



**HAL**  
open science

## Theory Meets Empiry: A Citation Network Analysis

Courtney Fitzpatrick, Elizabeth Hobson, Tamra Mendelson, Rafael Rodríguez, Rebecca Safran, Elizabeth Scordato, Maria Servedio, Caitlin Stern, Laurel Symes, Michael Kopp

► **To cite this version:**

Courtney Fitzpatrick, Elizabeth Hobson, Tamra Mendelson, Rafael Rodríguez, Rebecca Safran, et al.. Theory Meets Empiry: A Citation Network Analysis. *Bioscience*, 2018, 68 (10), pp.805-812. 10.1093/biosci/biy083 . hal-02075739

**HAL Id: hal-02075739**

**<https://hal.science/hal-02075739>**

Submitted on 7 Mar 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Theory meets empiry: a citation network analysis.

Journal:	<i>BioScience</i>
Manuscript ID	17-0163.R3
Manuscript Type:	Professional Biologist
Date Submitted by the Author:	n/a
Complete List of Authors:	<p>Fitzpatrick, Courtney; Indiana University, Biology          Hobson, Elizabeth; Santa Fe Institute, Santa Fe Institute          Mendelson, Tamra; University of Maryland Baltimore County, Department of Biological Sciences          Rodriguez, Rafael; University of Wisconsin-Milwaukee, Department of Biological Sciences          Safran, Rebecca; University of Colorado, Department of Ecology and Evolutionary Biology          Scordato, Elizabeth; University of Colorado, Department of Ecology and Evolutionary Biology; California State Polytechnic, Biological Sciences          Servedio, Maria; University of North Carolina at Chapel Hill, Department of Biology          Stern, Caitlin; Santa Fe Institute, Santa Fe Institute          Symes, Laurel; Dartmouth College, Biology; University of Wisconsin - Milwaukee, Biology          Kopp, Michael; Aix Marseille University, CNRS</p>
Key words:	science of science, sexual selection, speciation, citation network, theory
Abstract:	<p>According to a recent survey, ecologists and evolutionary biologists feel that theoretical and empirical research should coexist in a tight feedback loop, but believe that the two domains actually interact very little. We evaluate this perception using a citation network analysis for two datasets, representing the literature on sexual selection and speciation. Overall, 54-60% of citations come from a paper's own category, whereas 17-23% are citations across categories. These cross-citations tend to focus on highly-cited papers, and we observe a positive correlation between the numbers of citations a study receives within and across categories. We find evidence that reviews can function as integrators between the two literatures, argue that theoretical models are analogous to specific empirical study systems, and complement our analyses by studying a co-citation network. We conclude that theoretical and empirical research are more tightly connected than generally thought, but that avenues exist to further increase this integration.</p>

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

SCHOLARONE™  
Manuscripts

Draft Manuscript

1  
2  
3  
4 1 **Title: Theory meets empiry: a citation network analysis.**

5  
6  
7 2 **Authors**

8  
9 3 Courtney L. Fitzpatrick (corresponding author), Department of Biology, Indiana  
10 4 University, Bloomington, IN 47405, USA, email: [cfitzpa@indiana.edu](mailto:cfitzpa@indiana.edu),  
11 5 twitter: [@clynfitzpatrick](https://twitter.com/clynfitzpatrick)

12  
13 6 Elizabeth A. Hobson, Santa Fe Institute, Santa Fe, NM 87501, USA

14  
15 7 Tamra C. Mendelson, Department of Biological Sciences, University of Maryland  
16 8 Baltimore County, Baltimore, MD 21250, USA

17  
18 9 Rafael L. Rodríguez, Behavioral & Molecular Biology Group, Department of Biological  
19 10 Sciences, University of Wisconsin-Milwaukee, WI 53201, USA

20  
21 11 Rebecca J. Safran, Department of Ecology and Evolutionary Biology, University of  
22 12 Colorado, Boulder, CO 80309, USA

23  
24 13 Elizabeth Scordato, Department of Ecology and Evolutionary Biology, University of  
25 14 Colorado, Boulder, CO 80309, USA

26  
27 15 Now at: Department of Biological Sciences, California State Polytechnic  
28 16 University, Pomona, CA, 91768

29  
30 17 Maria R. Servedio, Department of Biology, University of North Carolina at Chapel Hill,  
31 18 Chapel Hill, NC 27514, USA

32  
33 19 Caitlin A. Stern, Santa Fe Institute, Santa Fe, NM 87501, USA

34  
35 20 Laurel Symes, B. Department of Psychological and Brain Sciences, Dartmouth College,  
36 21 New Hanover, NH 03755, USA

37  
38 22 Michael Kopp, Aix Marseille University, CNRS, Centrale Marseille, I2M, Marseille,  
39 23 France

40  
41  
42  
43  
44  
45 26 **Abstract**

46  
47  
48 27 According to a recent survey, ecologists and evolutionary biologists feel that theoretical  
49 28 and empirical research should coexist in a tight feedback loop, but believe that the two  
50 29 domains actually interact very little. We evaluate this perception using a citation network  
51 30 analysis for two datasets, representing the literature on sexual selection and speciation.

52  
53 31 Overall, 54-60% of citations come from a paper's own category, whereas 17-23% are

1  
2  
3 32 citations across categories. These cross-citations tend to focus on highly-cited papers,  
4 33 and we observe a positive correlation between the numbers of citations a study receives  
5 34 within and across categories. We find evidence that reviews can function as integrators  
6 35 between the two literatures, argue that theoretical models are analogous to specific  
7 36 empirical study systems, and complement our analyses by studying a co-citation  
8 37 network. We conclude that theoretical and empirical research are more tightly connected  
9 38 than generally thought, but that avenues exist to further increase this integration.

14  
15 39 **The ideal : a union between ‘theory’ and ‘empiry’.**

16  
17 40 Lasting advances in evolutionary biology and ecology often emerge from the  
18 41 complementarity of theoretical and empirical perspectives. Theoretical models use  
19 42 mathematics as a tool to formalize hypotheses, thereby distilling biological complexity  
20 43 down to the components that are thought to be most significant. Models can also  
21 44 increase our understanding of processes that are obscured because they happened in  
22 45 deep time or because they span many human generations, and can unify observations  
23 46 across diverse groups of organisms.

24  
25  
26  
27  
28 47  
29  
30 48 Of course, theoretical biology is only meaningful in so far as it relates—even if only as an  
31 49 abstraction—to real, observable phenomena. Empirical observations and experiments  
32 50 are essential to provide a foundation of realistic assumptions that can form the basis of  
33 51 theoretical studies. Likewise, controlled empirical experiments (analogous to the formal  
34 52 precision of theoretical research) tease signal from noise to identify the real drivers of  
35 53 biological processes. That is, the theoretical and the empirical approach each yields  
36 54 unique insights, and the integration of both is required for a full understanding of any  
37 55 biological phenomenon.

38  
39  
40  
41  
42 56  
43  
44 57 Because the word ‘theory’ is used in a variety of ways across disciplines—and even  
45 58 within disciplines—it is worth specifying our usage of the word. We are using the word  
46 59 theory in what has become its contemporary meaning within the fields of ecology and  
47 60 evolutionary biology. That is, we focus on the type of research that uses mathematics or  
48 61 computer simulations to represent and investigate biological processes.

49  
50  
51  
52 62  
53  
54 63 Indeed, the synergy between theoretical and empirical approaches (“theory” and  
55 64 “empiry”; we co-opt the latter word from German) has a long history in ecology and

1  
2  
3 65 evolutionary biology, which runs the gamut from productive and agreeable cooperation  
4 66 to an uneasy marriage fraught with rivalry and conflict. For example, Darwin's empirical  
5 67 observations and verbal models gave way to mathematical formulations by Fisher,  
6 68 Haldane and Wright, which, in turn, have inspired decades of empirical research. In the  
7 69 study of sexual selection, the classical empirical and theoretical literature made equally  
8 70 important contributions to a concrete understanding of how competition for mates  
9 71 shapes extravagant phenotypes (see Bennet 2000 for communication between them  
10 72 that is both productive and contentious). Thus, both theory and empiry are involved in  
11 73 the testing of hypotheses (Servedio et al. 2014). Theory's role is to make sure that  
12 74 hypotheses, especially complex ones, are logically consistent (i.e., what follows logically  
13 75 from a set of assumptions), whereas empiry tests whether they adequately explain  
14 76 nature.  
15 77

16 78 A recent survey confirms that ecologists and evolutionary biologists share the view that  
17 79 integration should exist between theoretical and empirical research (Haller 2014). An  
18 80 overwhelming majority of both self-identified theoreticians and self-identified empiricists  
19 81 reported that they believe theoretical and empirical work should "coexist in a tight  
20 82 feedback loop" (Haller 2014), with each one informing the other. However, results from  
21 83 this same study suggest that the community perceives a collective failure to achieve its  
22 84 reported ideal. While a small minority believe theory and empiry already coexist in a  
23 85 feedback loop, most reported that they believe the two approaches interact very little.  
24 86 This disconnect between the shared ideal (theory and empiry should be integrated) and  
25 87 the shared perception of reality (an absence of that integration) invites further  
26 88 investigation.

### 27 89 **Reality check: a citation network analysis**

28 90 A quantifiable – though obviously incomplete – metric of communication among  
29 91 scientists is provided by citations between published articles. Here, we quantify the  
30 92 connections between theoretical and empirical research by means of a citation network  
31 93 analysis. By way of example, we focus on a specific area of research that has an  
32 94 established history of drawing from both theoretical and empirical approaches: the  
33 95 literature on sexual selection and its relationship to speciation. We started out by  
34 96 generating two data sets: one that represents a sample of the general sexual-selection  
35 97 literature since the 1970s (SS), and a second one, more narrowly defined and more

1  
2  
3 98 complete, that reflects the increasing interest in sexual selection and speciation over the  
4 99 past 20 years (SS & S). A detailed description of our sampling approach and other  
5 100 methods is provided in Appendix 1. We then assigned all papers in the two datasets to  
6 101 one of four categories: “Empirical” (E), “Theoretical” (T), “Mixed” (M) and “Reviews and  
7 102 Ideas” (R); for detailed definitions, see Appendix 1. Finally, we constructed citation  
8 103 networks, in which nodes stand for published papers and directed links represent  
9 104 citations going from a cited paper (the “source”) to the citing paper (the “target”; see  
10 105 again Appendix 1).

## 106 **Basic results**

107 In both data sets, the number of citations received by individual papers—both locally  
108 (i.e., within the coded network) and globally (numbers taken directly from Web of  
109 Science; see Appendix 2 for some results from the global network)—is heavily biased  
110 towards a relatively small number of highly-cited papers (Fig. S2, S3). Although R  
111 papers tend to get more citations than both E and T papers, overall median citation  
112 numbers do not differ strongly between the three types. In particular, T papers receive at  
113 least as many citations as E papers (and, as we shall see below, a large proportion of  
114 their citations comes from empirical studies). These results do not change qualitatively  
115 when citation numbers are calculated on a per-year basis (results not shown). Basic  
116 citation patterns within and between categories are visualized in Fig. 1 and summarized  
117 by the following dyadic citation frequencies (see Appendix 1 for details of the  
118 calculation):

119  
120 **EMPIRY:** Of all citations in the coded Sexual Selection (SS) network made by E articles,  
121 the majority (54%) were of other E articles, while 22% were of R articles, and 20% were  
122 of T articles. Citations in the Sexual Selection and Speciation (SS & S) network by  
123 empirical research were distributed similarly: 59% were of other E articles, 23% were of  
124 R articles, and 17% were of T articles.

125  
126 **REVIEW:** Citation frequencies by R papers were more evenly distributed across article  
127 types as compared to citation frequencies by E and T papers. Of all citations in the SS  
128 network made by R papers, 40% were of E articles, 33% were of other R papers and  
129 24% were of T articles. A similar pattern was found in the SS & S data set; of the  
130 citations made by R papers, 48% were of E articles, 23% were of other R papers, and

1  
2  
3 131 28% were of T articles. (See Table S1 and Fig. 1 for full summary of dyadic citation  
4 132 frequencies.)

5  
6 133

7  
8 134 THEORY: Of all citations in the SS network made by T papers, the majority (56%) were of  
9 135 other T articles, while 21% were of R articles and 19% were of E articles. A similar  
10 136 pattern was found in the SS & S data set; of the citations made by T papers, 59% were  
11 137 of other T papers, 17% were of R papers, and 23% were of E articles.

12  
13 138

14  
15 139 MIXED: Of all citations in the SS network made by M papers, 33% were of E articles,  
16 140 while 24% were of R articles and 32% were of T articles. In the SS & S data set 42% of  
17 141 the citations made by M papers were of E papers, while 28% of the citations were  
18 142 allocated both to T and R papers.

19  
20 143

21  
22 144 The above frequencies differ somewhat from year to year, but we found no clear trends  
23 145 over time (Fig. S4), the only exception being E to E citations that increase over time,  
24 146 reflecting the overall increase in the number of published empirical studies (see Fig. S1).

25  
26 147

### 27 28 29 30 31 148 **Seven hypothesis about citation patterns**

32  
33 149 Citations within and between theoretical and empirical studies might be influenced by a  
34 150 number of causal factors, which might, in turn, generate a number of patterns. In the  
35 151 following, we discuss these as a set of non-mutually exclusive hypotheses and put them  
36 152 forward as potential explanations, though not necessarily as an exhaustive list (see  
37 153 Table 1).

38  
39 154

#### 40 41 42 43 155 *The lack-of-integration hypothesis (i)*

44  
45 156 The first hypothesis states that researchers cite primarily within their own category, such  
46 157 that empiry cites mostly empiry and theory cites mostly theory. This is the hypothesis  
47 158 that most closely reflects the perception reported by Haller (2014). Indeed, our results  
48 159 are consistent with this prediction (Fig. 1). However, this pattern is perhaps not  
49 160 surprising; some degree of increased citation frequency within-category should be  
50 161 expected, pointing to a question that, although at the heart of our study, is subjective and  
51 162 therefore empirically unanswerable. That is, what level of citation between categories



1  
2  
3 163 reflects an ideal integration? We do not propose to have an answer to that question, and  
4 164 instead, will discuss our results in the context of additional factors (hypothesis ii-vii) that  
5 165 have the potential to influence observed citation patterns. Subsequently, we add some  
6 166 insights gleaned from a co-citation network analysis, and close by highlighting both the  
7 167 ways that theory and empiry “talk” to each other as well as the ways in which that  
8 168 communication might be enhanced.

13  
14 169 *The “null” expectation hypothesis (ii)*

15  
16 170 Researchers might cite in proportion to the frequency of papers from each category in  
17 171 the literature. This represents something of a “null” hypothesis and predicts, for instance,  
18 172 that if 70% of published papers are empirical (as in the SS dataset), then 70% of  
19 173 citations from both theory and empiry would be of empirical research. However, this is  
20 174 not what we found in our datasets. While there is no *a priori* reason to view the null  
21 175 hypothesis as a reasonable expectation, one especially marked deviation merits  
22 176 discussion: the citations made by the theoretical literature. That is, theory cited theory at  
23 177 a substantially higher rate than it appeared in the literature and empiry at a lower rate  
24 178 than it appeared in the literature. While this is, of course, consistent with hypothesis (i),  
25 179 two other potential explanations are given by the “size-of-the-community” and the  
26 180 “theory-as-system” hypotheses, which we discuss next.

33  
34 181 *The “size of the community” hypothesis (iii)*

35  
36 182 Citation patterns might be influenced by the size of the research community. In  
37 183 particular, smaller communities might be more able to cite their own type exhaustively,  
38 184 because the number of papers is not overwhelming. In this scenario, citing exhaustively  
39 185 within a smaller community is feasible, and omitted citations are more noticeable.  
40 186 Because theoretical biology is a smaller community than empirical biology (as reflected  
41 187 by the difference between the number of theoretical and empirical papers in our data  
42 188 set), this hypothesis predicts that theory would cite theory at higher rates than it is  
43 189 represented in the literature. Indeed, this pattern is what we found.

