099

10. Isaac NJB, Mallet J, Mace GM. Taxonomic inflation: its influence on macroecology and conservation. Trends in Ecology & Evolution. 2004;19: 464–469.

11. Agapow P-M, Bininda-Emonds ORP, Crandall KA, Gittleman JL, Mace GM, Marshall JC, et al. The Impact of Species Concept on Biodiversity Studies. The Quarterly Review of Biology. 2004;79: 161–179.

12. Groves CP. Primate Taxonomy: Inflation or Real? Annual Review of Anthropology. 2014;43: 27–36.

13. Padial JM, Riva ID la. Taxonomic inflation and the stability of species lists: the perils of ostrich’s behavior. Systematic Biology. 2006;55: 859–867.

14. Heller R, Frandsen P, Lorenzen ED, Siegismund HR. Are there really twice as many bovid species as we thought? Systematic Biology. 2013;62: 490–493.

15. Sangster G. The application of species criteria in avian taxonomy and its implications for the debate over species concepts. Biological Reviews. 2014;89: 199–214.

16. Sangster G. Increasing numbers of bird species result from taxonomic progress, not taxonomic inflation. Proceedings of the Royal Society of London B. 2009;276: 3185–3191.

17. Gill FB. Species taxonomy of birds: Which null hypothesis? The Auk. 2014;131: 150–161.

18. American Ornithologists’ Union. Check-List of North American Birds. [s.l.] : American Ornithologists’ Union; 1889.

19. Chesser RT, Burns KJ, Cicero C, Dunn JL, Kratter AW, Lovette IJ, et al. Fifty-seventh Supplement to the American Ornithologists’ UnionCheck-list of North American Birds. The Auk. 2016;133: 544–560.

20. Bebber DP, Marriott FHC, Gaston KJ, Harris SA, Scotland RW. Predicting unknown species numbers using discovery curves. Proceedings of the Royal Society of London B. 2007;274: 1651–1658.

21. Catesby M. The natural history of Carolina, Florida and the Bahama Islands : containing the figures of birds, beasts, fishes, serpents, insects, and plants : particularly, the forest-trees, shrubs, and other plants, not hitherto described, or very incorrectly figured by author. 1731.

22. Mayr E. Systematics and the origin of species from the viewpoint of a zoologist. New York : Columbia University Press; 1942.

23. Cracraft J. Species Concepts and Speciation Analysis. Multilocus species delimitation in a complex of morphologically conserved trapdoor spiders (mygalomorphae, antrodiaetidae, aliatypus). Boston, MA: Springer US; 1983. pp. 159–187.

24. Lepage D. Avibase [Internet]. 2016.

25. Lepage D, Vaidya G, Guralnick R. Avibase a database system for managing and organizing taxonomic concepts. ZooKeys. 2014;420: 117–135.

26. Banks RC, Cicero C, Dunn JL, Kratter AW, Rasmussen PC, Remsen JV, et al. Forty-Fifth Supplement to the American Ornithologists’ Union Check-List of North American birds. The Auk. 2004;121: 985–995.

27. Eisenmann E, Monroe BL, Parkes KC, Short LL, Banks RC, Howell TR, et al. Thirty-Fourth Supplement to the American Ornithologists’ Union Check-List of North American Birds. The Auk. 1982;99: 1CC–16CC.

28. Eisenmann E, Parkes KC, Banks RC, Lowery GH, Howell TR, Monroe BL, et al. Thirty-Third Supplement to the American Ornithologists’ Union Check-List of North American Birds. The Auk. 1976;93: 875–879.

29. Team SD. \*RStan: the R interface to Stan\*.