Gender-based homophily in collaborations across a heterogeneous scholarly landscape

Carole Lee
Department of Philosophy
University of Washington

Metascience 2019 Symposium
Stanford University
September 5-8

Joint work with:

Y. Samuel Wang

Business School

U Chicago

Elena Erosheva

Statistics;
Center for Social Sciences and Statistics; Social Work

U Washington

Jevin West

Information School

U Washington

Carl Bergstrom

Biology

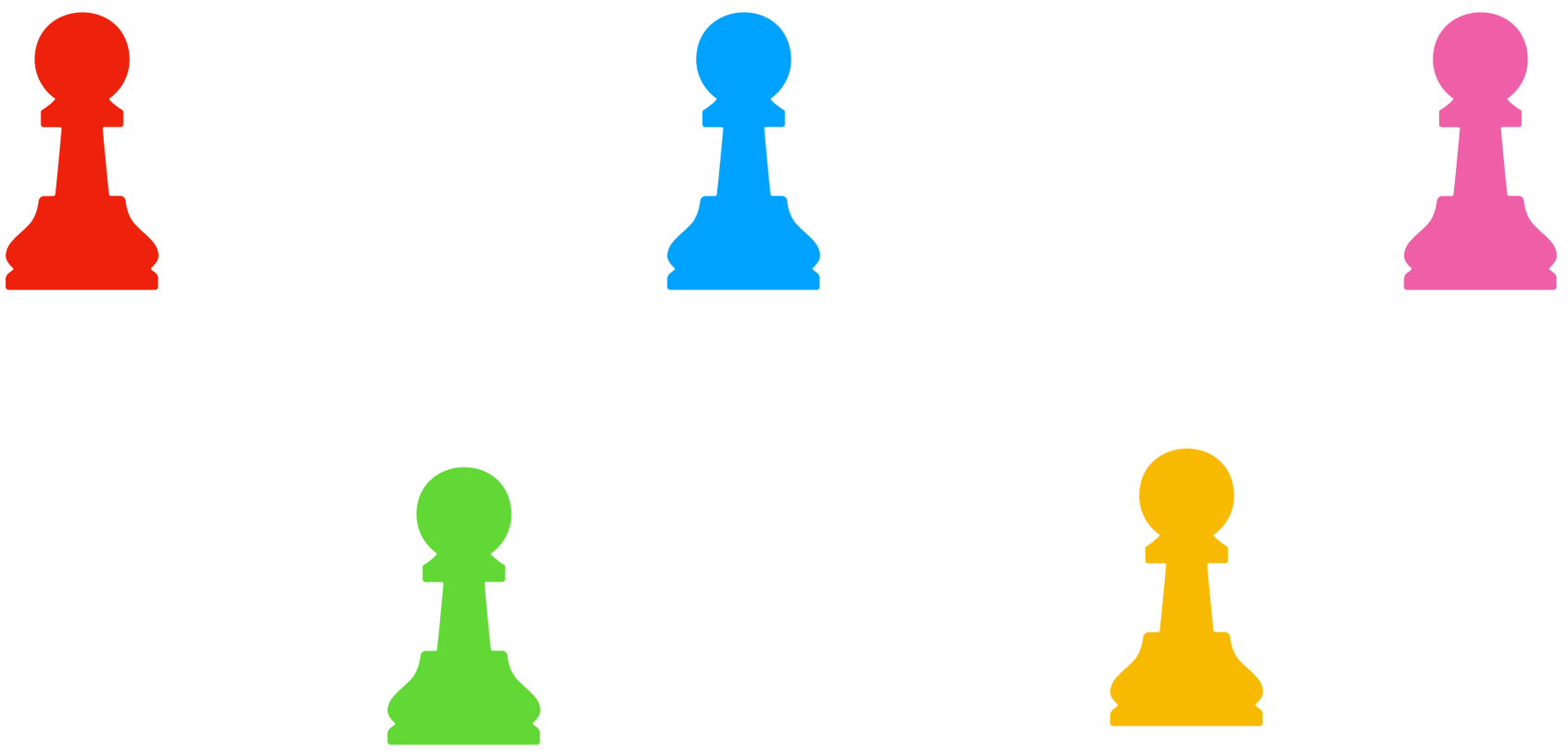
U Washington

Where do ideas come from?

Where do ideas come from?

Ideas come from people. . .

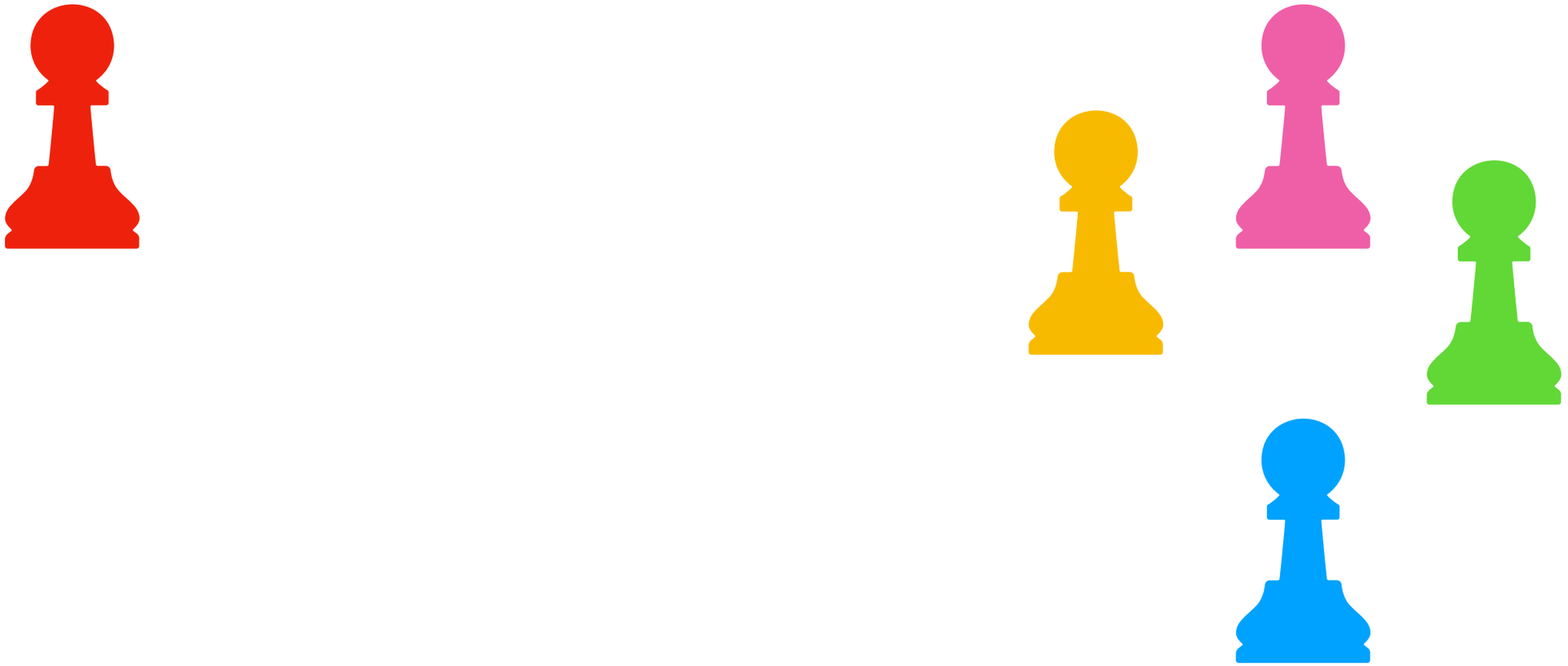
Where do ideas come from?



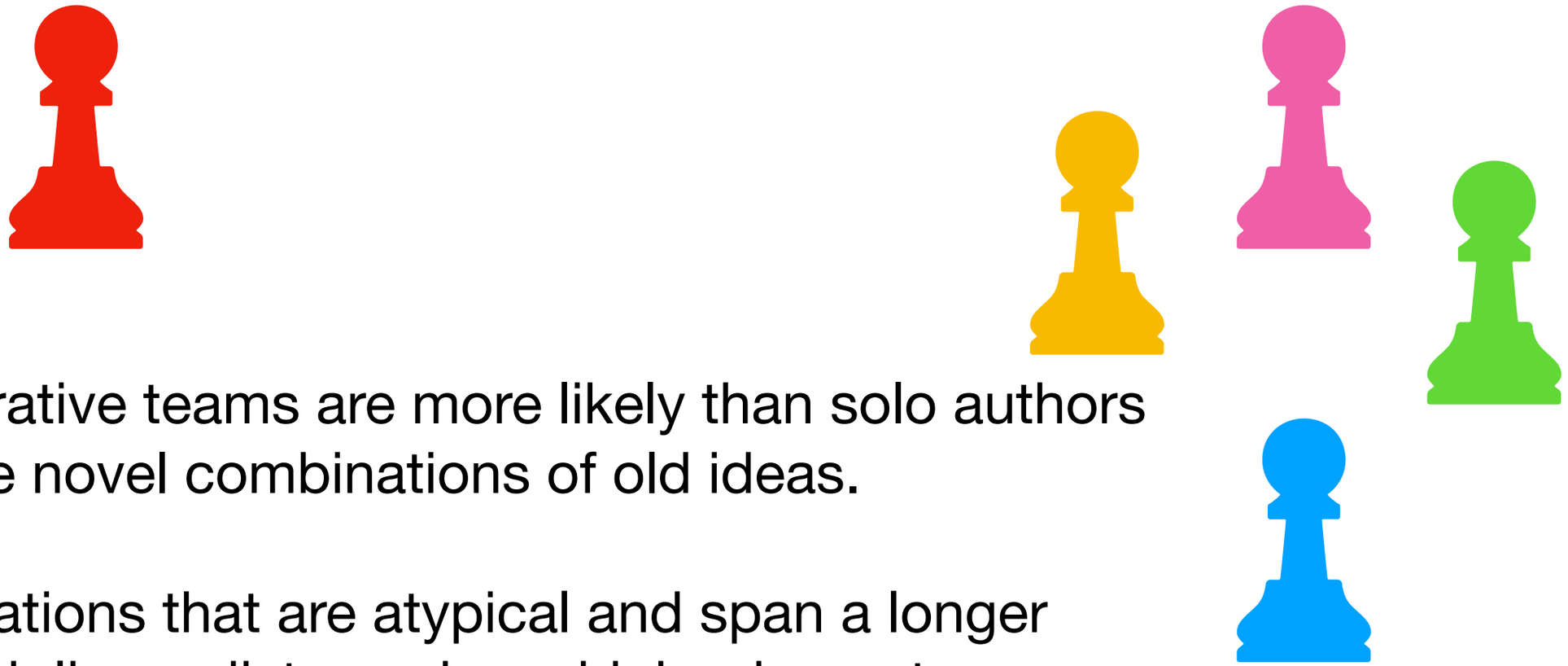
Different people have different forms of expertise.

Where do ideas come from?

We can create new ideas by combining folks with different forms of expertise into collaborative teams.



Where do ideas come from?



Collaborative teams are more likely than solo authors to create novel combinations of old ideas.

Combinations that are atypical and span a longer interdisciplinary distance have higher impact.

Uzzi, Brian, Satyam Mukherjee, Michael Stringer, & Ben Jones. "Atypical combinations and scientific impact." *Science* 342 (2013): 468-472.

Larivière, Vincent, Stefanie Haustein, & Katy Börner. "Long-distance interdisciplinarity leads to higher scientific impact." *Plos one* 10 (2015): e0122565.

Larivière, Vincent, Yves Gingras, Cassidy R. Sugimoto, and Andrew Tsou. "Team size matters: Collaboration and scientific impact since 1900." *JASIST* 66 (2015): 1323-1332.