44 190  
45 191 Moreover, the theory-biased citation pattern within theoretical research might result not  
46 192 only from the effect of working within a smaller community, but from an interaction  
47 193 between disparity in abundance (empiry is represented more heavily than theory) and  
48 194 familiarity (theoreticians are more familiar with theory and empiricists with empiry). In this

1  
2  
3 195 scenario, theoretical biologists are presented with a double obstacle; not only do they,  
4 196 like empiricists, face the challenge of learning material that is outside their area of  
5 197 expertise, but the unfamiliar body of literature (empiry) for theoreticians is many times  
6 198 larger than the unfamiliar body of literature (theory) for empiricists. This might effectively  
7 199 create a higher threshold for theoretical biologists than empiricists when it comes to  
8 200 learning enough about the other approach to reference it comprehensively. Here there  
9 201 would seem to be a great opportunity for reviews of the empirical literature in an effort to  
10 202 invite theoretical treatment of contemporary empirical patterns or puzzles (see below for  
11 203 further discussion of how reviews might facilitate integration).

12  
13  
14  
15  
16  
17  
18 204 *The “theory-as-system” hypothesis (iv)*

19  
20 205 Another potential explanation for the theory-biased citations of theory might be that  
21 206 theoretical approaches to a given research question are similar to work on that question  
22 207 in a specific empirical system. From this perspective, theory is analogous to research  
23 208 within a unique study system like *Drosophila* or sticklebacks. Although these research  
24 209 communities employ specific biological systems to examine a range of broader research  
25 210 questions, a given study will necessarily cite other research from that same system at a  
26 211 disproportionate rate in order to provide appropriate background and context for the  
27 212 specific work being presented. Extending this analogy, theoretical papers may cite other  
28 213 theoretical papers disproportionately.

29  
30  
31  
32  
33  
34 214  
35  
36 215 Indeed, of all the citations made by cichlid papers in the coded SS & S network, 68%  
37 216 were of other cichlid papers (1808 out of 2650). For the three other most frequently  
38 217 studied systems in that dataset, approximately 40% of the citations were from the same  
39 218 system (*Drosophila*: 228/538 citations, sticklebacks: 217/516 citations, *Heliconius*:  
40 219 48/121 citations; see Appendix 1 for details). These percentages are similar to the 56%  
41 220 (SS) and 59% (SS & S) at which T papers cited other T papers in our datasets, and are  
42 221 thus consistent with the idea that the specialization of a paper, whether on study system  
43 222 or methodology, means that a disproportionate number of the citations (around 50%) will  
44 223 be of papers with the same specialization.

45  
46  
47  
48  
49  
50  
51 224 *The “reviews-as-proxies” hypothesis (v)*

52  
53  
54 225 Researchers from each of the two main categories (empiry and theory) might use  
55 226 reviews as proxies for the primary literature from the other one. In other words, empiry

1  
2  
3 227 might cite reviews of theory *in lieu* of theory itself and vice versa. To investigate this  
4 228 hypothesis, we analyzed “indirect citation networks” in which papers are linked when one  
5 229 is cited by a review that is cited by the other paper (see Appendix 1). We found that  
6 230 reviews cited by E papers cite a higher proportion of T papers than E papers cite  
7 231 themselves, and R papers cited by T papers likewise cite a higher proportion of E papers  
8 232 than T papers cite themselves. As a consequence, patterns of indirect citations  
9 233 (citations by R papers that are cited either by E or T papers) differ less between types of  
10 234 target papers than do patterns of direct citations. Furthermore, T papers are  
11 235 overrepresented in indirect citations relative to their frequency in both datasets. (See Fig  
12 236 S5 for full comparisons).

13  
14  
15  
16  
17  
18  
19  
20 237 The above analysis provides some support for the idea that both empirical and  
21 238 theoretical studies cite reviews in lieu of original articles. Indeed, the purpose of a review  
22 239 is to consolidate information. Theory, in particular, might be more likely to cite reviews  
23 240 than the original empirical articles for the reasons outlined above (see the “size of the  
24 241 community” and the “theory-as-systems” hypotheses), but also because they highlight  
25 242 broad patterns that emerge from many individual studies. If the primary goal of  
26 243 theoretical studies is to investigate the processes that give rise to general patterns,  
27 244 review articles may provide better context than individual empirical studies because  
28 245 effective reviews summarize the broad patterns that are collectively demonstrated by  
29 246 empirical studies. Indeed, indirect citations of empirical work by theoretical studies were  
30 247 closer to the distribution expected under hypothesis (ii) (i.e., the representation in the  
31 248 literature), and the disparity between theoretical citations of theory and empiry was  
32 249 reduced. This analysis is consistent with the idea that reviews may function to integrate  
33 250 the theoretical and empirical domains by grouping relevant pieces of research together,  
34 251 which may, in turn, convert an unmanageable task (learning a large body of unfamiliar  
35 252 literature by reading each incremental contribution) into a manageable one. This  
36 253 hypothesis is further supported by our finding that, of all the categories of papers,  
37 254 reviews are globally cited most frequently (see top panels of Figures S2 and S3). Finally,  
38 255 the results from the co-citation network analysis presented below suggest that,  
39 256 especially in the SS data set, reviews often bridge otherwise disconnected groups of  
40 257 literature together, as would be the case if researchers were relying on reviews as  
41 258 representatives of collections of original research articles. Thus, review articles may play  
42 259 an important role in integrating across approaches.

1  
2  
3 260 *The “concentrated cross-citation” hypothesis (vi)*  
4

5 261 Our hypothesis (vi) states that cross-citations between empiry and theory might be  
6 262 concentrated on a small number of high-impact source papers, while the majority of  
7 263 papers from each category remains uncited by the other category. To investigate this  
8 264 hypothesis, we first calculated the number of papers from the E, R and T categories that  
9 265 are cited by at least one paper from each of these categories (see Venn diagrams in Fig.  
10 266 S6). The results show, in particular, that a large proportion of theoretical papers receive  
11 267 citations from empirical papers: in the SS dataset, 65 out of 94 theory papers (69%) are  
12 268 cited by E papers from the coded network; similarly, in the SS & S dataset, 90 out of 162  
13 269 T papers (58%) receive citations from E. In contrast, only 10% of E papers in SS (72 out  
14 270 of 713) and 14% in SS & S (138 out of 998) are cited by T papers. This asymmetry  
15 271 reflects, at least in part, the much greater number of E papers in the networks, but likely  
16 272 also the fact that many specialized E studies have low chances of getting read by  
17 273 theoreticians. Note that no such asymmetry exists for target papers: As shown in Fig.  
18 274 S7, about 50% of both E and T papers cite studies from the other category (SS:  
19 275  $297/713=42\%$  of E papers cite at least one T paper from the coded network, and  
20 276  $50/94=53\%$  of T papers cite at least one E paper; in the SS & S dataset, the  
21 277 corresponding figures are  $540/998=54\%$  for E citing T and  $90/162=56\%$  for T citing E).  
22 278

23 279 The preceding analysis (Fig. S6 and S7) only counted the number of papers that are part  
24 280 of at least one cross-citation within the coded network. This does not rule out that most  
25 281 cross-citations have the same source papers. To further quantify the distribution of  
26 282 cross-citations, we calculated Gini coefficients to compare the unevenness of the  
27 283 number of across- versus within-category citations received by both E and T papers (see  
28 284 Appendix 1 and 3). This analysis shows that, after correcting for differences in sample  
29 285 size, citations between categories have somewhat higher Gini coefficients than within-  
30 286 category citations (see Appendix 3 for detailed presentation of results). Thus, cross-  
31 287 citations are indeed biased towards highly-cited source papers, but not to the point of  
32 288 involving only a small number of citation classics. Interestingly, the difference between  
33 289 the Gini coefficients for within- and between-category citations seems to be higher for E  
34 290 than T source papers (compare left and right column of Fig. A3). This again suggests  
35 291 that empiricists access a larger proportion of the theoretical literature than vice versa.  
36 292 In summary, the idea that only a small number of source papers from each category are  
37 293 responsible for the lion’s share of communication between theory and empiry is only  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 294 partially supported by our analysis. In particular, a large proportion of theoretical papers  
4 295 receive citations from empirical studies in the network (Fig. S6). Nevertheless, the  
5 296 distribution of citations across categories is more uneven than within categories,  
6 297 indicating that many cross-citations are, indeed, directed towards highly-cited source  
7 298 papers (especially for empirical sources).

11  
12 299 *The “transcendence of influence” hypothesis (vii)*

13  
14 300

15 301 Our final hypothesis states that papers that are highly influential within their field tend to  
16 302 transcend their own category, exerting influence in other categories as well. The  
17 303 hypothesis thus predicts a positive correlation between the citation frequency for a given  
18 304 paper within its own category (e.g. theoretical papers cited by other theoretical papers)  
19 305 and the citation frequency for that same paper outside of its own category (e.g.  
20 306 theoretical papers cited by empirical research). Indeed, in three of the four cases we  
21 307 examined (see Appendix 1), we found evidence of this correlation (the only exception  
22 308 being empirical studies in the Sexual Selection dataset; Fig. 2). In particular, theoretical  
23 309 studies that are highly cited by other theoretical papers also tend to be highly cited by  
24 310 empirical papers. A positive correlation between theoretical and empirical citations exists  
25 311 also for highly-cited first-authors, be they theoreticians or empiricists (Appendix 2). This  
26 312 lends support to the idea that high-impact papers and authors tend to influence not only  
27 313 their own research community, but transcend those boundaries to influence thinking  
28 314 more generally. This finding is good news in light of the shared goal of integration  
29 315 between theoretical and empirical research approaches, because it provides evidence  
30 316 that high-impact research from each approach does have an impact on the other.

31 317

32 318

33 319 **Additional insights from a co-citation network**

34 320 In addition to the directed citation networks, we also studied undirected co-citation  
35 321 networks, which link source papers that are cited by the same target paper (for details,  
36 322 see Appendix 1 and 4). Consequently, co-citations reflect decisions in the community  
37 323 about which citations ‘belong together’ and illustrate patterns of general consensus  
38 324 about connections within the literature.

39 325

1  
2  
3 326 We measured the extent to which papers from the different categories contribute to the  
4 327 structure of the co-citation network, using three common measures of network centrality:  
5 328 degree strength, eigenvector centrality, and betweenness (see Appendix 1). The highest  
6 329 score for all three metrics was obtained by theory papers, followed by reviews. First,  
7 330 theory papers had the highest median degree strength, indicating that they are cited  
8 331 together with many other papers in our data set. Similarly, theory papers had the highest  
9 332 eigenvector centrality scores; indicating not only that they are often cited with other  
10 333 papers, but that they tend to be co-cited with other heavily co-cited papers. Finally, theory  
11 334 and reviews had the highest median betweenness, although the differences in this  
12 335 metric between categories was not as pronounced as in the other two metrics.  
13 336 Nonetheless, the high median betweenness values for theory and reviews indicate that  
14 337 they often bridge between groups of co-cited papers that are otherwise more isolated  
15 338 (see Fig S9 for full results). Specifically, the highest betweenness value was for a review  
16 339 paper (Emlen and Oring 1977) in the SS data set and a theoretical paper (Lande 1981)  
17 340 in the SS&S data set. Indeed, the high betweenness score for these papers is apparent  
18 341 when visualizing the networks (see Fig. S10 for full visualization of co-citation networks).  
19 342 Thus, our analysis of the co-citation networks provides some support for the idea that  
20 343 some paper types more often play critical roles in bridging gaps between otherwise less-  
21 344 connected factions within the literature. Specifically, in the SS dataset, theory and mixed  
22 345 papers were more likely to fill this role than other paper types. In the SS & S dataset,  
23 346 however, this role was filled mainly by review and theory papers. This bridging role was  
24 347 not filled by empirical papers in the SS dataset, where none of the empirical papers had  
25 348 high betweenness scores. In contrast, a few empirical papers in the SS & S data set had  
26 349 higher betweenness scores, which suggests that in that subset of literature, particular  
27 350 empirical papers may contribute to connecting the body of literature together, although to  
28 351 a lesser extent than review and theory papers.

352

353 **More integration exists between theory and empiry than anticipated. It can still be**  
354 **improved.**

355 In conclusion, using a citation network analysis to quantify citations across  
356 methodological approaches, we find reason to challenge the intuition that theory and  
357 empiry are poorly integrated in evolutionary biology (Haller 2014).



1  
2  
3 358 Indeed, the main message of our study is two-fold. 1) Despite our finding that empirical  
4 359 and theoretical research more often cite within category than not, theoretical and  
5 360 empirical biology might inform each other more than the community thinks. 2)  
6  
7  
8 361 Nonetheless, opportunities exist to enhance communication and integration between  
9 362 theoretical and empirical research.

10  
11 363  
12  
13 364 We found that neither theory nor empiry self-cited more than 60% of the time, meaning  
14 365 that—assuming citations are a reasonable proxy for integration—at least 40% of  
15 366 citations represented a form of integration (either direct or indirect via reviews). In  
16 367 addition, we found that both highly-cited papers and highly-cited authors from one  
17 368 domain tend to be well received by the other domain, too. Thus, the observed citation  
18 369 patterns indicate that, at least at some level, practitioners of each approach are following  
19 370 and valuing advances in the other approach.

20  
21  
22 371  
23  
24 372 Why then do many researchers report a disconnect between theory and empiry? There  
25 373 are at least two potential explanations. First, the perceived disconnect between  
26 374 theoretical and empirical approaches may reflect unrealistic expectations. Among the co-  
27 375 authors of this paper, we have had extensive and productive conversations about the  
28 376 goals of theory and empiry, centering around how closely theory and empiry match when  
29 377 they do inform each other. Many empiricists envision theoretical models that match  
30 378 empirical systems at the level of specific parameter values, while theorists often work at  
31 379 a higher level of abstraction, seeking to test the validity of concepts rather than arriving  
32 380 at specific numerical predictions (Servedio et al. 2014). Second, citation patterns may  
33 381 overestimate integration of theory and empiry, because cross-citations are mostly used  
34 382 to provide background and context to a paper, rather than being a direct inspiration for a  
35 383 framework of research. While measuring the importance of this effect is beyond the  
36 384 scope of this study, in the following we detail three general modes by which theory and  
37 385 empiry are, indeed, integrated/connected and, in each case, we advocate for explicit  
38 386 attempts by researchers to engage with each other more often.

39  
40  
41 387  
42  
43 388 First, integration between theory and empiry can happen sequentially, where one  
44 389 approach inspires the other. Formal theory is often motivated by long-standing empirical  
45 390 observations that may be especially difficult to explain. For example, the observation that  
46 391 females of many species prefer males with elaborate and exaggerated traits motivated

1  
2  
3 392 the now sizeable theoretical literature about sexual selection and, even more  
4 393 specifically, the somewhat technical studies investigating the relative strengths of direct  
5 394 and indirect selection (e.g. Weatherhead and Robertson 1979; Kirkpatrick 1985;  
6 395 Kirkpatrick and Barton 1997). In turn, new empirical areas of study are sometimes  
7 396 preceded by and born out of theoretical formulations. For example, the mathematical  
8 397 models that examine how the relative strengths of direct and indirect selection interact  
9 398 and how they can lead to reinforcement (Servedio 2001) directly motivated a subsequent  
10 399 empirical study (Albert and Schluter 2010).

11  
12  
13  
14  
15  
16 400

17 401 Second, as outlined above, review papers provide powerful opportunities for researchers  
18 402 to contribute to integration, and our results show both that their citations are more evenly  
19 403 distributed across domains than those of papers reporting original studies and that they  
20 404 often function as bridges between disparate groups of cited literature. Indeed, the job of  
21 405 condensing a large body of literature may become ever more important as publication  
22 406 rate increases and as computational and molecular technologies grow, making  
23 407 unfamiliar domains of science even more difficult to master. Although reviews are  
24 408 sometimes regarded with skepticism and viewed as a mechanism by which researchers  
25 409 can increase their citations (because reviews are cited more frequently than original  
26 410 studies), we do not endorse this cynical view. Instead, we encourage theoreticians and  
27 411 empiricists alike to continue the practice of, first, identifying places in the literature with  
28 412 an abundance of current research articles and, next, synthesizing the results for a  
29 413 broader audience. Further, we advocate explicit collaborations between empiricists and  
30 414 theoreticians in writing reviews that summarize contemporary empirical and theoretical  
31 415 research on a given topic in one paper (e.g. Kopp *et al.* 2017).

32  
33  
34  
35  
36  
37  
38  
39  
40  
41 416

42 417 Finally, explicit integration of theory and empiry occasionally happens *within one*  
43 418 *research paper* (i.e. our mixed (M) category). These papers were rare in our data sets,  
44 419 probably reflecting the fact that models of both sexual selection and speciation are  
45 420 situated at a rather abstract level and thus it is often challenging to draw direct analogies  
46 421 with empirical data from a specific study system. Furthermore, we know from our own  
47 422 experience as co-authors that collaborations across specialization can be difficult.  
48 423 Nonetheless, we contend that the relative dearth of these mixed papers represents an  
49 424 underutilized opportunity for increased collaboration and integration. It is interesting in



1  
2  
3 425 this context that citation patterns of mixed papers are closest to the “null-expectation”  
4 426 given by the distribution of studies in the literature.

5  
6 427

7  
8 428 We hope that our analysis will be of use to researchers that feel there is a lack of  
9 429 integration between theory and empiry: if you have the impression that the question of  
10 430 interest to you is not being addressed by authors who utilize the other approach, writing  
11 431 a review or a mixed paper might be an effective way to draw attention to it.

12  
13  
14 432

15  
16  
17 433 **Acknowledgements**

18  
19 434 This work was supported by the National Evolutionary Synthesis Center (NESCent),  
20 435 NSF #EF-0905606. We thank NESCent staff for their support and Angela Zoss of Duke  
21  
22 436 University Libraries for assistance with the Sci<sup>2</sup> tool. We thank resident NESCent  
23  
24 437 colleagues and additional members of the working group “Toward a unified evolutionary  
25 438 theory of decision making in animals” for helpful discussion, and the Santa Fe Institute  
26  
27 439 for funding a collaboration visit (CLF to work with EAH).

28  
29 440

30 441 **Cited Literature**

31  
32 442

33  
34 443 Albert, Author A Y K, and D Schluter. 2010. “Reproductive Character Displacement of  
35 444 Male Stickleback Mate Preference: Reinforcement or Direct Selection? *Evolution* 58  
36 445 (5): 1099–1107.

37  
38 446 Bateman, AJ. 1948. “Intra-Sexual Selection in *Drosophila*.” *Heredity* 2: 349–68.

39 447 Bennet, J. 2000. “No Title.” In *Complete Variorum Edition of Fisher’s The Genetical*  
40 448 *Theory of Natural Selection*, edited by J.H. Bennett, 308–9. Oxford University  
41  
42 449 Press.

43  
44 450 Colwell, Robert K. 2009. “Biodiversity: Concepts, Patterns, and Measurement.”

45 451 Deltas, George. 2003. “The Small-Sample Bias of the Gini Coefficient: Results and  
46  
47 452 Implications for Empirical Research.” *The Review of Economics and Statistics* 85  
48  
49 453 (1): 226–34. doi:10.1162/rest.2003.85.1.226.

50  
51 454 Fisher, R.A. 1930. *The Genetical Theory of Natural Selection*. Edited by JH Bennett.  
52  
53 455 Oxford: Oxford University Press.

54 456 Gini, Corrado. 1936. “On the Measure of Concentration with Special Reference to  
55  
56 457 Income and Statistics.” *Colorado College Publication, General Series* 208: 73–79.

- 1  
2  
3 458 Haller, BC. 2014. "Theoretical and Empirical Perspectives in Ecology and Evolution: A  
4 459 Survey." *BioScience* 64 (10): 907–16.  
5  
6 460 Kirkpatrick, M. 1985. "Evolution of Female Choice and Male Parental Investment in  
7 461 Polygynous Species: The Demise of the 'sexy Son.'" *American Naturalist* 125 (6):  
8 462 788–810.  
9  
10 463 Kirkpatrick, Mark, and N H Barton. 1997. "The Strength of Indirect Selection on Female  
11 464 Mating Preferences." *Proceedings of the National Academy of Sciences of the*  
12 465 *United States of America* 94 (4): 1282–86.  
13  
14 466 Servedio, Maria R. 2001. "Beyond Reinforcement: The Evolution of Premating Isolation  
15 467 by Direct Selection on Preferences and Postmating, Prezygotic Incompatibilities."  
16 468 *Evolution* 55 (10): 1909–20. doi:10.1111/j.0014-3820.2001.tb01309.x.  
17  
18 469 Servedio, Maria R., Yaniv Brandvain, Sumit Dhole, Courtney L. Fitzpatrick, Emma E.  
19 470 Goldberg, Caitlin A. Stern, Jeremy Van Cleve, and D. Justin Yeh. 2014. "Not Just a  
20 471 Theory—the Utility of Mathematical Models in Evolutionary Biology." *PLoS Biology*  
21 472 12 (12): e1002017. doi:10.1371/journal.pbio.1002017.  
22  
23 473 R Core Team. 2017. "R: A Language and Environment for Statistical Computing."  
24 474 Vienna, Austria. <https://www.r-project.org/>.  
25  
26 475 Sci2 Team. 2009. "Science of Science (Sci2) Tool." Indiana University and SciTech  
27 476 Strategies. <https://sci2.cns.iu.edu/user/index.php>.  
28  
29 477 Weatherhead, Patrick J, and Raleigh J Robertson. 1979. "Offspring Quality and the  
30 478 Polygyny threshold: 'The Sexy Son Hypothesis.'" *The American Naturalist* 113 (2):  
31 479 201–8.  
32  
33 480 Wittebolle, Lieven, Massimo Marzorati, Lieven Clement, Annalisa Balloi, Daniele  
34 481 Daffonchio, Kim Heylen, Paul De Vos, Willy Verstraete, and Nico Boon. 2009.  
35 482 "Initial Community Evenness Favours Functionality under Selective Stress." *Nature*  
36 483 458 (7238): 623–26. doi:10.1038/nature07840.  
37  
38 484  
39 485  
40 486

487 **Table 1.** Possible patterns of citation behavior. All scenarios hypothesize citation  
 488 patterns within and between theory and empiry except for scenario v, which posits a role  
 489 for review articles.

490

i	Researchers cite primarily within their own category (Haller 2014).
ii	Researchers cite each category (theory vs empiry) in proportion to its frequency in the literature (“null” hypothesis)
iii	Citation frequency is driven by the size of research community (“size of community” hypothesis).
iv	Theory functions analogously to a specific taxonomic study system in its citation patterns (“theory-as-system” hypothesis).
v	Reviews are cited in lieu of primary literature across categories (“reviews-as-proxies” hypothesis).
vi	Citations between categories are concentrated among a small number of publications (“concentrated cross-citation” hypothesis).
vii	Influential papers in one category are likely to transcend category and be highly influential in the other (“transcendence of influence” hypothesis).

491

492

493

1  
2  
3 494

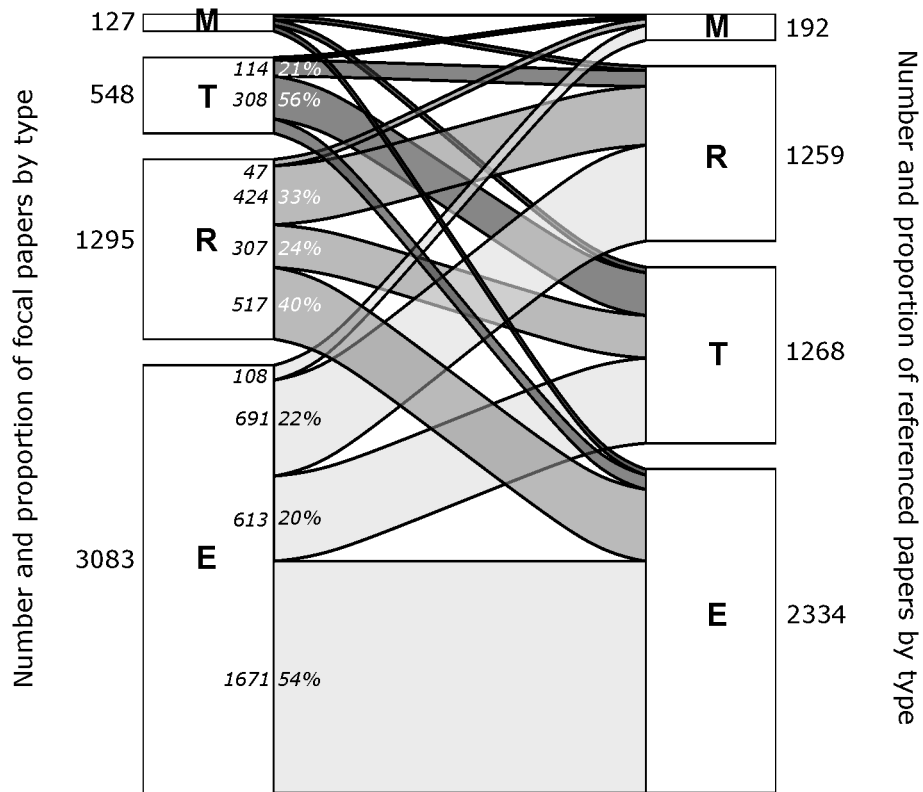
4 495 Figure 1. Citation patterns for different types of papers (E = empirical, R = review, T =  
5 496 theory, M = mixed), for the a) Sexual Selection (SS) and b) Sexual Selection and  
6 497 Speciation (SS & S) data sets. The figure shows the answers to the question: Which  
7 498 types of papers are papers of each category citing? Vertical white boxes are scaled by  
8 499 the proportion of papers by type for the papers that cite (targets) and the papers that are  
9 500 cited (source) in the coded network. The size of the ties that flow between these  
10 501 categories are proportional to the total number of citations made by a given type of  
11 502 paper (i.e. the four ties that flow from each white box sum to 100% of the citations made  
12 503 by that category type, and all 16 ties in the whole plot show 100% of total citations made  
13 504 in our coded network). Numbers in margins indicate the total number of citations made  
14 505 by category, italicized numbers within target paper boxes indicate the number of  
15 506 citations made by papers of each type cited in each type of paper, and major links are  
16 507 labeled with the percent of total citations for each paper type. For example, in the SS  
17 508 data (a), there were 3083 citations made by empirical papers, and of these citations,  
18 509 1671, or 54%, referenced other empirical papers. Note that not all links are labeled, see  
19 510 Table S1 for all values.

20  
21  
22 511

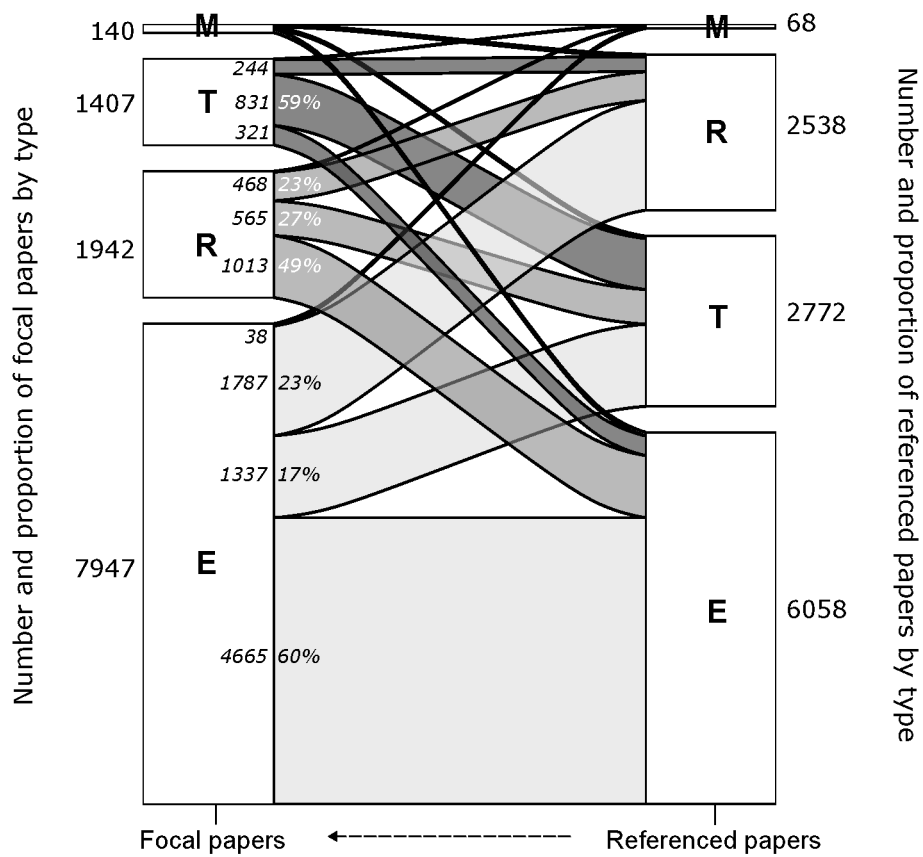
23 512 Figure 2. Relationship between same-category (e.g. theory citing theory) and other-  
24 513 category (e.g. empiry citing theory) influence for empirical (E) and theoretical (T) papers.  
25 514 Each data point represents one paper (the “focal” paper). The position of each data point  
26 515 describes i) the proportion of all same-category papers in the coded network that cite the  
27 516 focal paper (x-axis) and ii) the proportion of all other-category papers in the coded  
28 517 network that cite the focal paper (y-axis) for a) empirical papers (Spearman’s rank  
29 518 correlation;  $\rho = 0.10$ ) and b) theoretical papers (Spearman’s rank correlation;  $\rho = 0.49$ ) in  
30 519 the Sexual Selection (SS) data set and for c) empirical papers (Spearman’s rank  
31 520 correlation;  $\rho = 0.49$ ) and d) theoretical papers (Spearman’s rank correlation;  $\rho = 0.57$ ) in  
32 521 the Sexual Selection and Speciation (SS & S) data set.