How do teams self-assemble?

- Intellectual factors: expertise
- Instrumental factors: resources
- Social factors: respect, friendship, curiosity-seeking, fun

Katz, J. Sylvan, and Ben R. Martin. "What is research collaboration?." *Research policy* 26 (1997): 1-18.

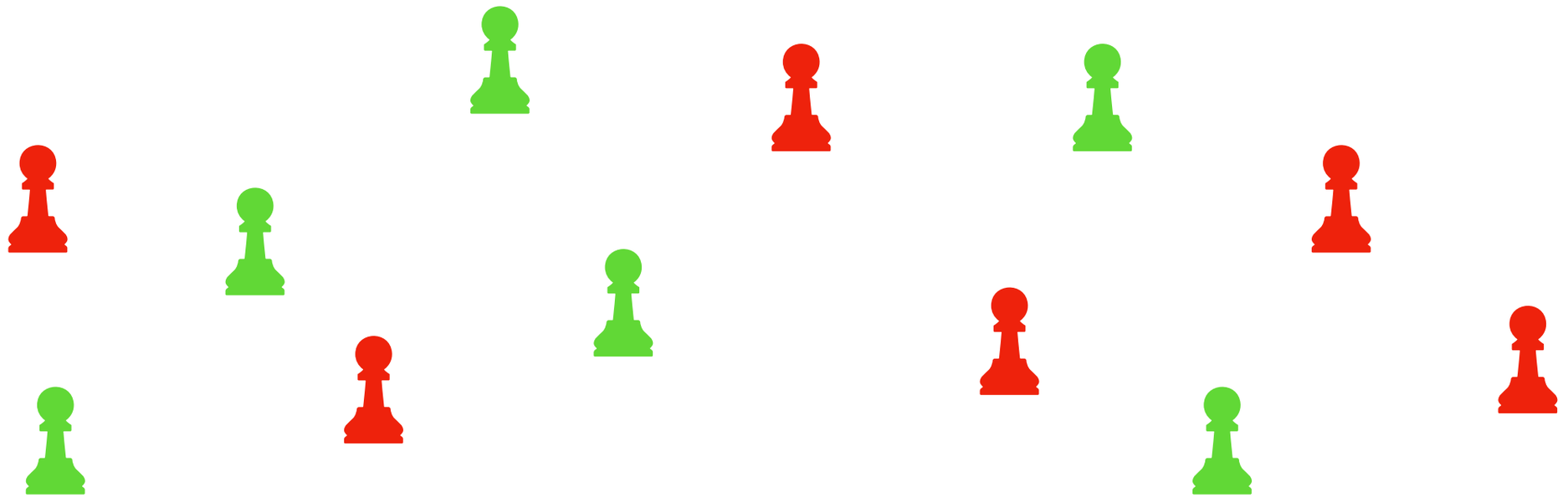
Melin, Göran. "Pragmatism and self-organization: Research collaboration on the individual level." *Research policy* 29 (2000): 31-40.

Maglaughlin, Kelly L., and Diane H. Sonnenwald. "Factors that impact interdisciplinary scientific research collaboration: Focus on the natural sciences in academia." (2005).

Hara, Noriko, Paul Solomon, Seung−Lye Kim, and Diane H. Sonnenwald. "An emerging view of scientific collaboration: Scientists' perspectives on collaboration and factors that impact collaboration." *Journal of the American Society for Information science and Technology* 54 (2003): 952-965.

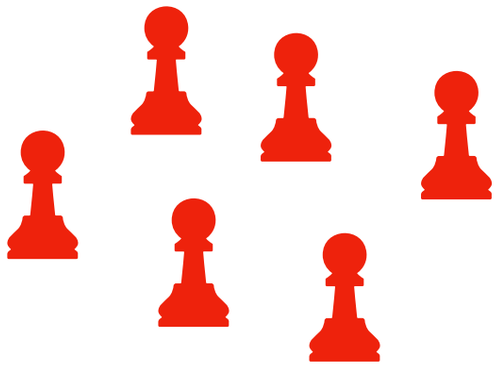
Collaboration involves social discretion

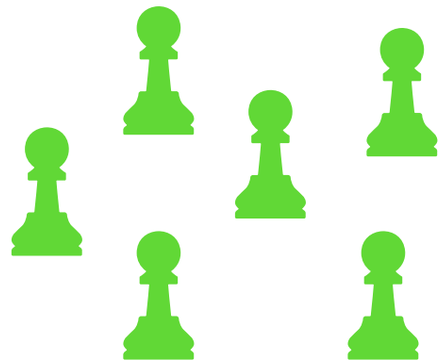
Does decreasing social distance increase collaboration?



Collaborative teams and social distance

- Homophily: Social similarly breeds connection.





- Gender homophily divides work environments, voluntary associations, and friendships.

McPherson, Miller, Lynn Smith-Lovin, and James M. Cook. "Birds of a feather: Homophily in social networks." *Annual review of sociology* 27 (2001): 415-444.

Kalleberg, Arne L., David Knoke, Peter V. Marsden, and Joe L. Spaeth, eds. *Organizations in America.* Sage, 1996.

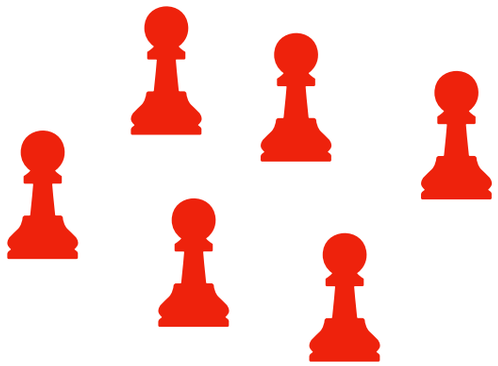
McPherson, J. Miller, and Lynn Smith-Lovin. "Sex segregation in voluntary associations." *American Sociological Review* (1986): 61-79.

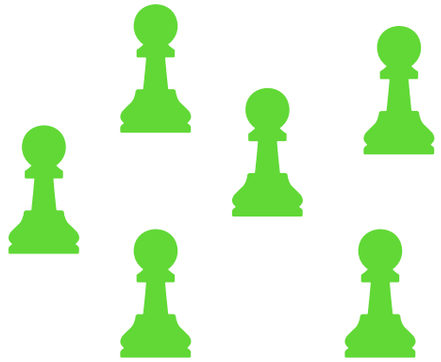
Ibarra, Herminia. "Homophily and differential returns: Sex differences in network structure and access in an advertising firm." *Administrative science quarterly* (1992): 422-447.

Ibarra, Herminia. "Paving an alternative route: Gender differences in managerial networks." *Social psychology quarterly* (1997): 91-102.

Question

- Is there gender-based homophily in collaborations across the heterogeneous scholarly landscape at varying levels of granularity?





- Women are less likely to co-author and serve as first or last author

- Given professional advantages of collaboration, it's important to understand how and under what conditions women collaborate

Larivière, Vincent, Chaoqun Ni, Yves Gingras, Blaise Cronin, and Cassidy R. Sugimoto. "Bibliometrics: Global gender disparities in science." *Nature News* 504, no. 7479 (2013): 211.

West, Jevin D., Jennifer Jacquet, Molly M. King, Shelley J. Correll, and Carl T. Bergstrom. "The role of gender in scholarly authorship." *PloS one* 8, no. 7 (2013): e66212.

Data

- JSTOR, a repository of papers across the humanities, social sciences, and natural sciences
    - We consider cited papers from 1960 onward
    - 252,413 multi-author papers with 807,588 authorships (i.e., instances of co-authoring)

Hierarchical clustering

- Apply hierarchical implementation of the InfoMap network algorithm to the citation network on the JSTOR corpus

- This reveals the hierarchical structure of corpus through the efficient coding of random walks on the citation network

Rosvall, Martin, and Carl T. Bergstrom. "Maps of random walks on complex networks reveal community structure." *Proceedings of the National Academy of Sciences* 105, no. 4 (2008): 1118-1123.

Rosvall, Martin, and Carl T. Bergstrom. "Multilevel compression of random walks on networks reveals hierarchical organization in large integrated systems." *PloS one* 6, no. 4 (2011): e18209.

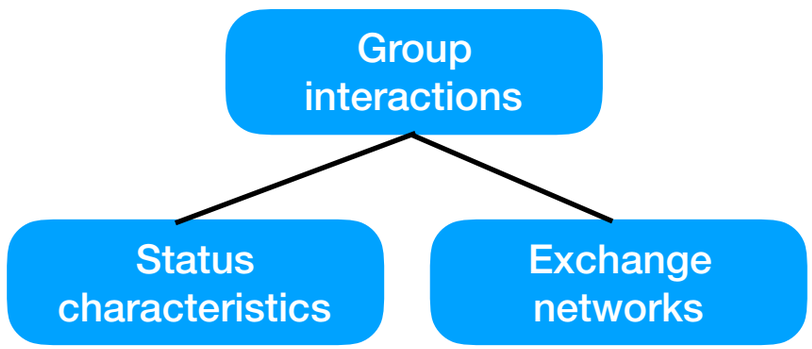
Hierarchical clustering

Status characteristics

Exchange networks

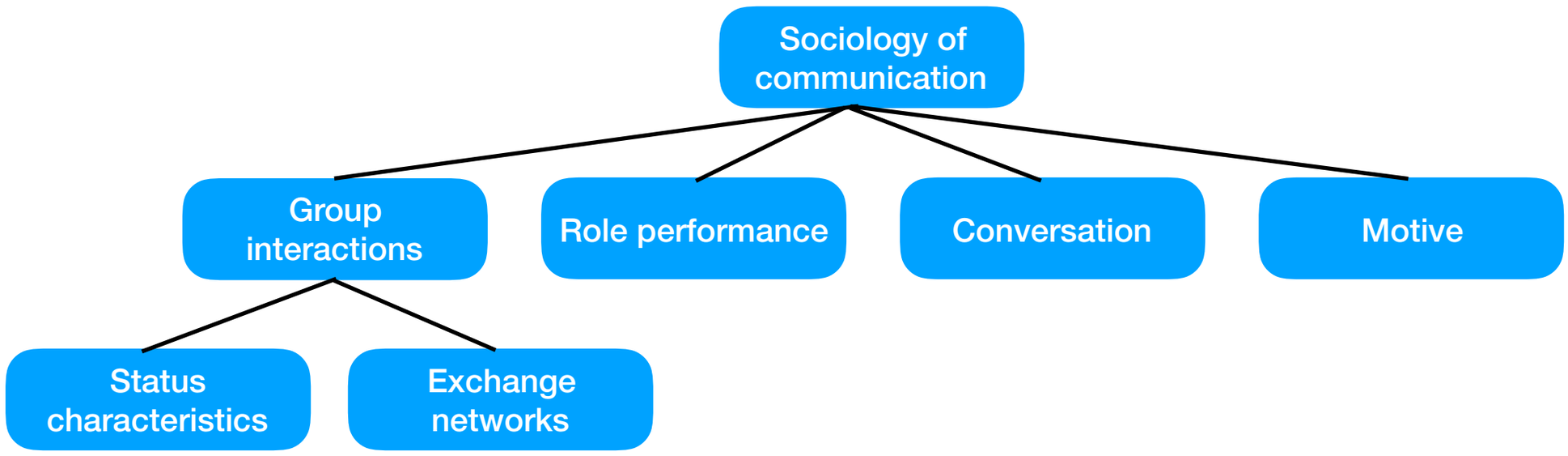
At the lowest level of clustering, each paper is grouped into one of 1,450 *terminal fields* which form the finest partition of the data.