33  
34  
35 52236  
37  
38 523  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

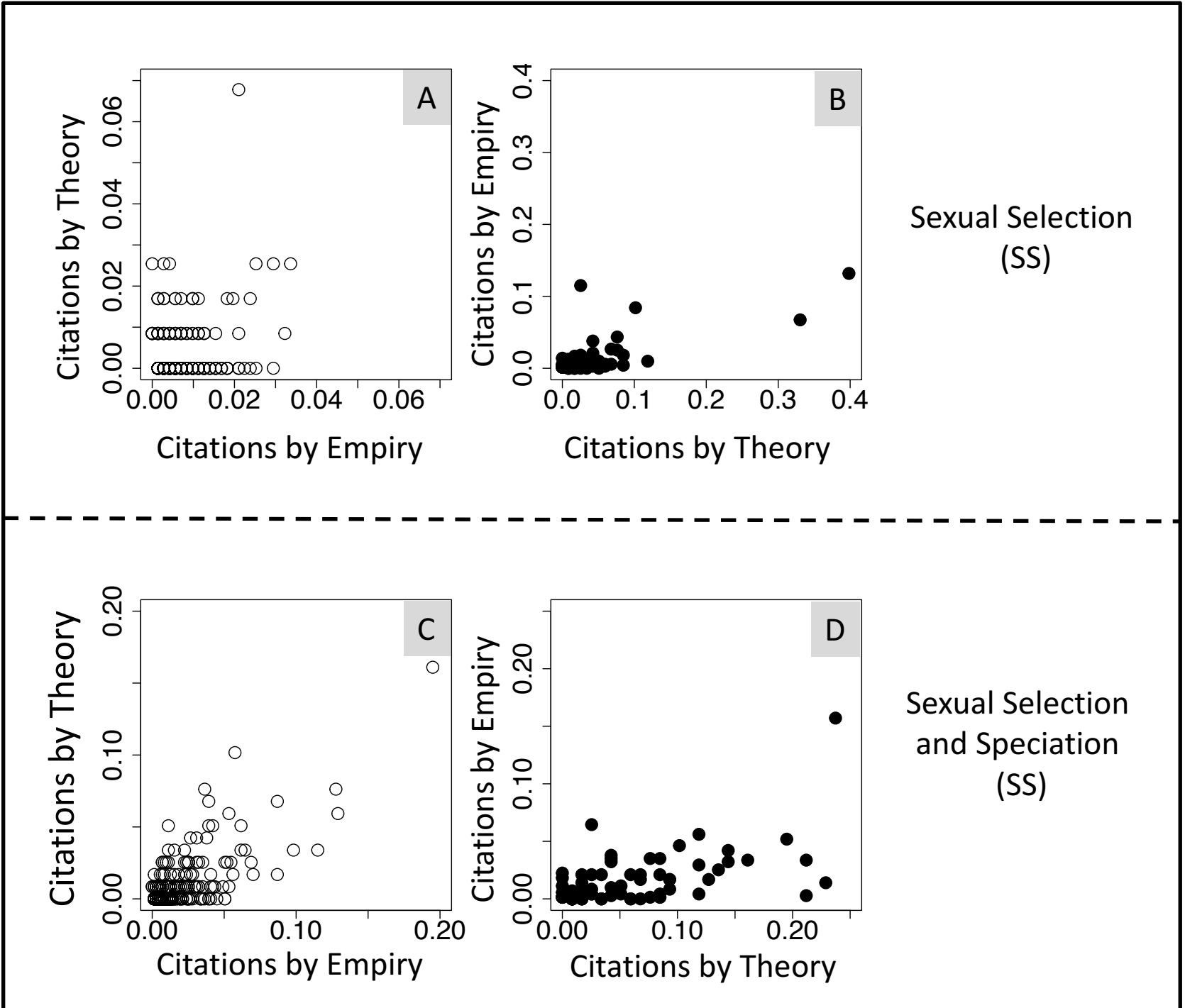
(a) Sexual selection (SS)



(b) Sexual selection & speciation (SS&S)

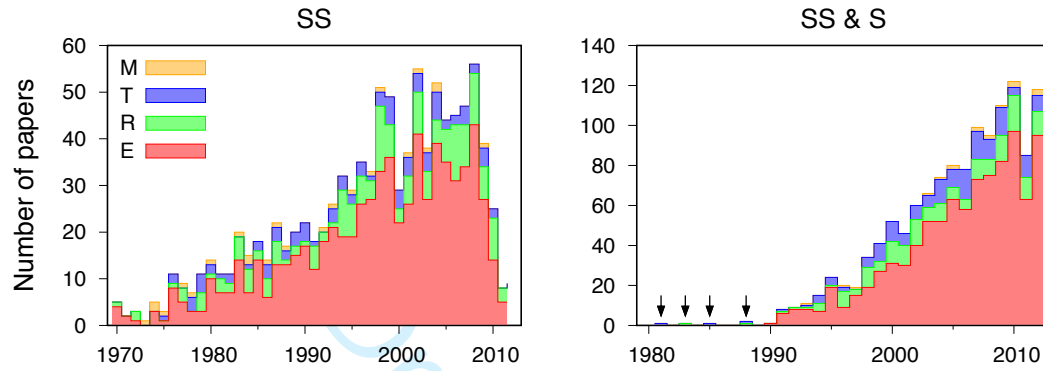


Focal papers ←----- Referenced papers



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47

## Supplementary Figures:

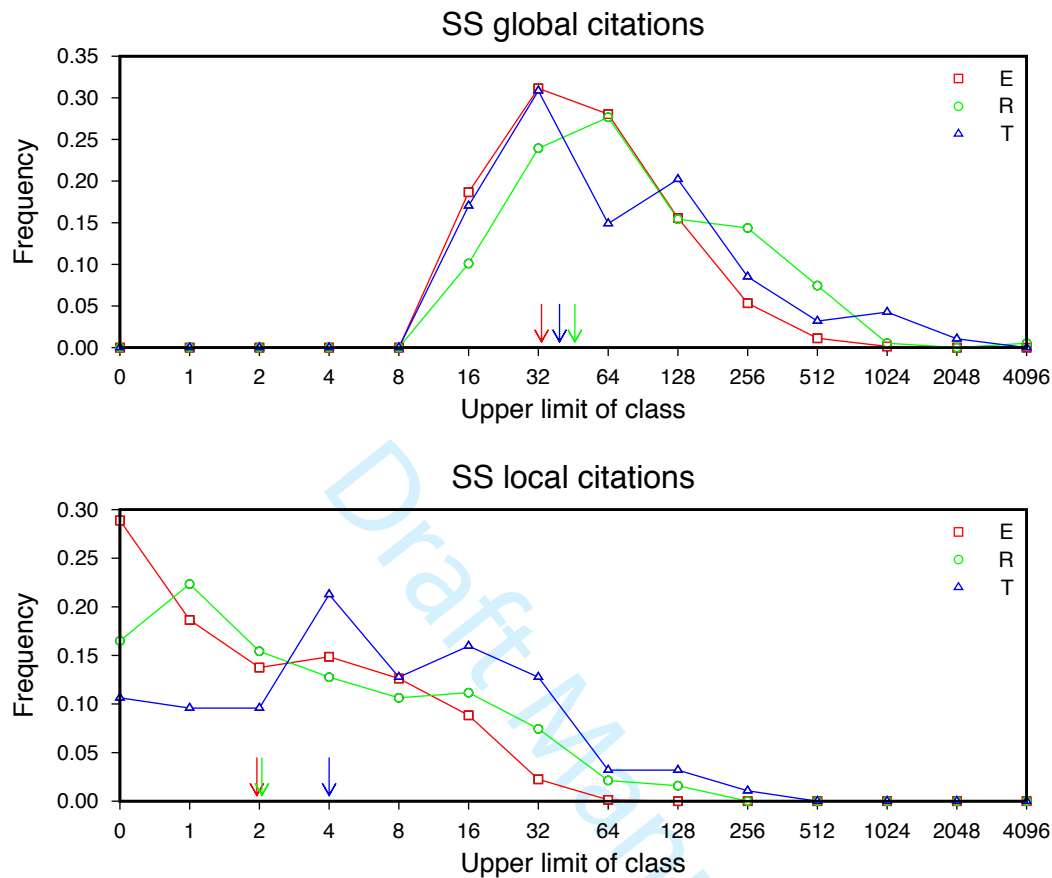


**Figure S1:** Numbers of coded papers for the two datasets over time. Arrows in the right-hand plot highlight early papers for the sexual selection and speciation (SS&S) dataset (Lande 1981, West-Eberhard 1983, Wu 1985, Lande and Kirkpatrick 1988, Verrell 1988).

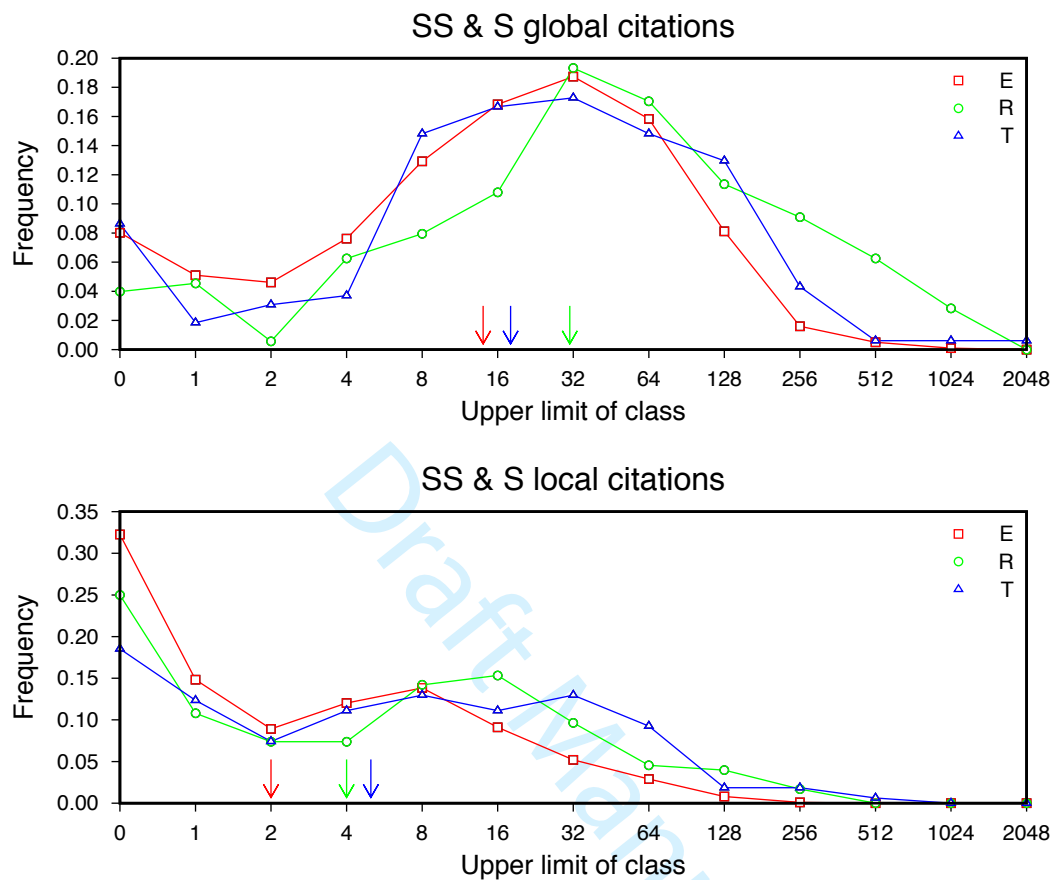
**Table S1.** Summary of citation dyad links by paper type. From left to right, columns show the number of links (citations) in the coded network, proportion of total citations that fall into each category, the proportion of papers in each category that are cited at least once, and the difference between observed citation frequencies and the citation frequencies that would be 'expected' if citations reflected prevalence in the original data set. For example, in the SS dataset, empirical papers cite another empirical paper 1671 times which is 54% of all citations by empirical papers; empirical papers comprise 69% of papers in the dataset, and the observed E cites E link percent is 15% lower than what would be 'expected' if citation frequencies reflected the frequency of paper types in the data set.

Citation dyad	Sexual Selection (SS)				Sexual Selection & Speciation (SS & S)			
	# links	% links	% papers cited in dataset	% difference	# links	% links	% papers cited in dataset	% difference
E cites E	1671	54%	69%	-15%	4665	60%	74%	-14%
E cites R	691	22%	19%	+3%	1787	23%	13%	+10%
E cites T	613	20%	10%	+10%	1337	17%	12%	+5%
E cites M	108	4%	2%	+2%	38	1%	1%	-1%
Total E cites	3083				7947			
R cites E	517	40%	69%	-29%	1013	49%	74%	-25%
R cites R	424	33%	19%	+14%	468	23%	13%	+10%
R cites T	307	24%	10%	+14%	565	27%	12%	+15%
R cites M	47	4%	2%	+2%	16	1%	1%	0%
Total R cites	1295				1942			
T cites E	104	19%	69%	-50%	321	23%	74%	-51%
T cites R	114	21%	19%	+2%	244	17%	13%	+4%
T cites T	308	56%	10%	+46%	831	59%	12%	+47%
T cites M	22	4%	2%	+2%	11	1%	1%	0%
Total T cites	548				1407			
M cites E	42	33%	69%	-36%	59	42%	74%	-32%
M cites R	30	24%	19%	+5%	39	28%	13%	+15%
M cites T	40	32%	10%	+22%	39	28%	12%	+16%
M cites M	15	12%	2%	+10%	3	2%	1%	+1%
Total M cites	127				140			

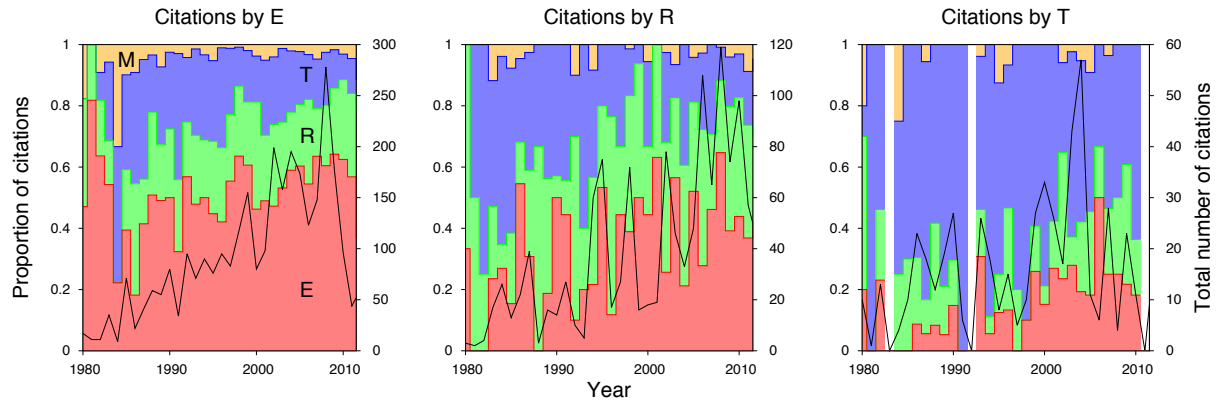




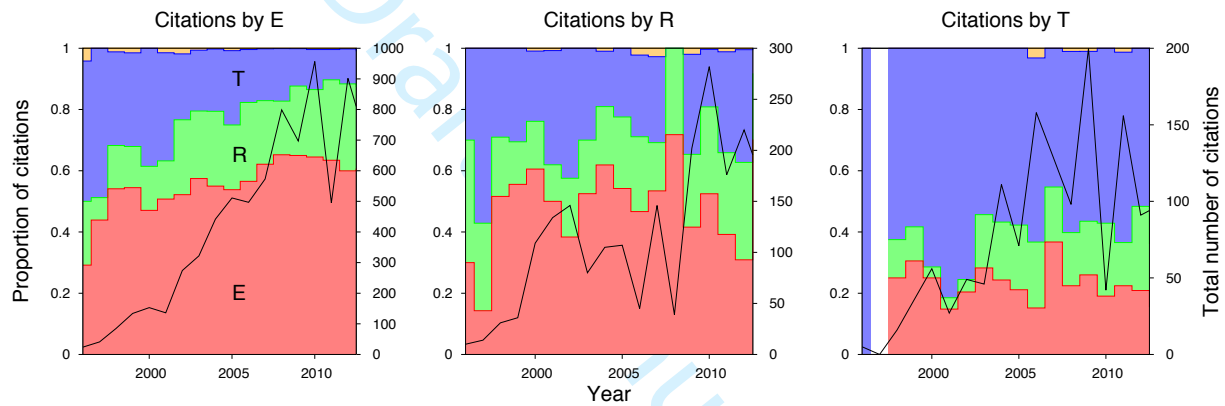
**Figure S2:** Distribution of global (overall) and local (within-network) citation numbers for empirical (E) and theoretical (T) papers as well as reviews (R) from the sexual-selection (SS) dataset. “Mixed” papers (M) are excluded due to small sample size. Curves represent frequencies of logarithmically-sized classes. Arrows indicate medians (global: 33 (E), 46 (R), 39.5 (T); local: 2 (E), 2 (R), 4 (T); red and green arrows are slightly displaced for visibility). Classes for less than eight global citations are empty because only papers with at least 10 global citations were included in the dataset.



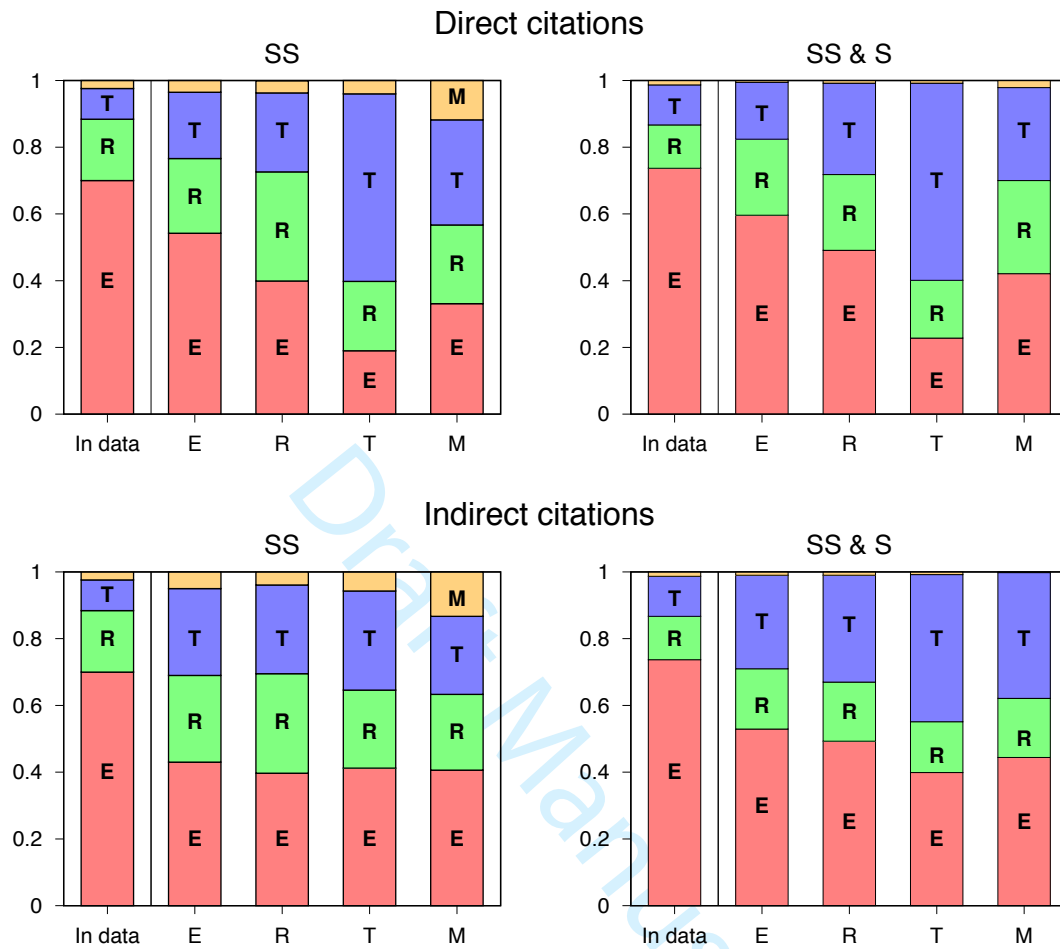
**Figure S3.** Distribution of global (overall) and local (within-network) citation numbers for empirical (E) and theoretical (T) papers as well as reviews (R) from the sexual-selection and speciation (SS & S) dataset. Mixed papers (M) are excluded due to small sample size. Curves represent frequencies of logarithmically-sized classes. Arrows indicate medians (global: 14 (E), 31 (R), 18 (T); local: 2 (E), 4 (R), 5 (T)).



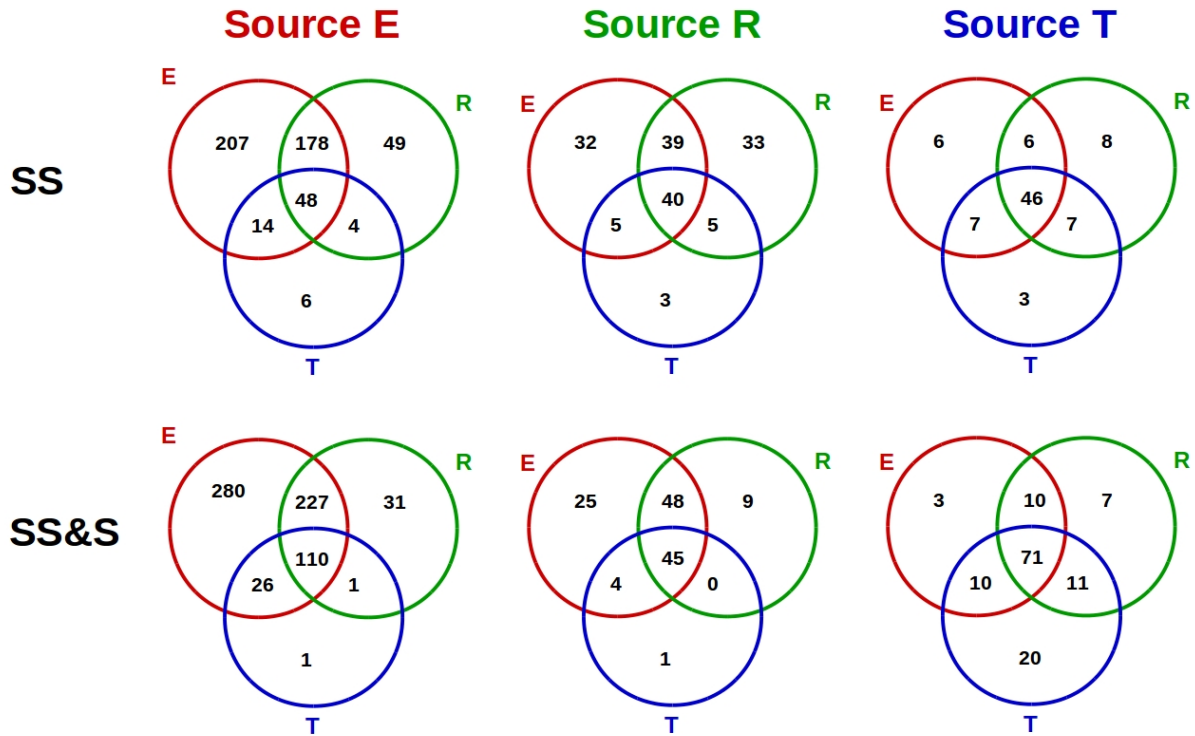
**Figure S4a:** Dyadic citation frequencies over time for sexual selection (SS) data set.



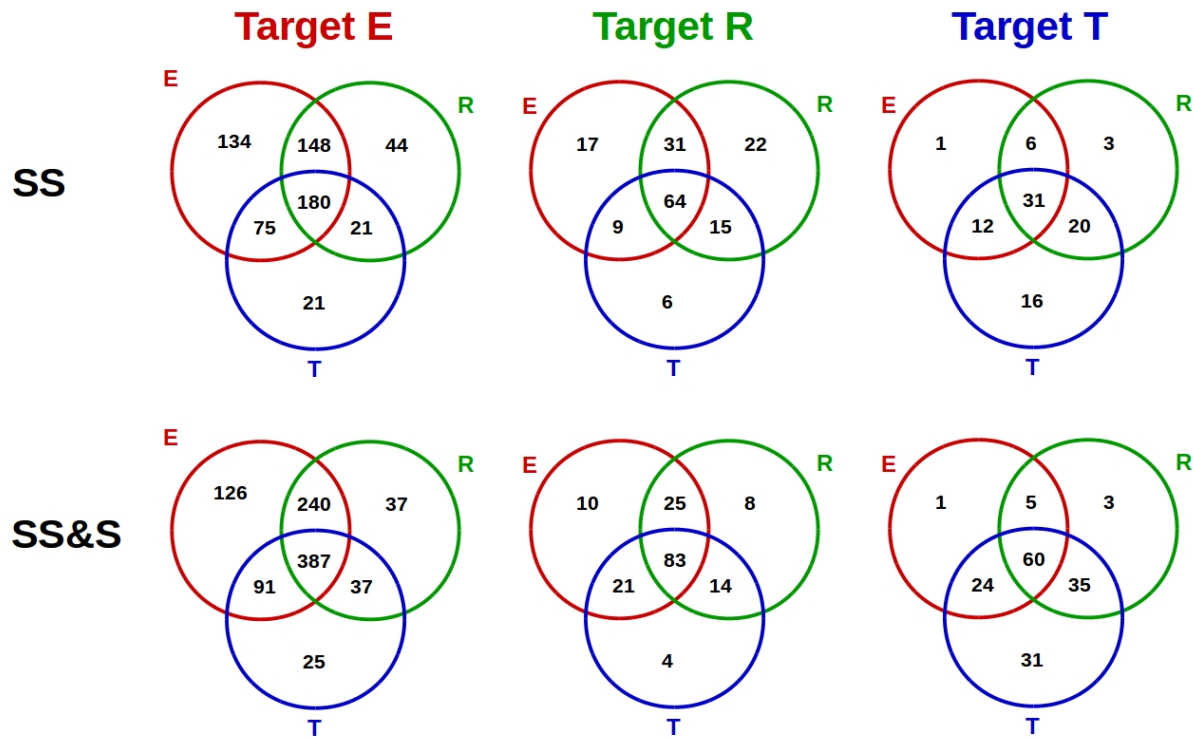
**Figure S4b:** Dyadic citation frequencies over time for sexual selection and speciation (SS & S) data set.



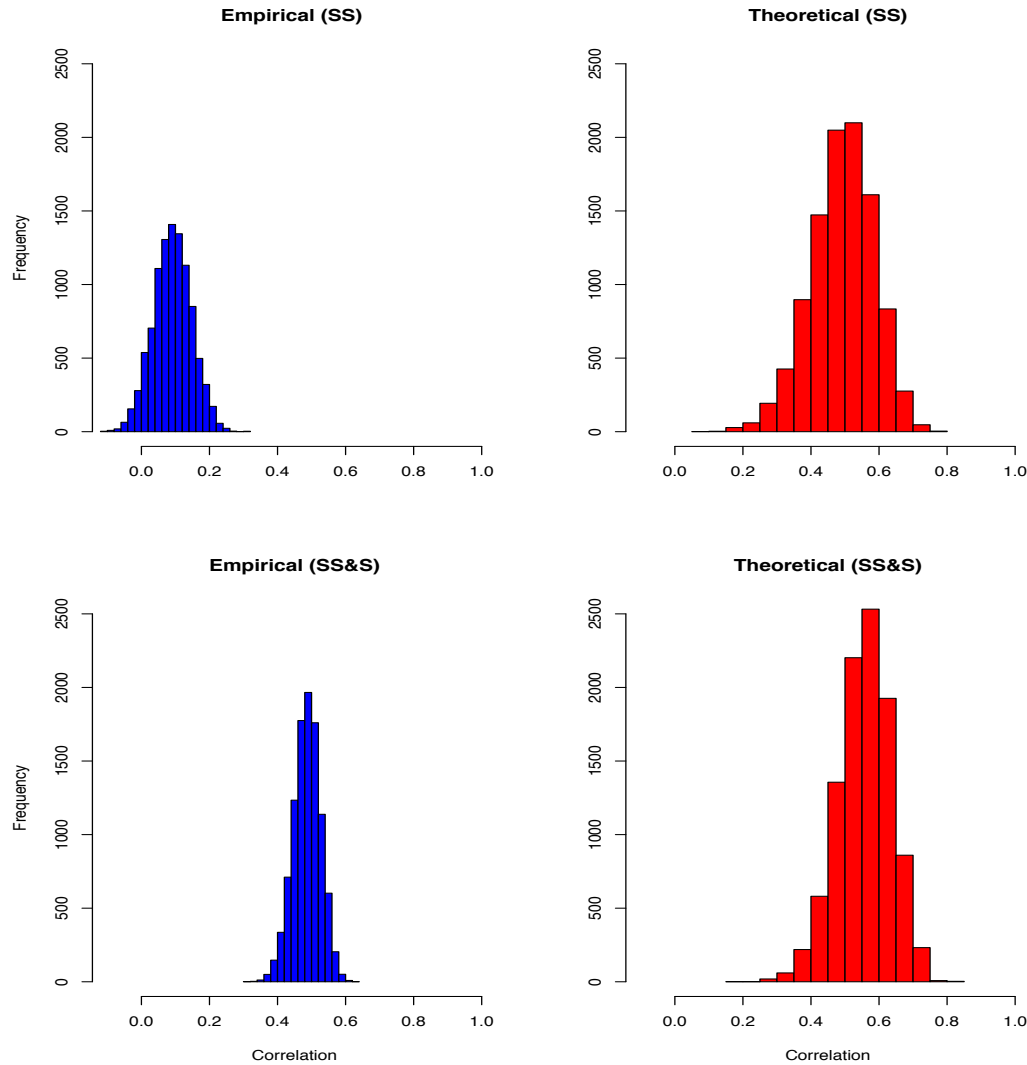
**Figure S5.** Patterns of indirect citations for different types of target papers (E = empirical, R = review, T = theory, M = mixed), for the sexual selection (SS) and sexual selection and speciation (SS & S) data sets. The first column in each plot shows the distribution of papers in the coded network. The following columns show the distribution of the types of (in-network) papers cited indirectly via reviews by papers in the different categories; that is, the column label represents the type of the citing paper (target), and the differently colored parts of each column represent the different types of indirectly cited papers (sources cited by reviews cited by the target). The plot shows the answers to the question: When papers from the different categories cite reviews, what kind of papers are cited by these reviews? Note that sample size for the M category (right-most column) is very small.



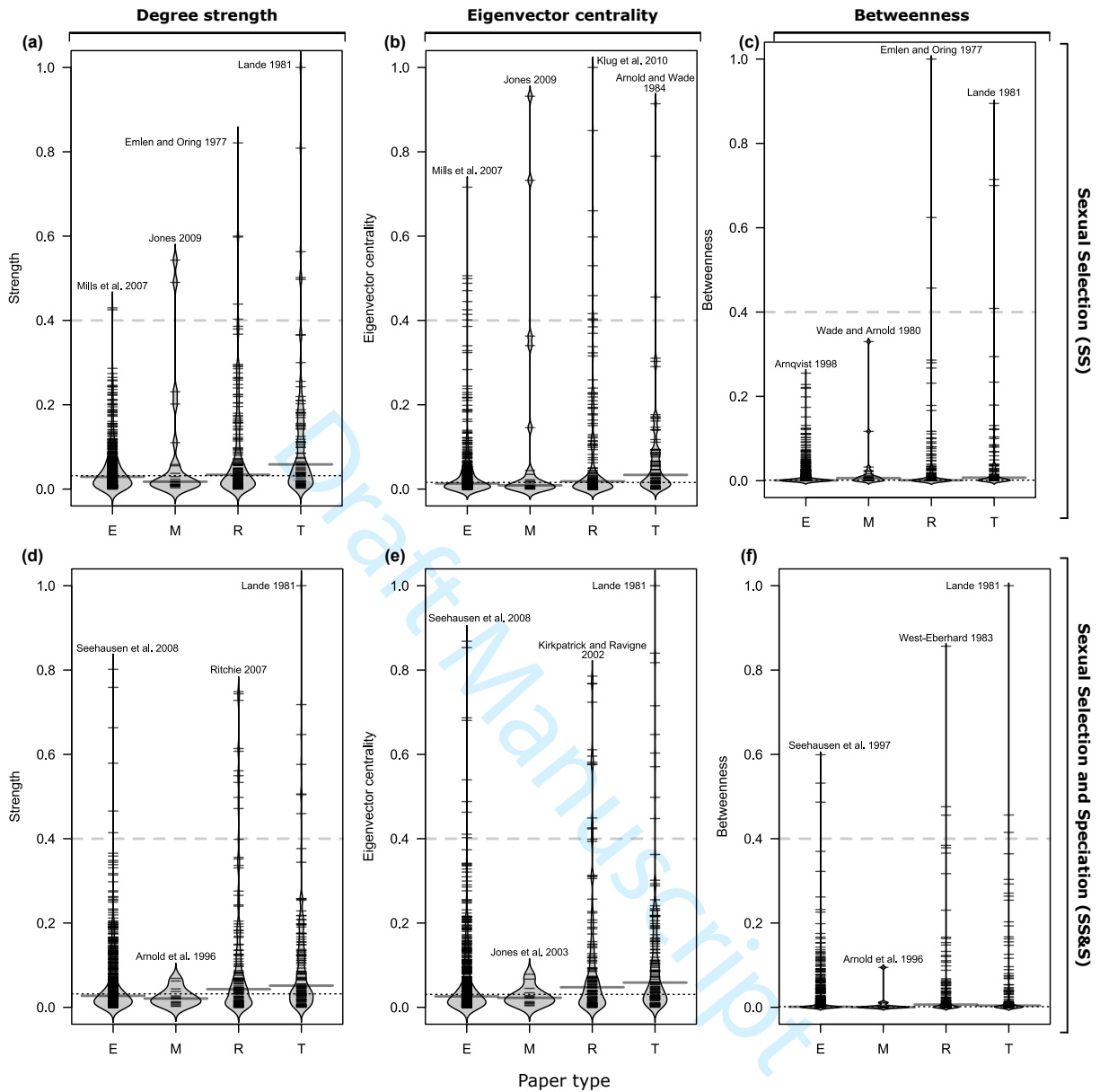
**Figure S6:** Venn-diagrams showing the numbers of source papers of a given category (columns) that have been cited by at least one target paper from the E, R, or T category (cycles), or combinations thereof. For example, in the top-left panel, 207 empirical papers from the SS dataset have been cited only by empirical papers (within our coded network), whereas 14 have been cited by both empirical and theoretical papers, and 48 have been cited by at least one paper from each category. The mixed category M was excluded from this analysis. Note that the total number of source papers in each plot (the sum of all numbers) is usually less than then total number of papers in the respective category (table 1), because some papers do not receive any citations from either E, R or T within the coded network.