Hierarchical clustering



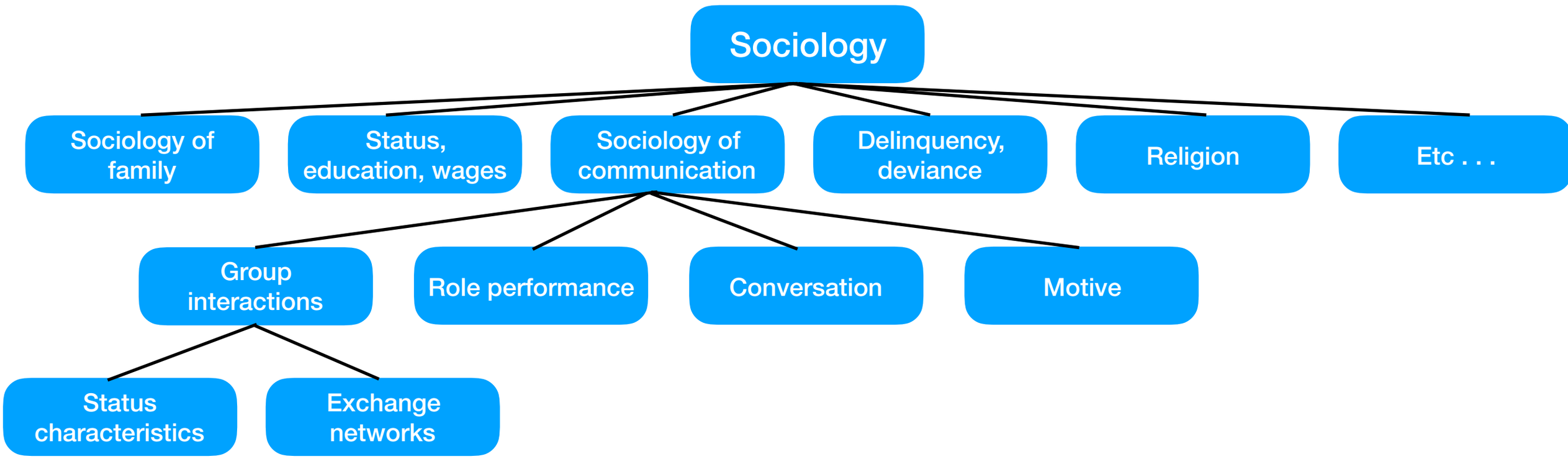
Each higher clustering forms progressively coarser partitions of the documents by aggregating terminal fields into the 280 *composite fields*.

Hierarchical clustering

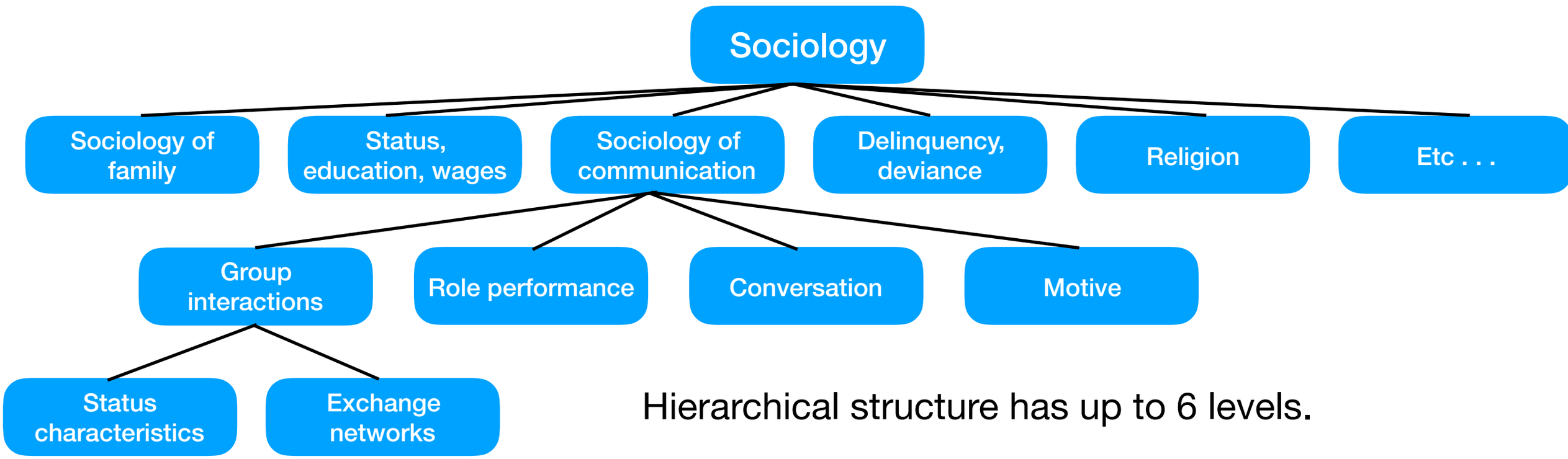


Hierarchical clustering

At the top-level, we have 24 *major fields*.



Hierarchical clustering

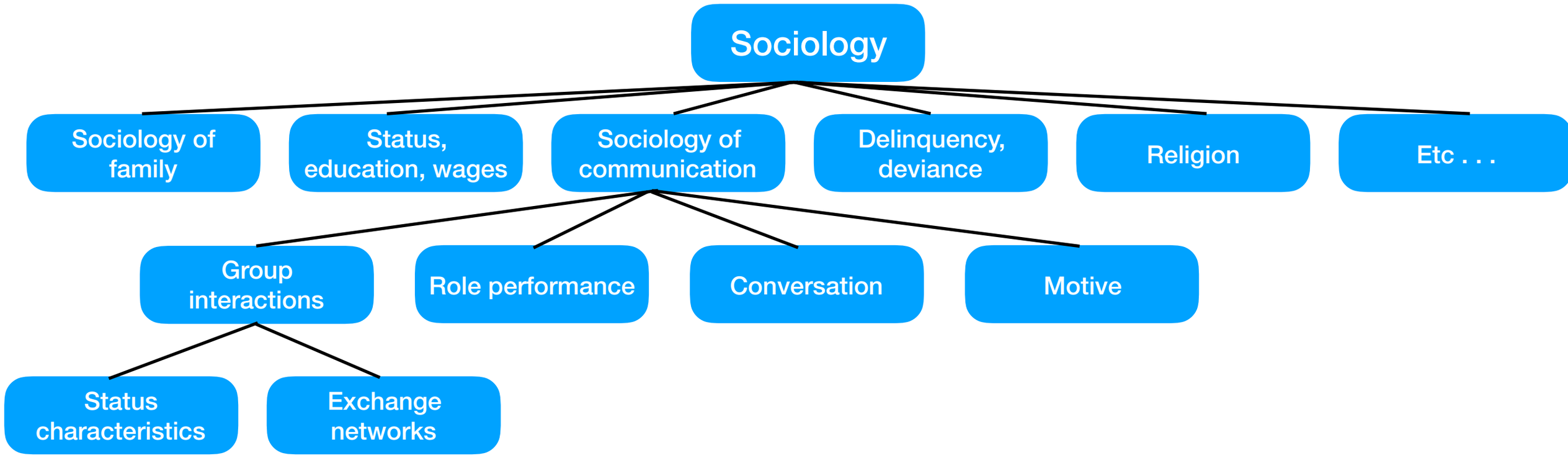


Hierarchical structure has up to 6 levels.

Hierarchical clustering

At any level, papers in a common field are more connected via citations than to papers from neighboring fields.

Fields at finer levels of the hierarchy are more connected than fields at coarser levels.



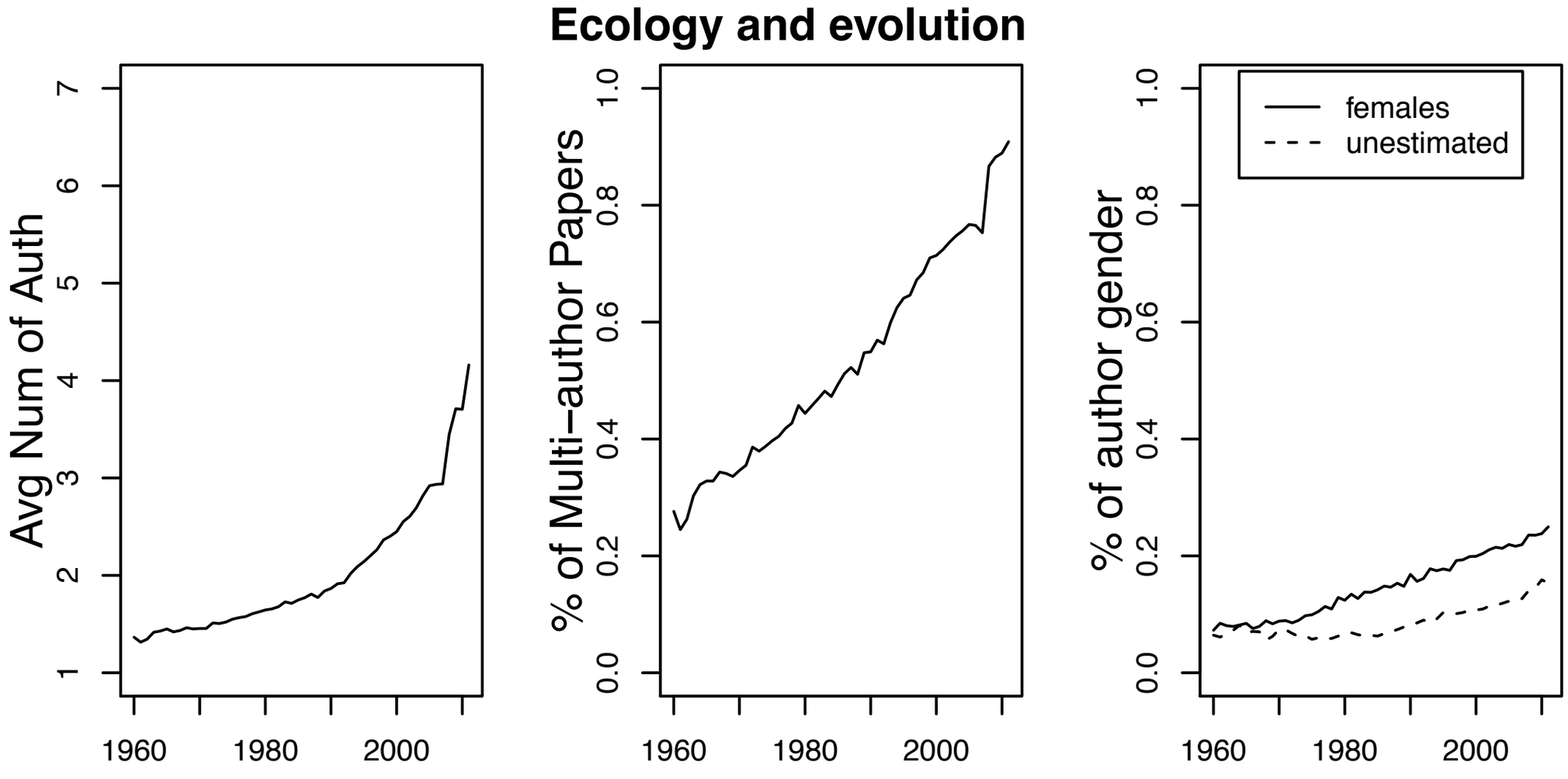
Determining gender

- Gender inferred from first names (≥95% certainty)
  - 75.3% gender determined via Social Security records
  - 12.6% determined via genderizeR (user profiles from social networks)
  - 12.1% gender unknown and omitted from our analysis:
    - 7.6% appear in neither data base
    - 4.5% are used for males and females with > 5% certainty
    - Rate of missingness compares favorably to previous studies