**Figure S7:** Venn-diagrams showing the numbers of target papers of a given category (columns) citing at least one source paper from the E, R, or T category (cycles), or combinations thereof. For example, in the top-left panel, 134 empirical papers from the SS dataset have cited only empirical papers (within our coded network), whereas 75 have cited both empirical and theoretical papers, and 180 have cited at least one paper from each category. The mixed category M was excluded from this analysis. Note that the total number of target papers in each plot (the sum of all numbers) is usually less than then total number of papers in the respective category (table 1), because some papers do not cite any paper from either E, R or T within the coded network.

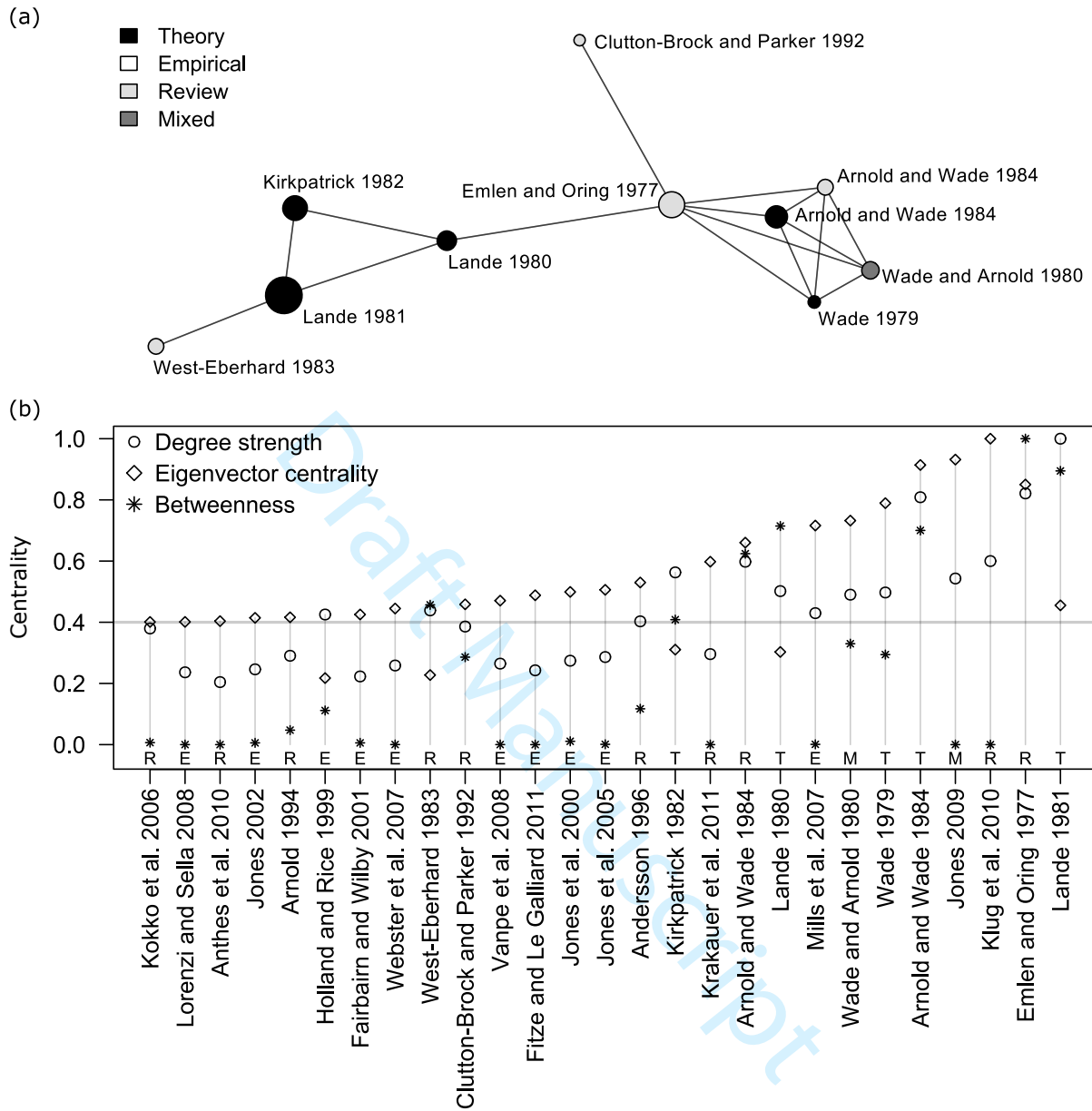


**Figure S8.** Histograms of bootstrapped Spearman rank correlation coefficients between proportion of citations received within a category and proportion of citations received across a category, as represented by Figure 2 in the main text. With the exception of empirical papers from the sexual selection (SS) dataset, all values are clearly positive.

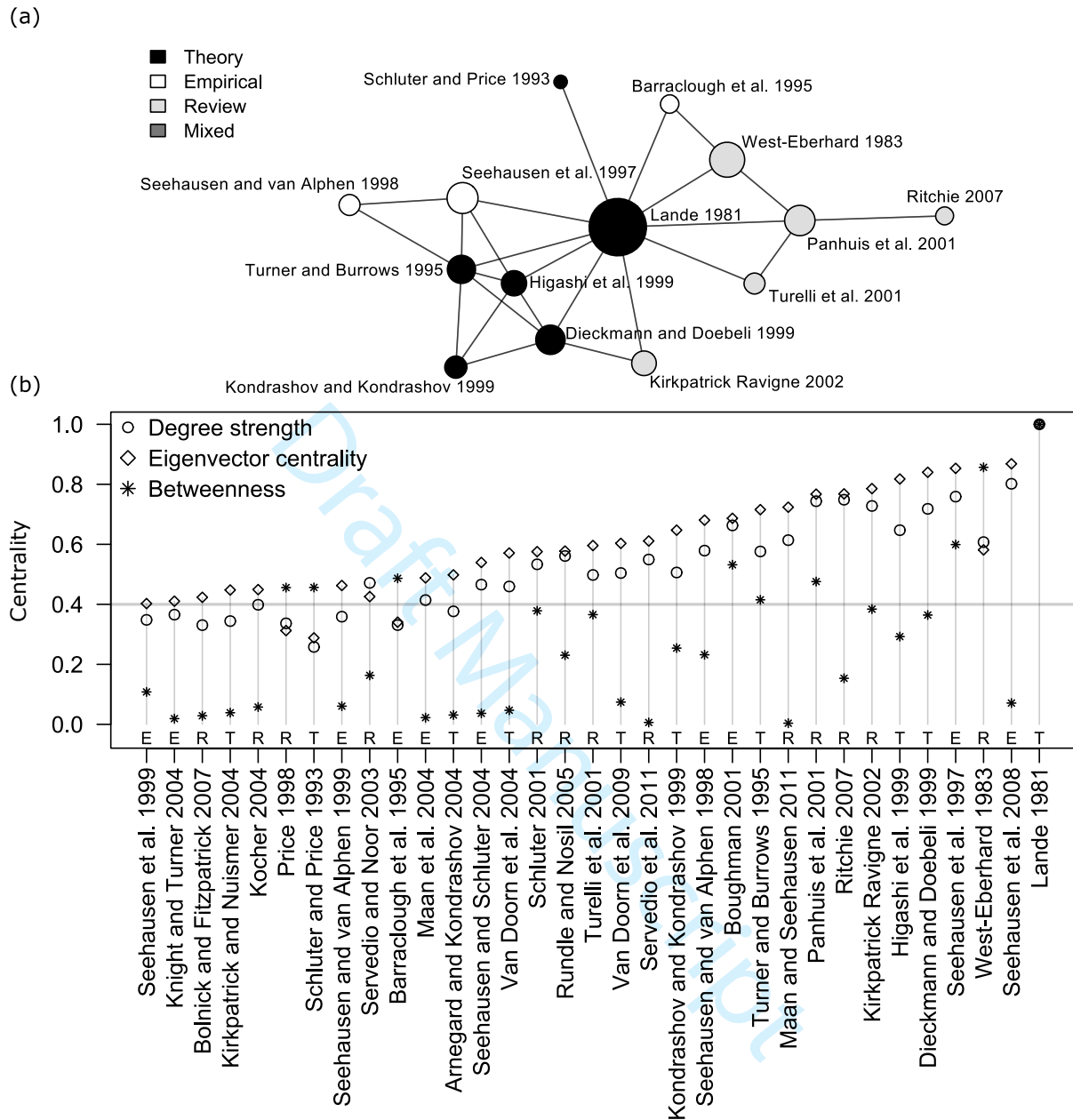


**Figure S9.** Distribution of centrality scores for SS and SS & S papers in the co-citation network across all three centrality measures (strength, eigenvector centrality, and betweenness). These beanplots are a combination of a one-dimensional scatter plot (short horizontal line segments) and the distribution of those plots as a density shape (grey polygons); also shown are the group median (grey horizontal segment for each bean), and the global median across all datapoints (black horizontal line). The paper with maximum centrality per paper type is labeled. The grey dashed line indicates the “high centrality” threshold used to subset centrality metrics for Figures S10 and S11. Centrality values for strength and betweenness rescaled to range between 0 and 1 to facilitate comparison.





**Figure S10.** Highly central Sexual Selection (SS) papers in the co-citation network. For all papers with at least one centrality measure above 0.4 (see threshold, Figure S9), we plotted the values of all three centrality metrics (b). The co-citation network in panel (a) shows a subset of these papers, where papers needed to be co-cited within the SS dataset more than 15 times (networks pruned in order to more effectively visualize patterns). Node size is proportional to the total number of times each paper was locally cited in the database; node color indicates paper type.



**Figure S11.** Highly central Sexual Selection & Speciation (SS & S) papers in the co-citation network. For all papers with at least one centrality measure above 0.4 (see threshold, Figure S9), we plotted the values of all three centrality metrics (b). The co-citation network in panel (a) shows a subset of these papers, where papers needed to be co-cited within the SS & S dataset more than 40 times (networks pruned in order to more effectively visualize patterns). Node size is proportional to the total number of times each paper was locally cited in the database; node color indicates paper type.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Draft Manuscript

## Appendix 1

### *Data collection*

To create our data sets, we performed two Web of Science searches. For the Sexual Selection dataset (SS), an initial search for the topic area “sexual selection” yielded ~22,000 results, a sample size that was prohibitively large (because we subsequently had to score each paper in the data set; see below). We therefore restricted our search by including only those papers that contained “sexual selection” in the title, were published between 1970 and 2013, and had received at least 10 citations at the time of the search (April 2013). We created the Sexual Selection and Speciation data set (SS & S) by performing a search with “sexual selection” and “speciation” in the topic fields and restricting the search to papers that were published between 1980 (because this is when the study of sexual selection and speciation really took off) and April 2013. Next, we downloaded the ISI files, including abstracts, which yielded initial data sets for both sexual selection (n = 1069 papers) and sexual selection and speciation (n = 1405 papers). Our final SS dataset retained 1019 papers and the SS & S dataset 1354. For both data sets, empirical research was the most prevalent category (see Table A1 for complete composition of these datasets). The fields represented by each data set have grown over time. The SS dataset started in the 1970s and increased steadily throughout the sample period. Likewise, the SS & S dataset started with a small number of seminal papers in the 1980s and increased sharply and steadily from the 1990s onward (Fig. S1).

### *Coding*

We assigned each paper in both data sets to one of four categories. The empirical category (E) includes those papers that perform original analyses on new data, as well as formal meta-analysis on published empirical data. The theoretical category (T) contains papers that present a new mathematical model and, although simulated data may be used, do not analyze new or previously published empirical data. A third category represents reviews and ideas (R; called reviews from here forward for brevity); more specifically, this category includes traditional review articles as well as conceptual, verbal, and graphical models, perspectives, and opinion pieces, but excludes letters, commentaries, and replies. A final “mixed” category (M) contains papers that present a new mathematical model and new data, or that use previously published data to demonstrate, illustrate, or test a new model. These types of papers might represent ideal examples of integration between theory and empiry, but we excluded them from some of our analyses for two reasons. First, our central question was about interactions between theoretical and empirical research as measured by citation patterns; these blended papers represent integration within one study and therefore do not fit squarely within one domain. Second, they represent such a small proportion of the

1  
2  
3 literature that many of the analyses would have been uninformative (see Table A1 ; raw data to be  
4 uploaded to Dryad upon acceptance).  
5  
6

7  
8 Because almost all authors contributed to the coding effort, we assessed inter-observer reliability. To do  
9 so, we randomly generated a subset of papers for each of the larger data sets. Each of the papers in the  
10 subsets were then coded independently by two different researchers via random assignment across 9  
11 participants. Although the original subsets contained 240 papers each, we subsequently reduced them by  
12 removing papers from the M category, leaving 224 and 231 papers for the SS and SS & S data sets,  
13 respectively. Because we removed the papers from the M category, which are more ambiguous to code,  
14 our inter-reliability measures may be somewhat inflated. Nonetheless, we obtained high inter-observer  
15 reliability of 0.96 (agreement for 214 out of 224 papers) for the SS data set and 0.97 (agreement for 225  
16 out of 231) for the SS & S data set.  
17  
18  
19  
20  
21  
22

### 23 *Citation networks*

24 We created citation networks by importing both data sets into the Science of Science (Sci<sup>2</sup>) Tool (Sci2  
25 Team 2009) and producing a “directed citation network”; that is, each citation in the network was  
26 represented by a unidirectional arrow pointing from the paper being cited (“source”) to the paper  
27 performing the citation (“target”). The initial networks included all articles that the papers in our coded  
28 data sets had cited. These initial networks included > 30,000 articles, more than an order of magnitude  
29 larger than the coded data sets. However, because the additional articles incorporated into the network by  
30 Sci2 had not been assigned a category code (E, T, R, or M) and were often focused on topics other than  
31 sexual selection and speciation, we excluded those citations (“links”) from the subsequent analysis. In  
32 other words, we analyzed only those links where both the paper performing the citation (“target”) and the  
33 paper being cited (“source”) were in our original, coded data set. Henceforth, we refer to this subset of the  
34 total network as the “coded network” (see Table A1 for number of papers in each category; some  
35 additional analyses using the entire networks are presented in Appendix 2; coded network files to be  
36 uploaded to Dryad upon acceptance). Finally, throughout this paper, we use the Sci2 network language —  
37 “target” paper cites “source” paper—to refer to the directionality of citation behavior. We recognize that  
38 this language is not completely intuitive but were unable to identify a preferable alternative that is both  
39 intuitive and precise.  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50

### 51 *Dyadic citation frequencies*

52  
53 To analyze basic citation patterns, we calculated citation frequencies for each of the sixteen possible  
54 citation dyads (target cites source) for both data sets. That is, the dyads take the form of: E cites E, E cites  
55  
56  
57  
58  
59  
60

1  
2  
3 T, E cites R, E cites M, T cites E, T cites T, and so on. Frequencies were calculated as the number of  
4 times a target ( $t$ ) of type  $x$  ( $t_x$ ) cited a source ( $s$ ) of type  $y$  ( $s_y$ ) out of all the possible citations by targets of  
5 type  $x$ , where  $x$  and  $y$  can be E, R, T, or M. In other words, when considering the citation pattern for  
6 empirical citations of theoretical research, this calculation asks, “of all the times E papers performed an  
7 act of citation, what fraction of those citations were of T papers?” We also considered the inverse  
8 question; that is, “of all the times E papers were cited, what fraction of those citations came *from* T  
9 papers?” These alternative perspectives did not produce substantially different patterns and we therefore  
10 focus on the former for the remainder of this article, referencing the latter only in a summary figure (Fig.  
11 1).  
12  
13  
14  
15  
16

### 17 18 *Citation rates for specific empirical systems*

19  
20 In order to probe hypothesis (iv) from the main text—that citation patterns might be influenced by the  
21 narrower specialization of theoretical biology—we followed up with an additional analysis. Specifically,  
22 we compared the citation rates of theory by theory to the citations rates within specific biological systems.  
23 We used only one of our data sets for this comparison, the SS & S data set, because the empirical papers  
24 in that data set were distributed across fewer study organisms, forming unambiguous clusters of papers  
25 focused on particular systems. We identified the four most common study systems: cichlids (182 articles),  
26 *Drosophila* (112 articles), sticklebacks (51 articles), and *Heliconius* (13 articles). For each paper within  
27 these subsets, we identified the number of citations that were of other studies from the same system. In  
28 other words, we asked, “how frequently do cichlid papers cite other cichlid papers, and how often do  
29 *Drosophila* papers cite other *Drosophila* papers, etc?” We were then able to compare the within-system  
30 citation rates for empirical study systems to the citation rates within theory.  
31  
32  
33  
34  
35  
36  
37  
38

### 39 *Analysis of indirect citations*

40  
41 Hypothesis (v) from the main text required an investigation of “indirect citations” that occur if researchers  
42 utilize review articles to summarize general patterns that are evidenced by individual articles. For this  
43 purpose, we created additional “indirect-citation networks”, whose links represent interactions of the kind  
44 “target cites review that cites source”. We then calculated dyadic frequencies as above (e.g., E cites R that  
45 cites T). Note that for this analysis we *only* included indirect citations through review articles.  
46  
47  
48  
49

### 50 *Analysis of cross-citations*

51  
52 H hypothesis (vi) from the main text posits that cross-citations between E and T might be concentrated on  
53 a small number of source papers. To investigate this hypothesis, we conducted two complementary  
54 analyses. First, we counted the number of E, R and T papers that cite or are cited by at least one paper  
55  
56  
57  
58  
59  
60

1  
2  
3 from each of these categories. In particular, this analysis shows how many empirical and theoretical  
4 papers are involved in cross-citations at all. Second, we compared the unevenness of the number of  
5 citations received by source papers from both within and across categories (again leaving out the M  
6 category). As a measure of unevenness, we use the Gini coefficient, which is most well-known as a  
7 measure of income inequality in economics (Gini 1936), but has also been used as a measure of species  
8 unevenness in studies of biodiversity (Wittebolle et al. 2009) and even in citation network analyses  
9 (Clough et al. 2015). The Gini coefficient estimates statistical dispersion and takes values ranging from 0  
10 (complete evenness) to 1 (maximal unevenness). In our case, the coefficient is equal to 0 if all papers  
11 receive the same number of citations, and equal to 1 if all citations are concentrated on a single paper.  
12 Because the Gini coefficient strongly depends on sample size (Deltas 2003)—in our case, on the total  
13 number of observed citations—we used a rarefaction technique (e.g. Colwell 2009) to control for this  
14 effect (see Appendix 2 for detailed description of these methods).  
15  
16  
17  
18  
19  
20  
21  
22

### 23 *Analysis of highly influential papers*

24  
25 To investigate hypothesis (vii) from the main text—that influential papers in one category are likely to  
26 also be influential in the other—we excluded papers from the R and M categories and compared only T  
27 and E. For each paper in the network of either type, we created estimates of “influence” by calculating the  
28 proportions of total E papers and total T papers in the network that cited it. We were then able to compare  
29 the proportion of citations that each paper received from its own category to the number of citations that it  
30 received from papers in the other category. This comparison asks, for example, “are E papers that are  
31 highly cited by other E papers also cited heavily by T?” Because recently published papers, which have  
32 not yet had time to become high-impact papers, might artificially bias the distribution toward apparent  
33 “low-impact” papers, we excluded from this analysis any source paper that was published within the last  
34 five years of our study interval (i.e., between 2009 and 2013). We also excluded papers that have received  
35 zero citations from both categories, since these would artificially inflate the correlation. To corroborate  
36 our conclusions, we performed a bootstrap analysis, which produced a distribution of correlation  
37 coefficients describing the relationship between influence in one category and influence in the other. This  
38 analysis is complemented by an analysis of highly-cited first authors in the global networks (described in  
39 Appendix 3; raw data files to be uploaded to Dryad upon acceptance).  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49

### 50 *Co-citation networks*

51  
52 Finally, as a complement to the analysis of the directed citation networks, we generated co-citation  
53 networks, which are weighted *undirected* networks of cited papers. To construct such a network, an edge  
54 is drawn between two papers if they were both referenced by the same target paper. Consequently, co-  
55  
56  
57  
58  
59  
60



1  
2  
3 citations reflect decisions about which citations ‘belong together’ in the literature and illustrate patterns of  
4 general consensus about connections between papers. This type of network allows for inferences about  
5 how papers from the different categories are viewed by the community, providing additional insight into  
6 some of our hypotheses. After creating the co-citation network, we then summed the co-citation events  
7 across all target papers. Finally, to correct for the association between co-citation counts and the age of a  
8 paper, we re-scaled the total citation count between papers to co-citations per year.  
9  
10  
11  
12  
13

14 We analyzed these co-citation networks in two ways. First, we measured the extent to which papers  
15 contribute to the structure of the co-citation network using three common measures of network centrality:  
16 degree strength, eigenvector centrality, and betweenness (see Appendix 4 for more details on these  
17 metrics). For each network metric, we used the undirected weighted co-citation networks (with edge  
18 weights re-scaled by the paper age, see above) to find the centrality of each node (using the R package  
19 ‘igraph’, Csardi & Nepusz 2006). To facilitate comparisons across centrality metrics, we renormalized all  
20 centrality values to range between 0 and 1, where 1 is the maximum observed value for each metric.  
21  
22  
23  
24  
25  
26

27 Second, to convey an impression of the network structure, we created graphical representations of subsets  
28 of the co-citation networks. To prevent these visualizations from becoming too dense and difficult to  
29 interpret, we pruned the networks using two criteria. Papers were included in the visualized co-citation  
30 networks 1) if they were unusually central to a network (by at least one of the three metrics) and 2) if they  
31 were co-cited with other papers above a certain minimum threshold (15 for the SS co-citation network  
32 and 40 for the SS & S data set). These pruning decisions result in visualized networks of papers that are  
33 both highly co-cited and highly central in the network. Because we visualized the network based on the  
34 total number of local citations (the number of times a paper was cited by another paper in the network),  
35 the papers included are somewhat age-biased towards older papers. Therefore, as a complement to the  
36 visualized network, which can be viewed as an illustration of the ‘foundational literature’ of a field, we  
37 also present the relative ranking of papers that were highly central (see co-citation in Results section).  
38  
39  
40  
41  
42  
43  
44  
45

#### 46 *Quantitative approach*

47  
48 In all of the analyses presented here, we take a descriptive approach. Although we employ some statistical  
49 tools (e.g. correlation coefficients and bootstrap analyses; (R CoreTeam 2017)) in order to illuminate  
50 patterns in the data, we intentionally do not test for statistical significance and thus our results do not  
51 present p-values. We opted for this approach for a number of reasons. First, our datasets do not represent  
52 random samples of a larger population; specifically, the SS data set was restricted in a non-random way  
53 and the SS & S data set is exhaustive within the chosen criteria. Second, our datasets are so large that  
54  
55  
56  
57  
58  
59  
60



1  
2  
3 almost any difference will turn out to be statistically significant. Finally, to present p-values would be to  
4 assume a meaningful null hypothesis against which to compare alternatives. Although our second  
5 hypothesis—that citation frequencies are merely driven by the frequencies of each paper type in the  
6 literature—represents a null hypothesis in the strictest sense, it does not represent a reasonable  
7 expectation; even in the idealized scenario, where theory and empiry are optimally integrated, we should  
8 expect (and even hope for) some separation between the two approaches. That is, empirical and  
9 theoretical approaches are different, and some signature of that difference in the citation patterns is  
10 expected even if theory and empiry were highly integrated. Thus, we present quantitative descriptions  
11 where appropriate and refrain from comparing against an artificial and therefore inappropriate null  
12 hypothesis.  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Draft Manuscript

Table 2: Number and proportion of papers from each category type in the data sets. The column labeled “Total papers” indicates the number of papers by type contained in our original dataset. The column labeled “cited papers” indicates the number of papers in the dataset that were cited at least once by another paper in the dataset. The numbers in each column are also presented as proportions, represented in parentheses.

Paper category	# SS papers (proportion)		# SS & S papers (proportion)	
	Total papers	Cited papers	Total papers	Cited papers
Empirical (E)	713 (0.70)	676 (0.69)	998 (0.74)	969 (0.74)
Review (R)	188 (0.18)	183 (0.19)	176 (0.13)	171 (0.13)
Theory (T)	94 (0.09)	93 (0.10)	162 (0.12)	160 (0.12)
Mixed (M)	24 (0.02)	24 (0.02)	18 (0.01)	18 (0.01)
Total papers	1019	976	1354	1318

## Appendix 2

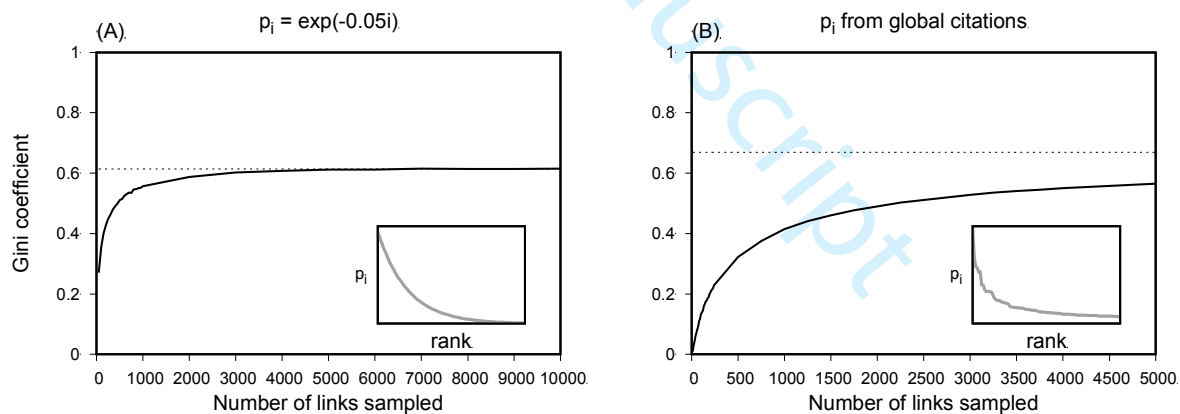
### *Unevenness of the distribution of cross-citations*

To test our hypothesis vi (“Cross-citations concentrate on small number of ‘crossover’ publications”), we compared the unevenness of the distribution of cross-citations versus within-category citations using the Gini coefficient (Gini 1936). The Gini coefficient can be calculated in the following way (in practice, we used the `ineq` library for R): One first determines the number of citations from, say, empirical papers received by each theoretical paper (where papers without citations are excluded from the analysis). These numbers are then arranged in ascending order, and the proportion of all  $E \rightarrow T$  citations is plotted as a function of the proportion of all T papers (e.g. the 50% lowest-cited T papers receive a total of 10% of all citations from E), yielding the so-called Lorenz curve. The Gini index is given by twice the area between the Lorenz curve and the main diagonal. It is equal to 0 if all papers receive the same number of citations, and equal to 1 if all citations are concentrated on a single source paper.

A major problem with this approach is that estimates of the Gini coefficient strongly depend on sample size (in our case, on the total number of observed citations; Deltas 2003). To better understand this issue, we consider the following simple model. Let there be  $n$  source papers and a pool of potential target papers. We will make the simplifying assumption that each source paper has a fixed probability  $p_i$  ( $i = 1..n$ ) of being cited by any given target paper (so target papers make an independent citation “decision” for each potential source paper, based on the respective probability). It is the Gini coefficient of the vector  $(p_1, \dots, p_n)$  that we wish to estimate, and the target papers (or rather, the citations they create) are a sample that allows us to do so (by providing estimates of the  $p_i$ ). How does the estimated Gini coefficient depend on sample size? – Figure A1 shows simulation results for two different  $p$ -vectors, the first one proportional to an exponential function and the second based on the distribution of *global* citations among all papers in the SS&S data set. In both cases, we simulated target papers until a given number of citations was reached. We also compared different mean numbers of citations per target paper, by multiplying the  $p$ -vectors with a constant  $c$ . Our results show that the estimated Gini coefficient increases strongly with the total number of citations in the sample, eventually approaching the true Gini coefficient specified by the  $p$ -vector. (In the example in Fig. A1(A), intermediate sample sizes lead to a very slight overestimation of the Gini coefficient, but this effect is negligible relative to the underestimation at small sample sizes.) In contrast, the mean number of citations per target paper (the parameter  $c$ ) has almost no influence on the estimate, as long as the number of target papers is not too small (not shown; the reason is that a single target paper necessarily cites all its sources with complete

evenness).