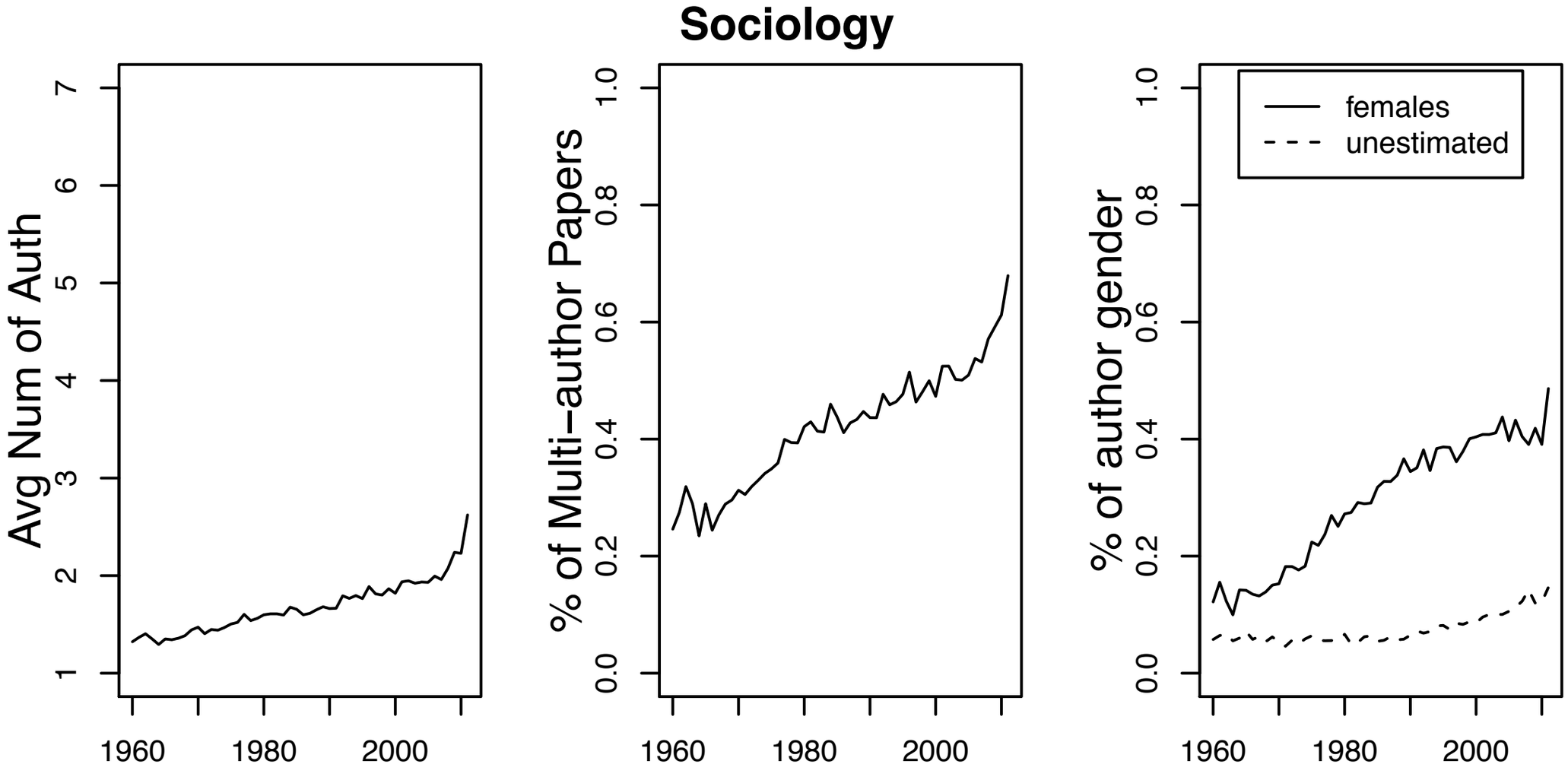
West, Jevin D., Jennifer Jacquet, Molly M. King, Shelley J. Correll, and Carl T. Bergstrom. "The role of gender in scholarly authorship." *PloS one* 8, no. 7 (2013): e66212.

Larivière, Vincent, Chaoqun Ni, Yves Gingras, Blaise Cronin, and Cassidy R. Sugimoto. "Bibliometrics: Global gender disparities in science." *Nature News* 504, no. 7479 (2013): 211.

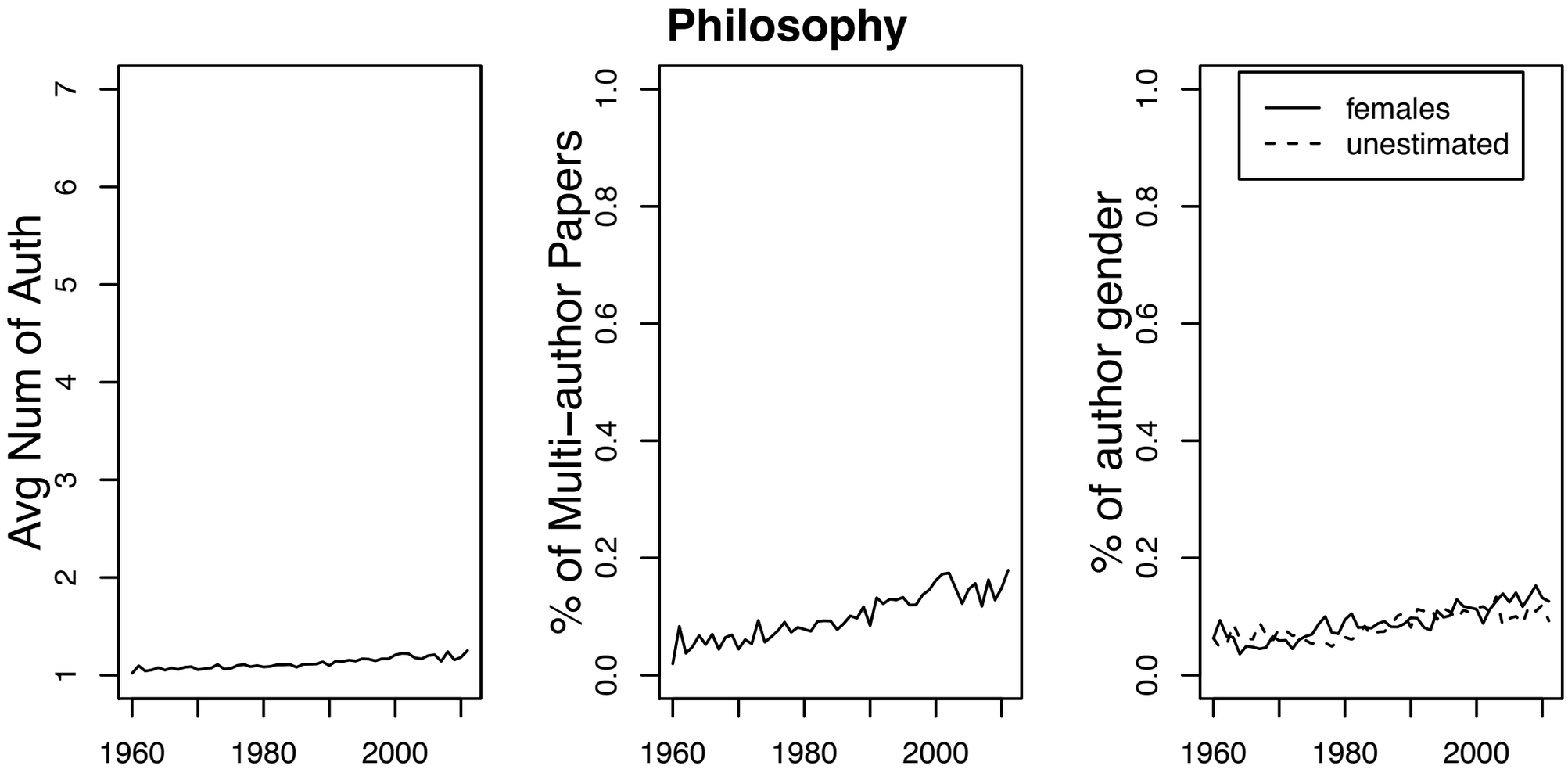
Gender and authorship practices across fields



Gender and authorship practices across fields



Gender and authorship practices across fields



Question

- Do males co-author with males (and females with females) more often than we would otherwise expect?

- We need to be careful about how we measure:

    1. How often do males co-author w/ males (and females w/ females)?

    2. Counterfactual: Is this more often than we would otherwise expect?

    3. If yes, then we have *behavioral homophily*.

1. Tendency for same-gender authorships to co-author

- We measure homophily by computing the difference in risks:
    - α = *P*(random co-author of random male is male) - *P* (random co-author of a random female is male)

1. Tendency for same-gender authorships to co-author

- We measure homophily by computing the difference in risks:
    - α = *P*(random co-author of random male is male) - *P* (random co-author of a random female is male)
- This measure is equivalent to:
    - Pearson correlation of gender indicators for random co-authorship pairs
    - Wright's *F* Coefficient of Inbreeding when all papers have two authors
    - Newman's network-based assortativity coefficient in an appropriately weighted network

Bergstrom, Theodore C. "The algebra of assortative encounters and the evolution of cooperation." *International Game Theory Review* 5, no. 03 (2003): 211-228.

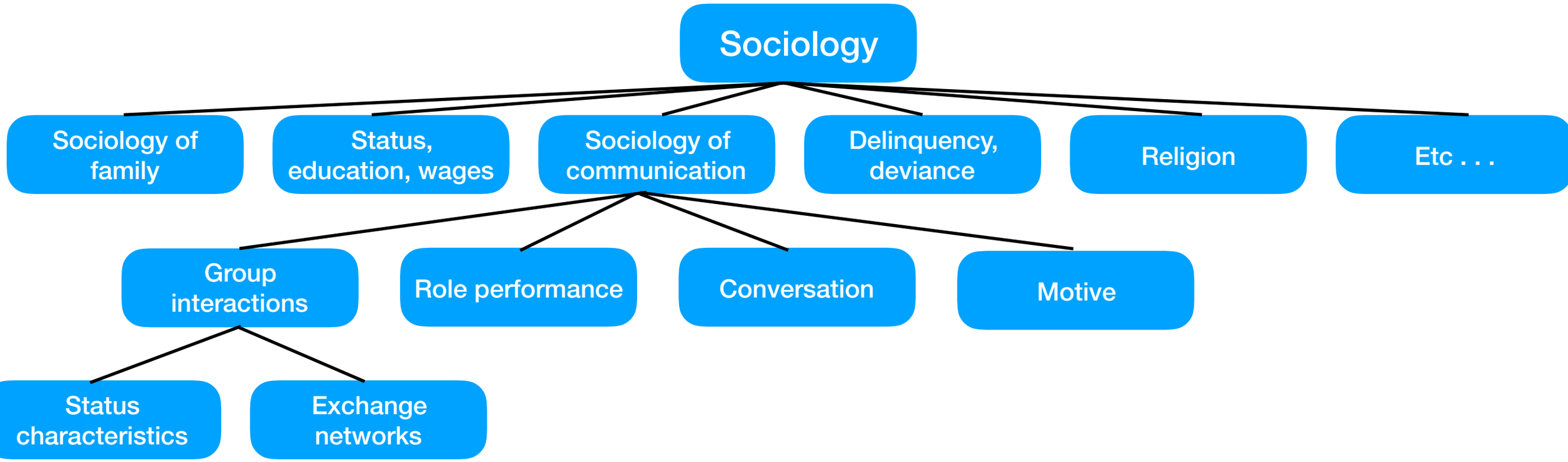
Wright, Sewall. "The genetical structure of populations." *Annals of eugenics* 15, no. 1 (1949): 323-354.

Bergstrom, T., M. Bergstrom, M. King, J. Jacquet, J. West, and S. Correll. "A note on measuring gender homophily among scholarly authors." (2016).

Wang, Y. Samuel, and Elena A. Erosheva. "On the relationship between set-based and network-based measures of gender homophily in scholarly publications." *arXiv preprint arXiv:1610.09026* (2016).

Newman, Mark EJ. "Mixing patterns in networks." *Physical Review E* 67, no. 2 (2003): 026126.

1. Tendency for same-gender authorships to co-author



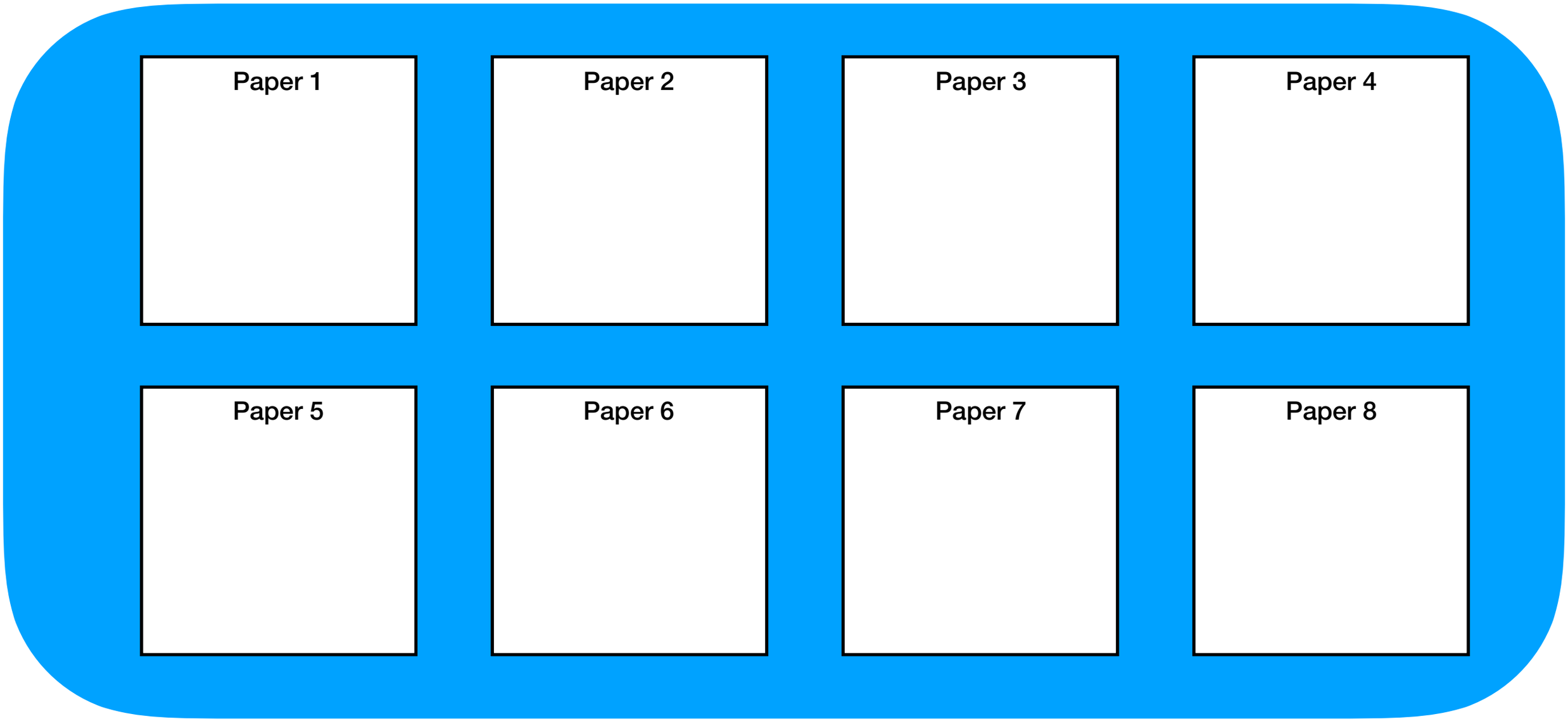
1. Tendency for same-gender authorships to co-author

Sociology

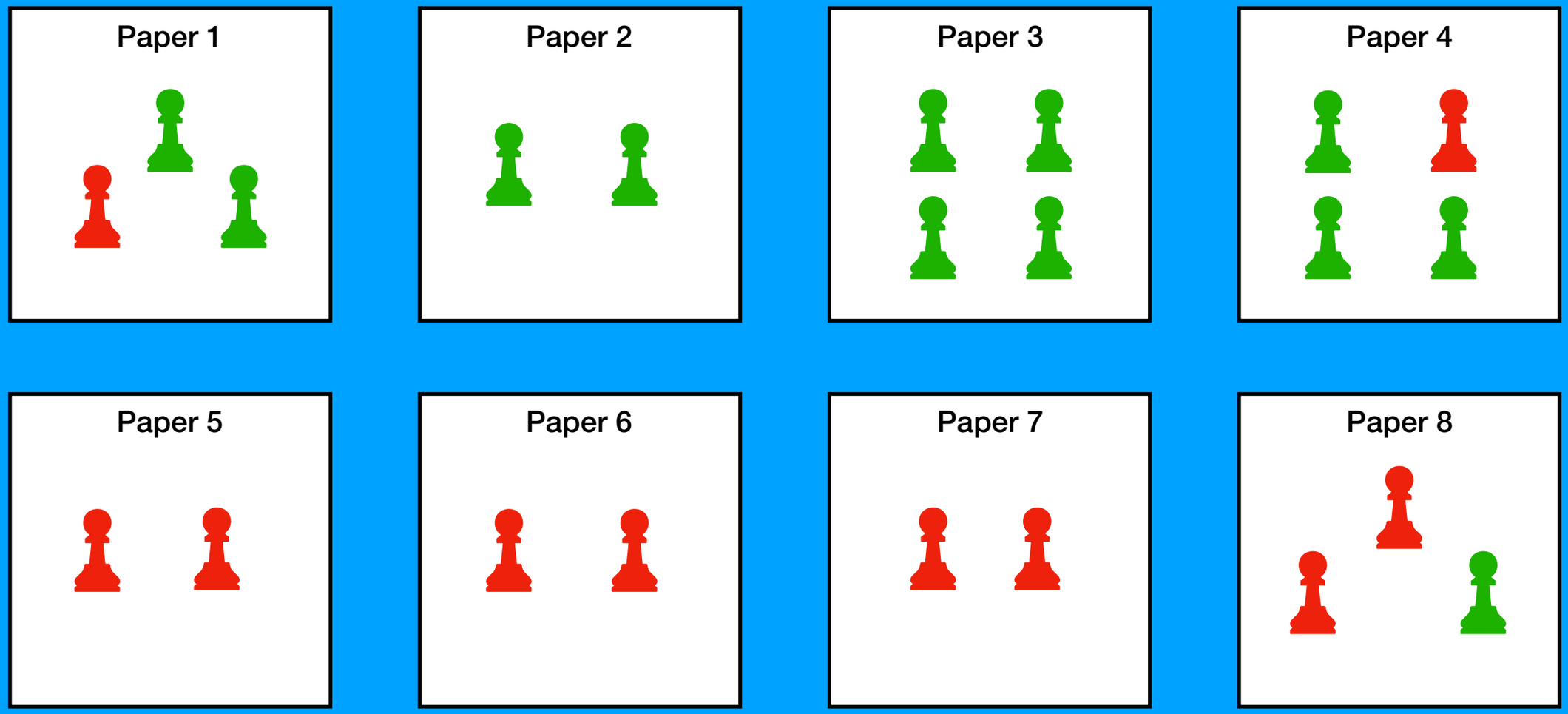
1. Tendency for same-gender authorships to co-author