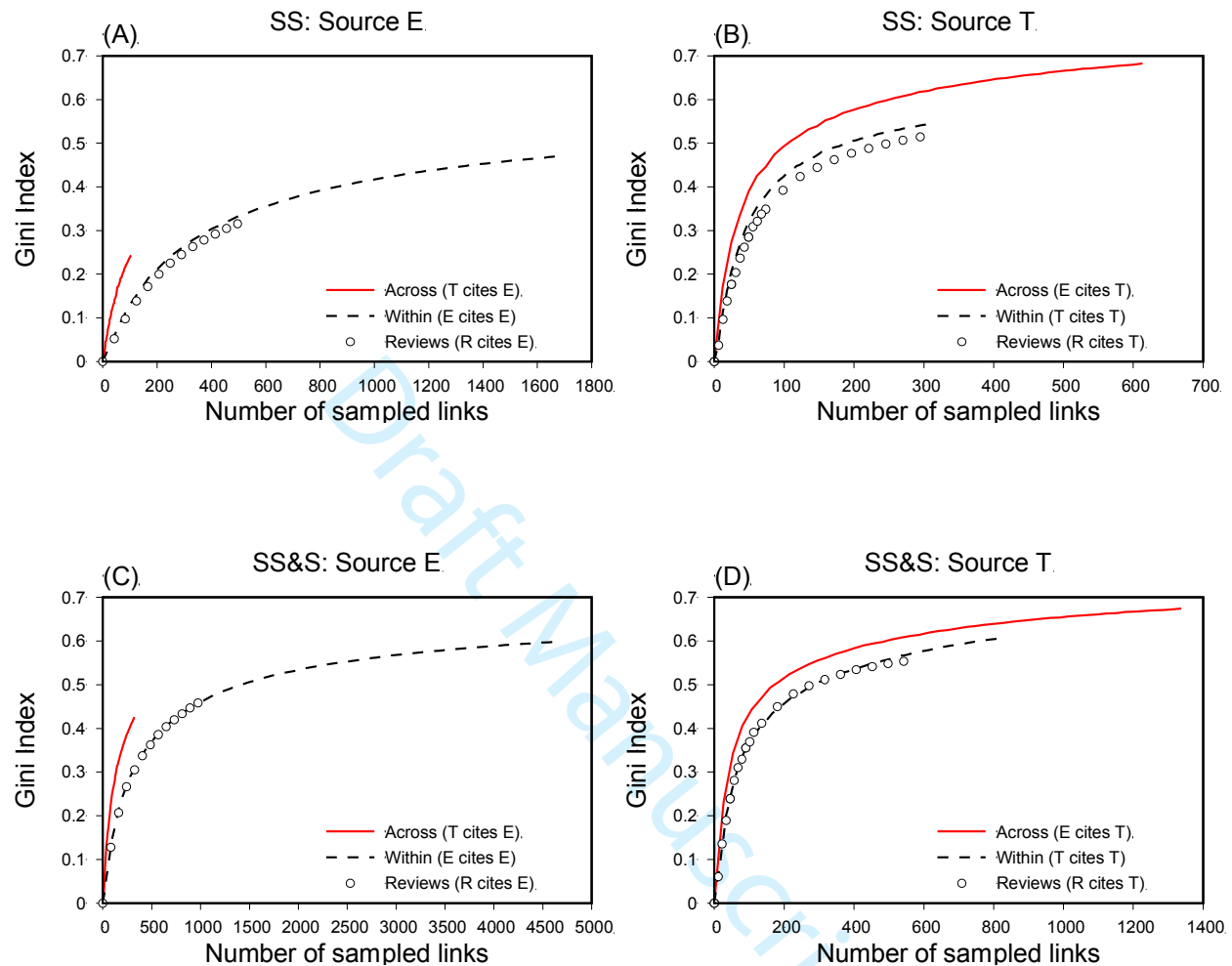
Coming back to our data set, we have three different sets of target papers E, R and T (leaving out M) with different associated  $p$ -vectors. Our aim is to estimate and compare the Gini coefficients of these  $p$ -vectors, but we need to account for the different sample sizes (different numbers of links created per target set). To do so, we use a technique of rarefaction (well-known for estimates of biodiversity, e.g. Colwell 2009), that is, we artificially create equal sample sizes by subsampling the target sets. Fig. A2 shows the resulting rarefaction curves, which show the Gini coefficients as a function of the number of sampled links, averaged over 500 different samples (each sample was produced by randomly sampling target papers without replacement until the desired number of links was reached). As can be seen, the curve for cross-citations is always above the one for within-category citations, confirming our hypothesis that cross-citations are indeed more biased towards heavily cited papers. Nevertheless, the difference is arguably not as large as one might have expected. This is also highlighted by the fact that a large proportion of T papers receives at least one cross-citation, as shown in Fig. S5 and S6. It is also noteworthy that citations by reviews create Gini coefficients comparable to those of within-category citations (or even a bit lower, in the case of the SS dataset).



**Figure A2:** Simulation results illustrating the dependence of the estimated Gini coefficient on the total number of links in the sample. Insets show the shape of the citation probabilities specified by the  $p$ -vector. The Gini coefficient is a measure for the unevenness of this vector. (A) 100 source papers with ordered citation probabilities  $p_i = 0.1 * \exp(-0.05*i)$ . (B) 1354 source papers with citation probabilities proportional to the number of global citations in the SS&S dataset (the highest-cited paper has  $p_1 = 0.1$ ).

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Draft Manuscript



**Fig. A3:** Rarefaction curves for Gini coefficients for different target/source combinations (see Appendix text for details). When corrected for sample size, across-category citations ( $E \rightarrow T$  or  $T \rightarrow E$ ) have higher Gini coefficients (higher unevenness) than within-category citations ( $E \rightarrow E$  or  $T \rightarrow T$ ) or citations by Reviews. The different lengths of the curves reflect the different sample sizes: The right-most end of each curve shows the Gini-coefficient for the entire data set.

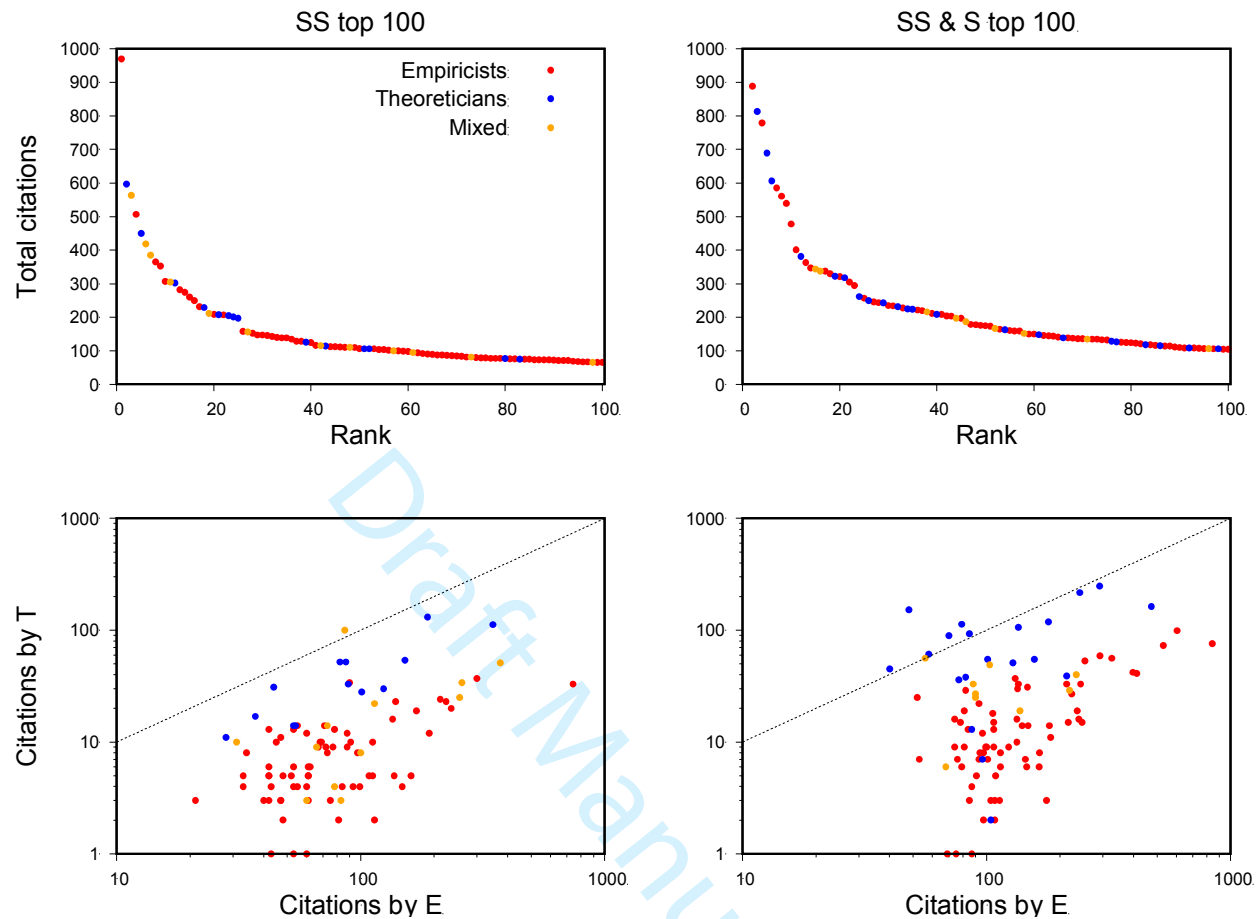
### Appendix 3

#### *Some results from the global network*

The majority of our analyses are based exclusively on the coded network (where both target and source

1  
2  
3 papers had been assigned a category). Nevertheless, the dataset extracted from the Web of Science also  
4 contains information about non-coded source papers, even though this information is in less accessible  
5 form (e.g., first authors only, not corrected for citation errors, alternative spellings etc.). To get at least  
6 some insight from these data, we conducted an additional analysis.  
7  
8  
9

10  
11 We first grouped citations by first author (the Web of Science data do not provide the names of co-  
12 authors). We then determined the top 100 first authors cited by papers from our datasets (see Fig. A1), as  
13 well as the top 100 first authors cited by papers from the E, R and T categories separately (the M category  
14 was excluded from this analysis). In addition, we designated authors as either empiricists, theoreticians or  
15 “mixed”, based on their lists of publications. The complete results are available in supplementary file  
16 authors\_top100.xls. In addition, Figure A1 and Table A2 and A3 analyze the distribution of the citations  
17 going to the top 100 overall first authors. The results largely confirm our conclusions from the coded  
18 network: Both empiricists and theoreticians are well-represented among highly cited authors, but each  
19 category cites more authors from their own category (e.g., there are about twice as many theoreticians  
20 among the top 100 authors cited by theoretical as opposed to empirical papers). In addition, there is  
21 considerable overlap and correlation between citations received from the different classes of papers (e.g.,  
22 in the SS & S dataset, the top-ranked author as cited by empirical papers, O. Seehausen, is ranked 11th  
23 regarding citations by theoretical papers, and the top-ranked author as cited by theoretical papers, M.  
24 Kirkpatrick, is number 9 by empirical papers).  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



**Figure A1:** Top hundred first authors cited by papers from our two datasets, with authors color-coded as empiricists (red), theoreticians (blue) or “mixed” (orange). The top row shows number of citations ordered by rank. The bottom row shows the relationship between citations received from empirical and theoretical papers in the two datasets on a double logarithmic scale. Spearman rank correlation coefficients are 0.394 for the SS dataset and 0.278 for the SS&S dataset, across all types of authors (but note that the top 100 authors are not a random sample of all authors).



		First author				
			E	T	M	Total
Source category	Top 100	All	76	15	12	103
		E	77	13	14	104
		R	82	15	8	105
		T	63	36	21	120
	Overlap	E/R	45	11	8	64
		E/T	41	12	8	61
		R/T	43	14	7	64

**Table A2:** Analysis of the top 100 first authors cited by papers in the sexual selection (SS) dataset. Authors are classified as empiricists (E), theoreticians (T) or “mixed” (M). First four rows: Top 100 authors as cited by all papers, or by papers from the E, R and T categories separately. For example, among the 100 most-cited authors by papers in the SS dataset, 76 are empiricists, and among the 100 most-cited authors by theoretical papers from this dataset, 36 are theoreticians. Note that the total number may exceed 100 due to ties (e.g., in the first row, authors ranked 101-103 have the same number of citations as the author on rank 100). Last three rows: Numbers of authors appearing in the top 100 lists of two categories of papers. For example, 41 empiricists are among the 100 most-cited authors by both empirical and theoretical papers from the SS dataset.

		First author				
			E	T	M	Total
Source category	Top 100	All	70	22	9	101
		E	75	20	7	102
		R	78	19	10	107
		T	42	54	11	107
	Overlap	E/R	55	13	6	74
		E/T	32	14	7	53
		R/T	34	19	8	61

**Table A3:** Same as Table A2, but for the source papers from the sexual selection and speciation (SS & S) dataset. Note that, among the 20 theoreticians within the top 100 list of empirical papers, only 14 are also

1  
2  
3 in the top 100 list of theory papers (compare second and sixth entry of the T column). It turns out that the  
4 6 theoreticians cited more by empirical than theoretical papers mostly develop phylogenetic or other  
5 computational methods. For more details, see Table A2.  
6  
7  
8  
9  
10  
11

## 12 **Appendix 4**

### 13 *Some methods for the co-citation networks*

14  
15  
16 First, we quantified each focal paper's degree strength, which is the sum of the scaled co-citation counts  
17 for all papers directly connected to the focal paper. Degree strength is a local measure of a node's  
18 centrality in a network because it only measures how each node is directly connected to other nodes in the  
19 network. In the context of the co-citation network, a paper with high degree strength is one that is cited  
20 together with other papers in our dataset many times. We also quantified two more global network  
21 metrics, which both take the whole structure of the co-citation network into account: eigenvector  
22 centrality and betweenness. Eigenvector centrality is based on the concept that the strength of edge  
23 weights is transferred along or percolated through network connections. Nodes with high eigenvector  
24 centrality have strong connections to their neighbors, who have strong connections to their neighbors, and  
25 on through the network. In the context of the co-citation network, a paper with high eigenvector centrality  
26 is one that is co-cited many times with others, who are also heavily co-cited with others. Betweenness  
27 quantifies the extent to which a node serves as a bridge between less-connected components of the  
28 network, and is based on the number of 'shortest paths' in the network that include the node of interest. In  
29 the context of the co-citation network, a paper with high betweenness is one that bridges between  
30 subgroups of papers in the network that are not often co-cited.  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42

43 We used the R package 'beanplot' v1.2 to visualize differences in the distribution of co-citation network  
44 metrics by paper type (Kampstra 2008). Because the co-citation network was too dense to plot in its  
45 entirety, we made the several decisions about how to prune the network to facilitate visualization. We  
46 filtered each co-citation network to include only papers which were unusually central in the network in at  
47 least one of the three centrality measures. We set this threshold at 0.4 (i.e. a paper's centrality score  
48 needed to be more than 40% of the maximum centrality score, see threshold in Figures S10 and S11. We  
49 plotted all three centrality values for papers that met these conditions.  
50  
51  
52  
53  
54

### 55 *Some results from the co-citation networks*

1  
2  
3  
4  
5 Theory papers in both the SS and SS & S datasets had higher median degree strength and higher median  
6 eigenvector centrality than the other paper types (Figure S9, a, b, d, e). High degree centrality is a direct  
7 measure of how often papers are co-cited, and indicates that theory papers often occur together with many  
8 other papers. Eigenvector centrality summarizes both direct and indirect co-citation patterns, and a high  
9 score indicates that both the direct co-citation links and the co-citation patterns of neighbors and  
10 neighbors of neighbors are also strong.  
11  
12  
13  
14

15  
16 High betweenness scores indicate papers which are critical in bridging between more isolated groups of  
17 co-cited papers. Median betweenness values were similar across all paper types, but slightly higher for  
18 theory and mixed papers in SS and for review papers in SS & S (Figure S9 c, f). The highest betweenness  
19 value was a review paper (Emlen and Oring 1977) in the SS dataset and a theoretical paper (Lande 1981)  
20 in the SS & S dataset. In contrast, betweenness scores for empirical papers in both datasets were in  
21 general lower, and had a smaller maximum value, than both theory and review papers, indicating that  
22 empirical papers are less likely to be used to bridge between less-connected factions within the literature.  
23 Mixed papers in the SS & S dataset had the smallest range of centrality values across all three centrality  
24 measures, except for the SS dataset, where the maximum value of betweenness was higher than the  
25 maximum betweenness score for any empirical paper. In both the SS and SS & S datasets, mixed papers  
26 comprised a small fraction of high-centrality papers; in the SS dataset, only two papers had centrality  
27 scores above 0.4 (Figure S10b), while the SS & S dataset, none of the mixed papers had centrality higher  
28 than 0.4 for any of the centrality metrics (Figure S11b).  
29  
30  
31  
32  
33  
34  
35  
36  
37

38 The role of high betweenness is visualized in Figure S10a, where Emlen and Oring 1977 sit squarely  
39 between two subsets of highly cited papers. In the SS & S co-citation network, Lande 1981 was the most  
40 central paper across all centrality measures. In the visualized co-citation network, Lande's role in  
41 connecting two factions within the literature is also apparent, as the paper sits at the center of the network  
42 and links the left and right subsections (which are often cited with other papers from their 'side' but not  
43 often with papers from the other 'side') together.  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60