Sociology

1. Tendency for same-gender authorships to co-author

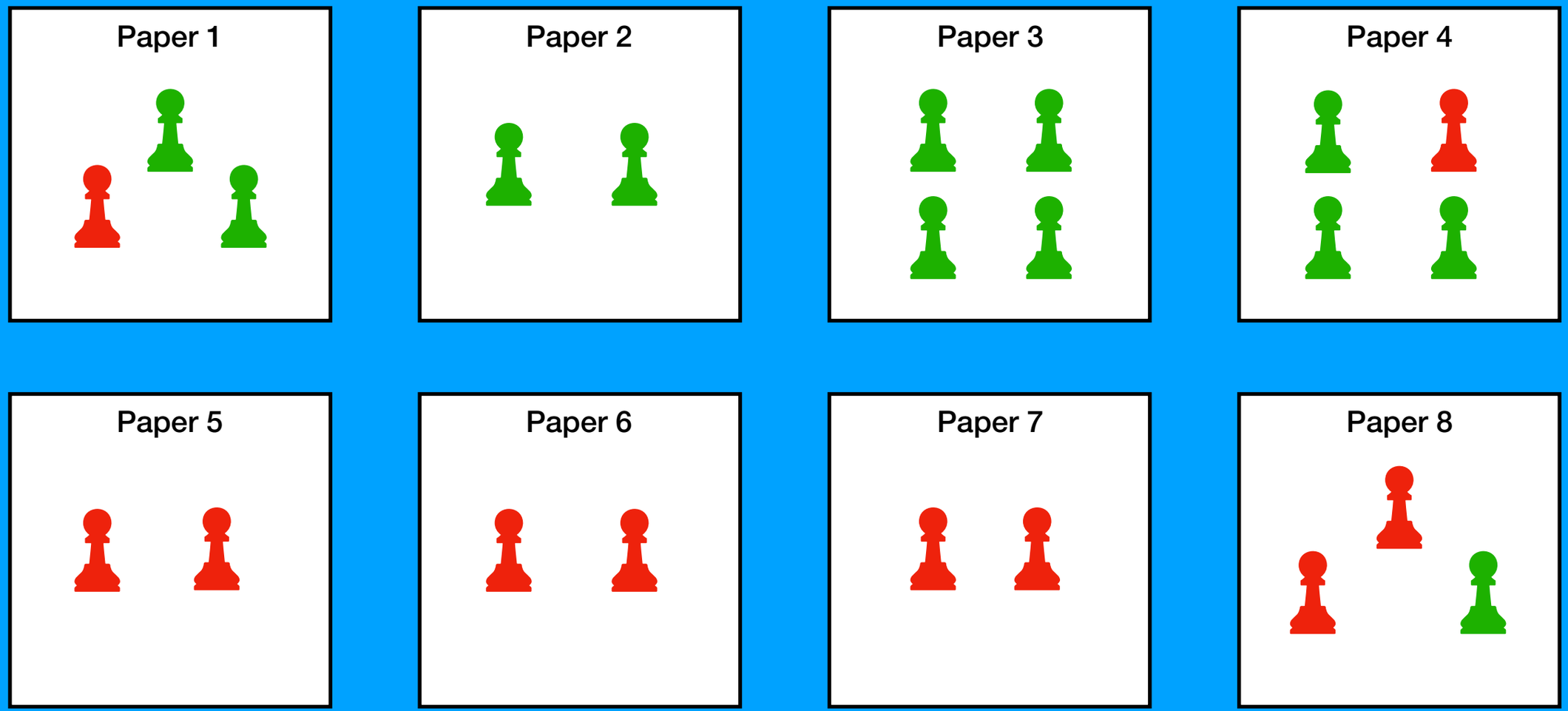


1. Tendency for same-gender authorships to co-author



female
male

1. Tendency for same-gender authorships to co-author



P(co-author = male | author = male) = 3/4

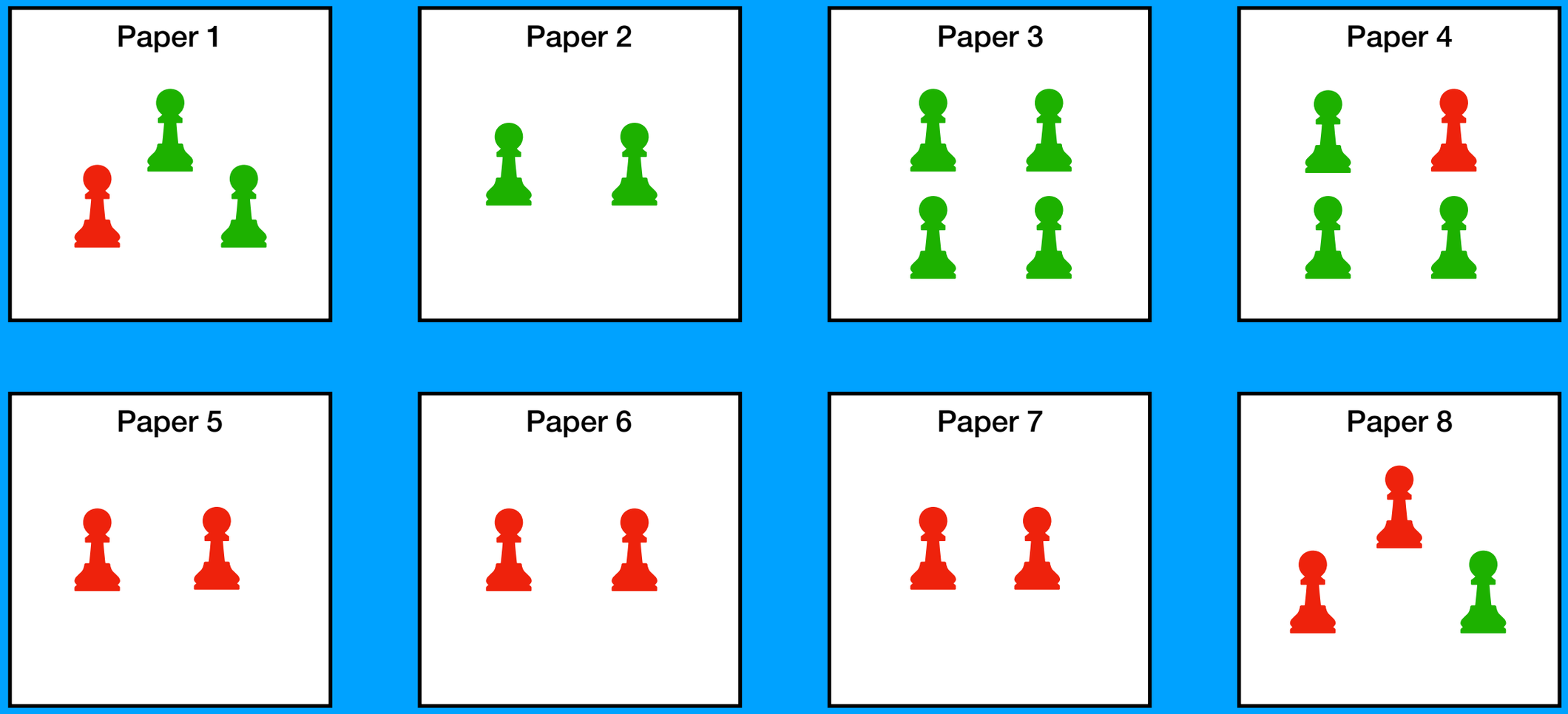
P(co-author = male | author = female) = 3/10

α = 0.45

Question

1. How often do males co-author w/ males (and females w/ females)?
2. Counterfactual: Is this more often than we would otherwise expect?
    1. To do this, we must account for:
        1. Structural Homophily
        2. Compositional Homophily

2. Is this more often than we would otherwise expect?

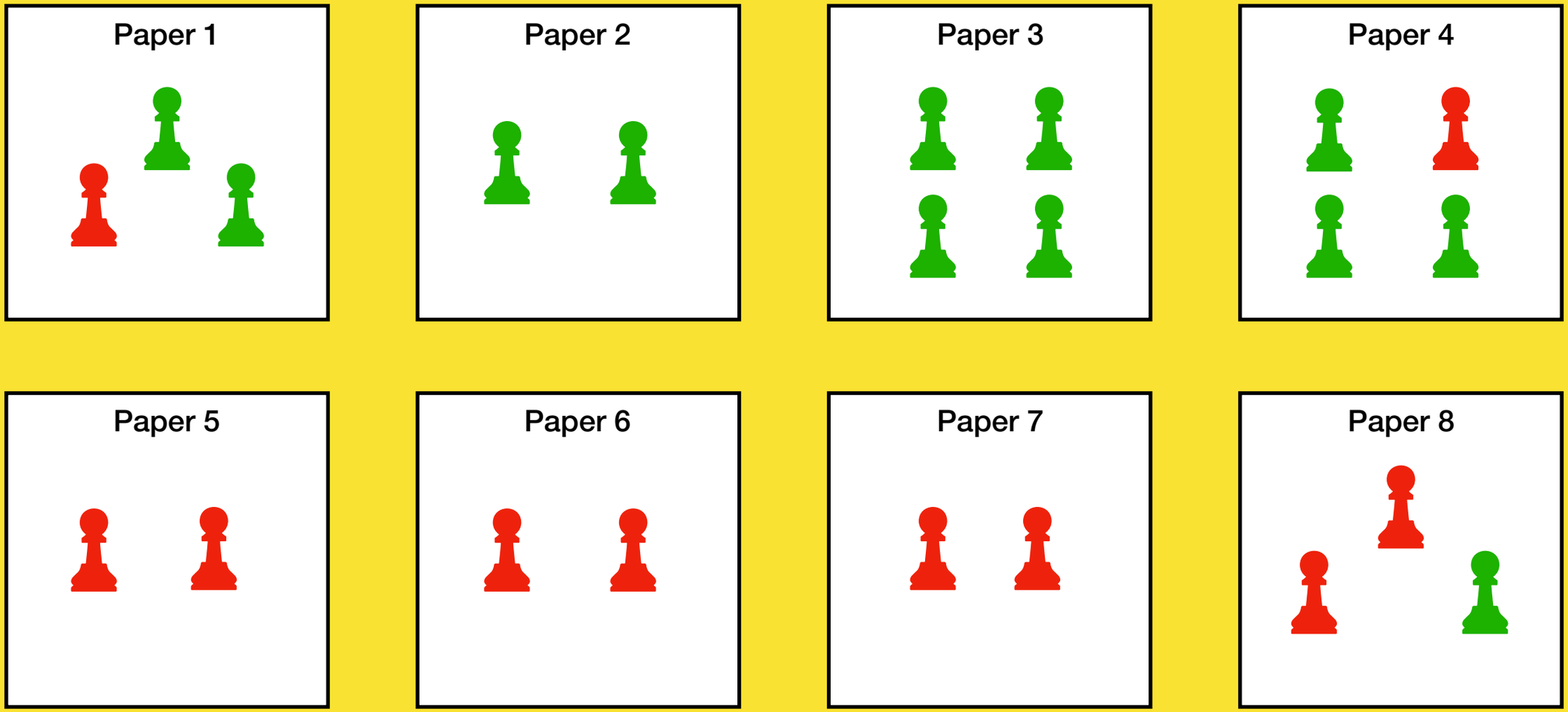


P(co-author = male | author = male) = 3/4

P(co-author = male | author = female) = 3/10

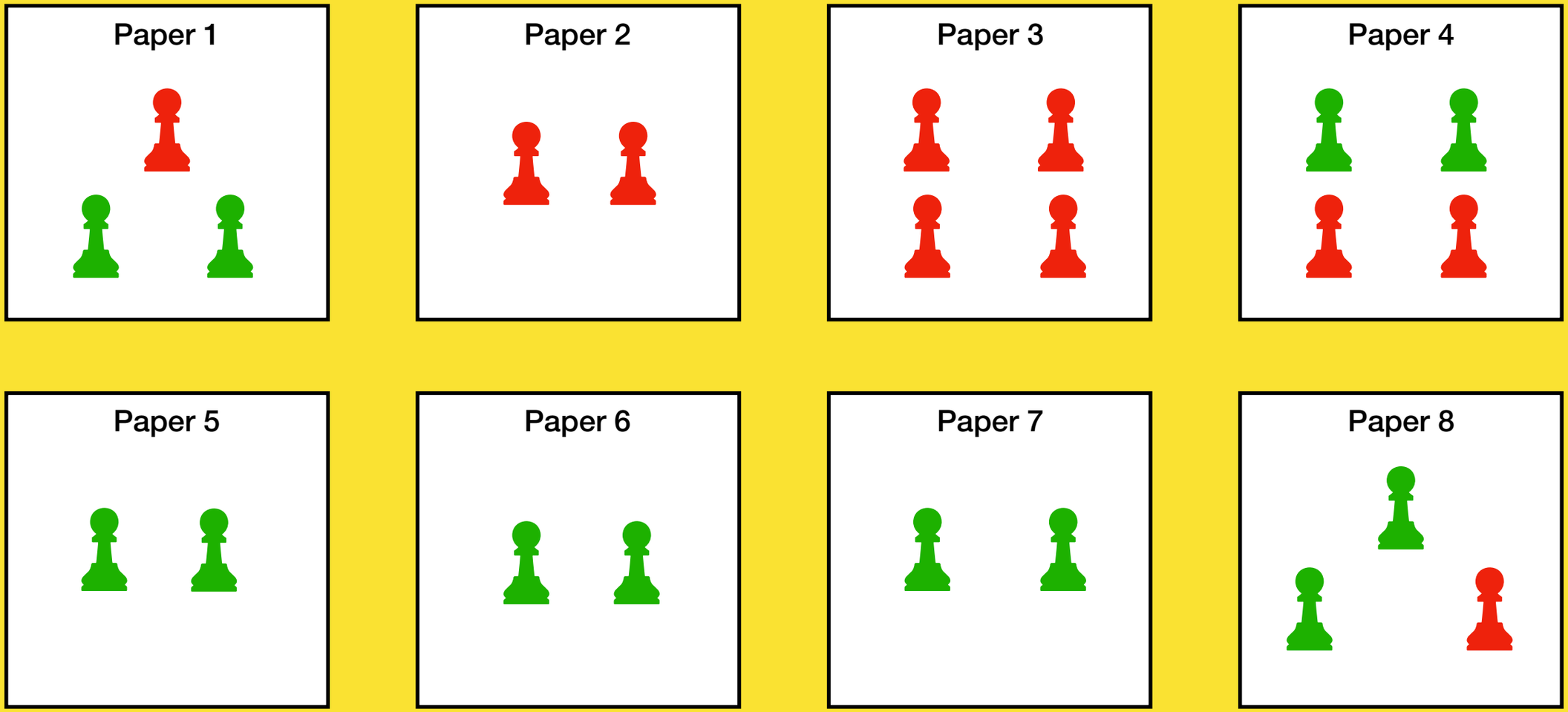
α = 0.45

2. Is this more often than we would otherwise expect?



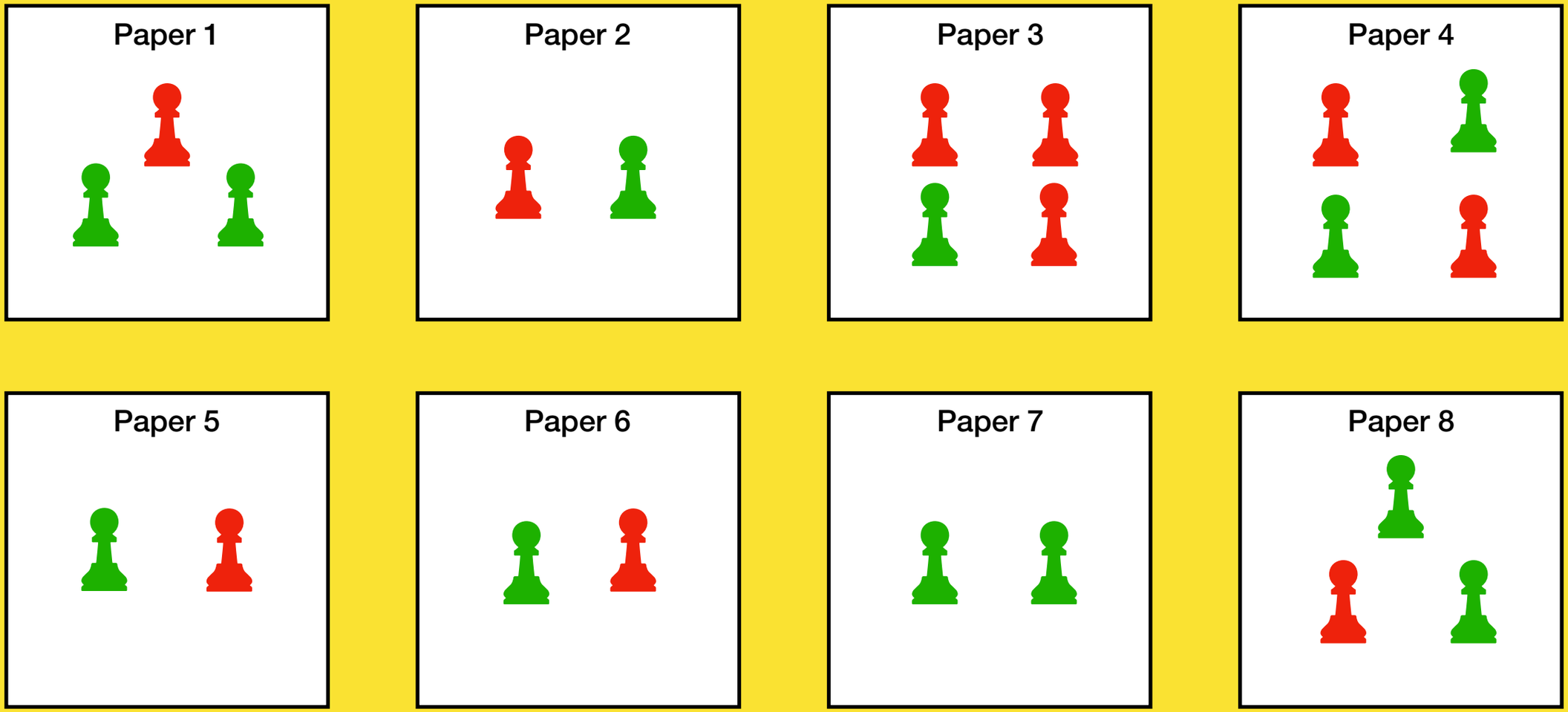
Fix gender ratio, number of authorships, number of authorships per paper.

2. Is this more often than we would otherwise expect?



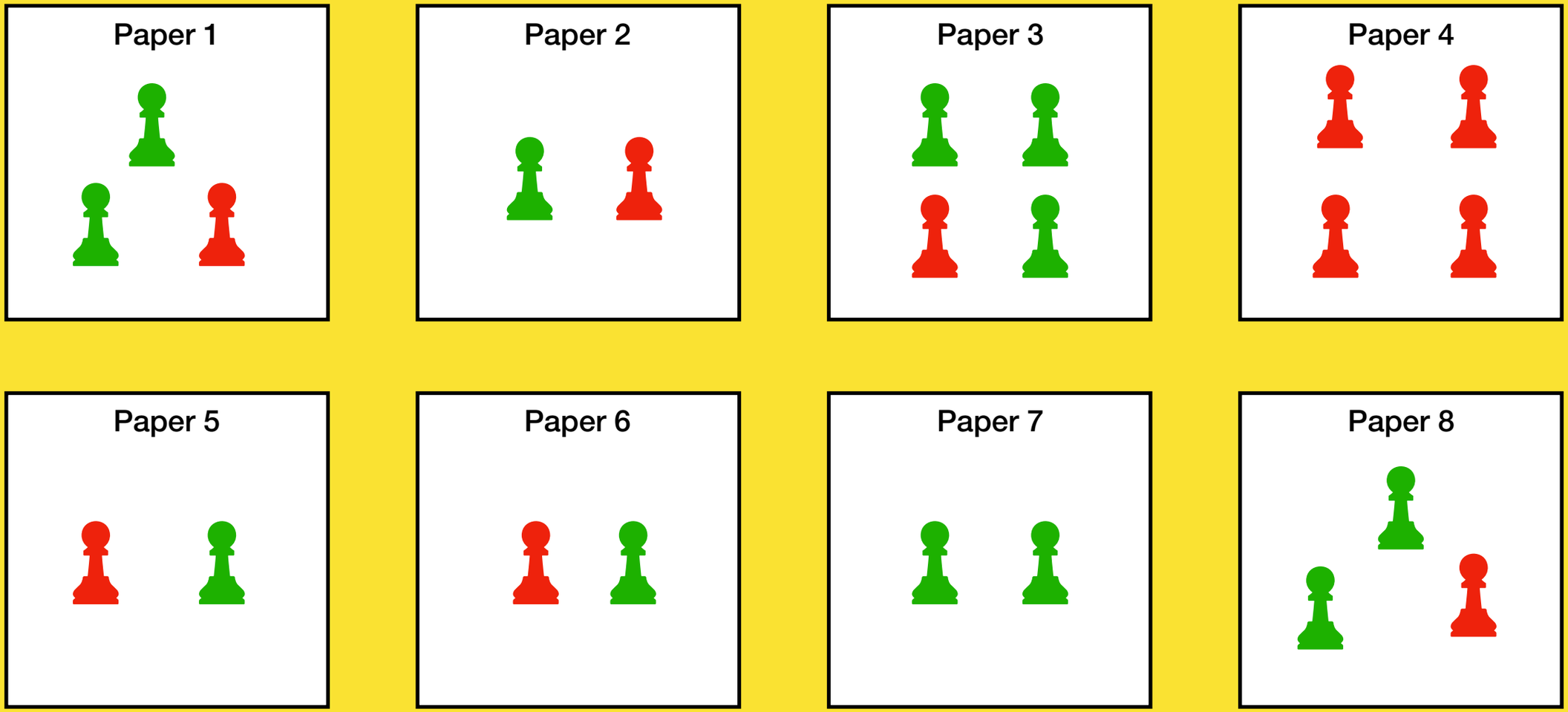
Randomly reassign authorships and measure counterfactual alpha value.

2. Is this more often than we would otherwise expect?



Randomly reassign authorships and measure counterfactual alpha value.

2. Is this more often than we would otherwise expect?

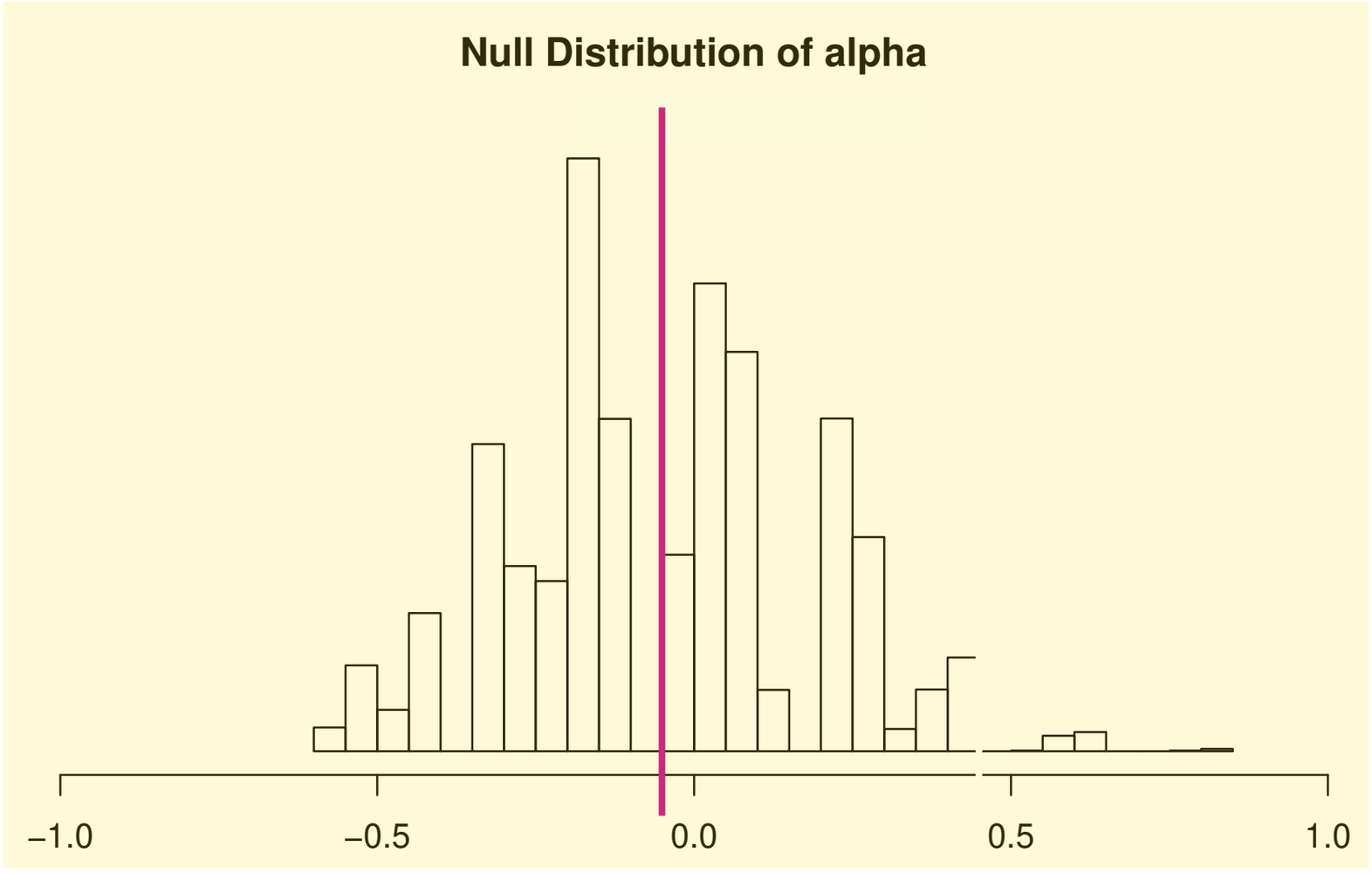


Randomly reassign authorships and measure counterfactual alpha value.

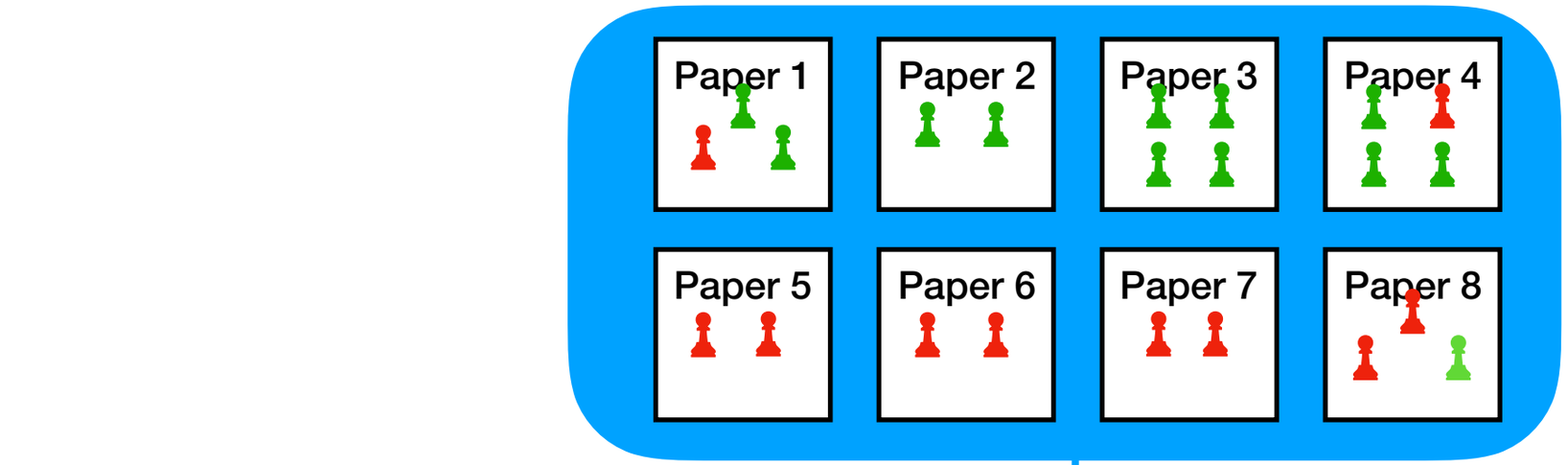
2. Is this more often than we would otherwise expect?

Structural homophily =

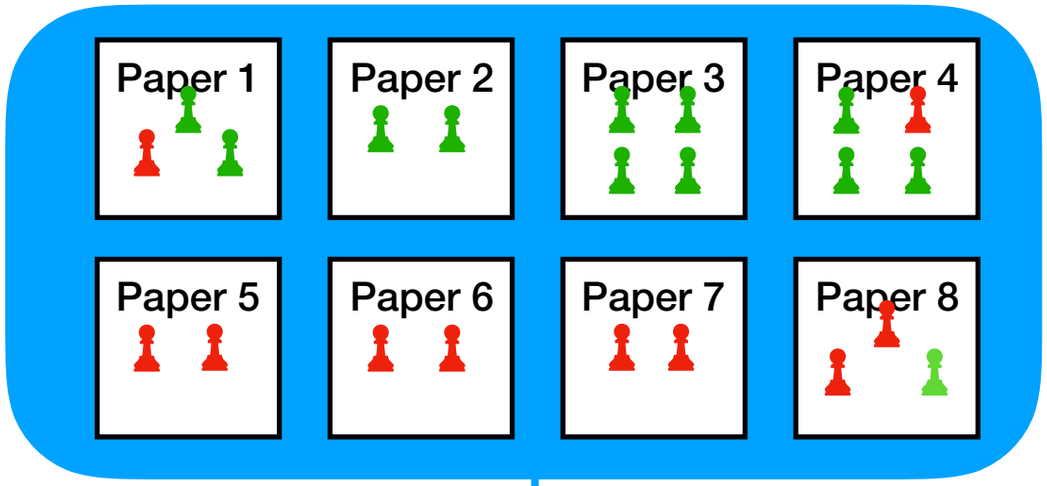
Deviation of alpha from zero due to structural aspects (gender ratio, number of authorships, number of authorships per paper)



2. Is this more often than we would otherwise expect?

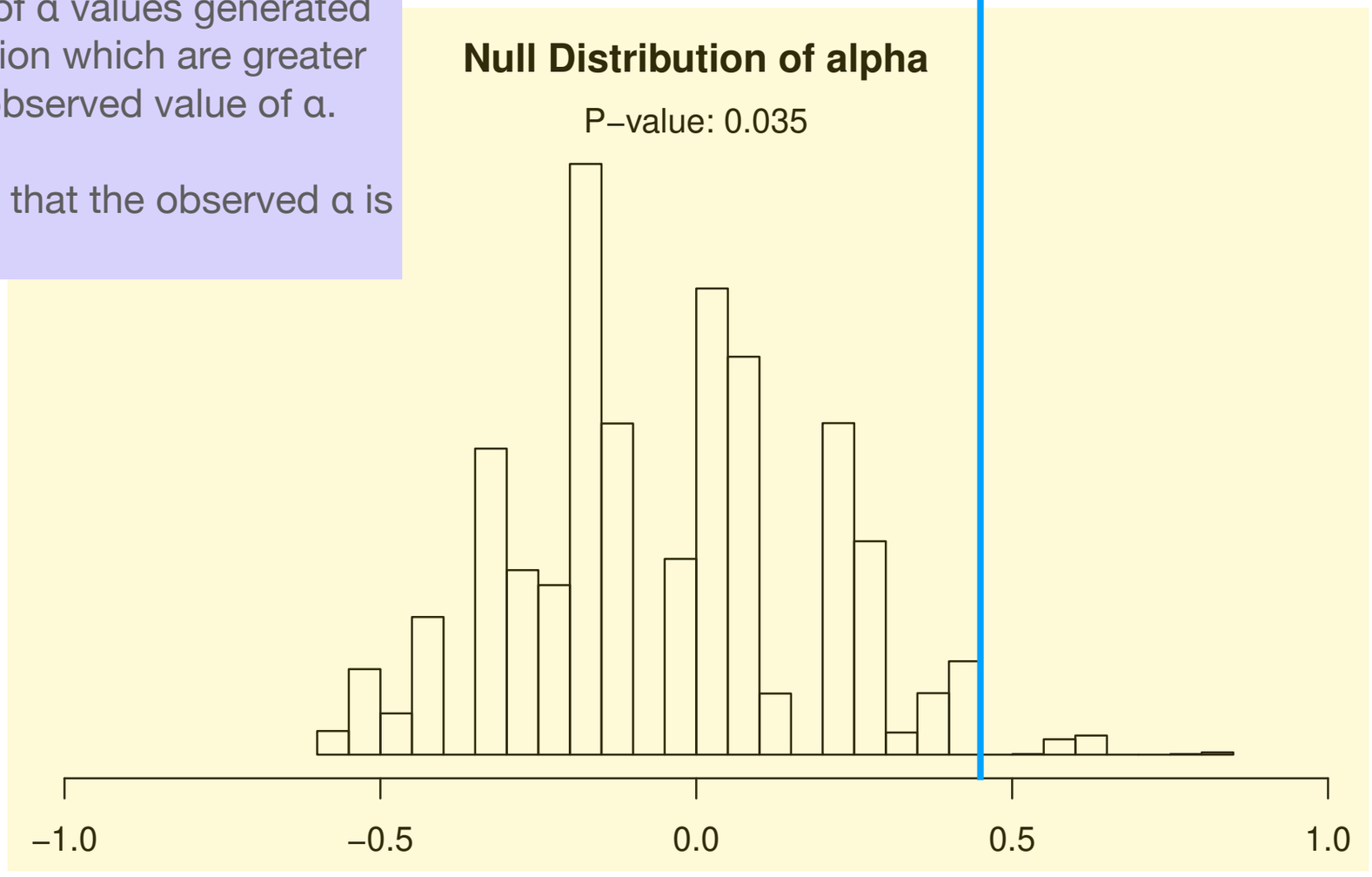


2. Is this more often than we would otherwise expect?

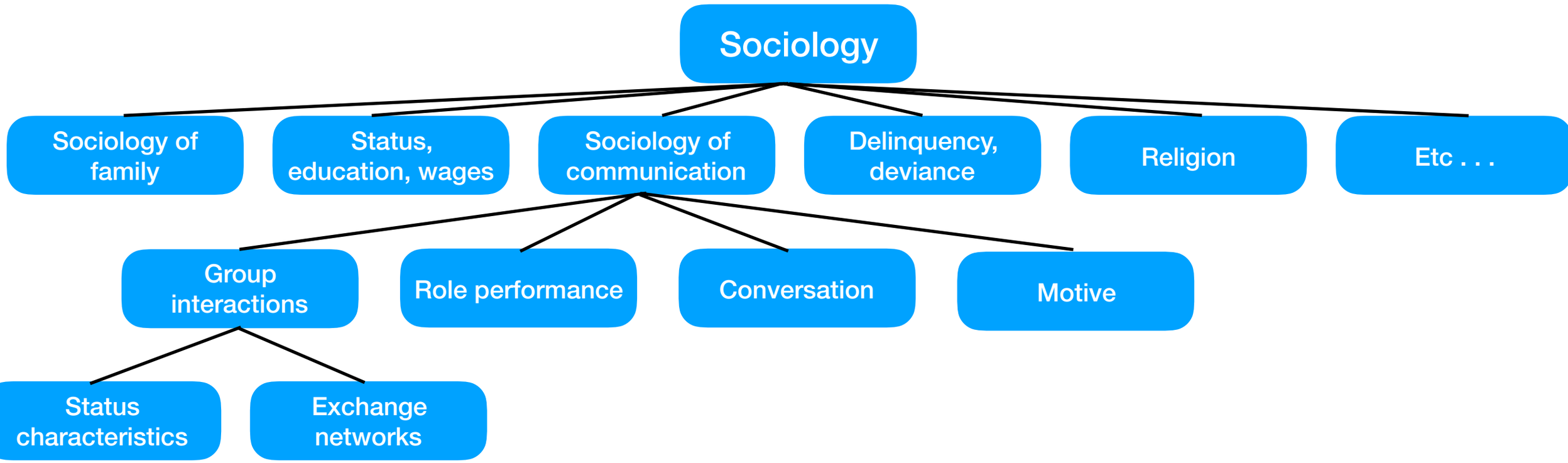


p value = proportion of α values generated from the null distribution which are greater than or equal to the observed value of α.

Small p-value implies that the observed α is unlikely to occur.



2. Is this more often than we would otherwise expect?

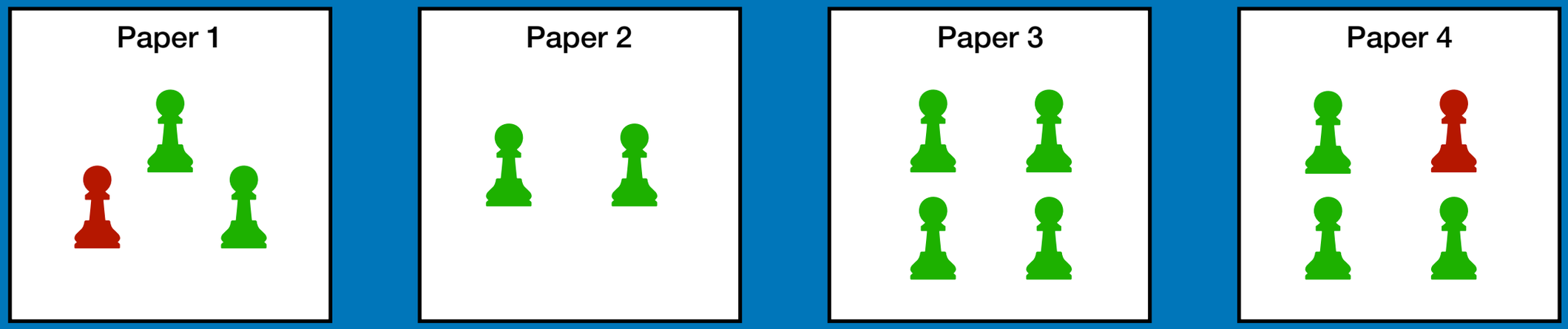


2. Is this more often than we would otherwise expect?

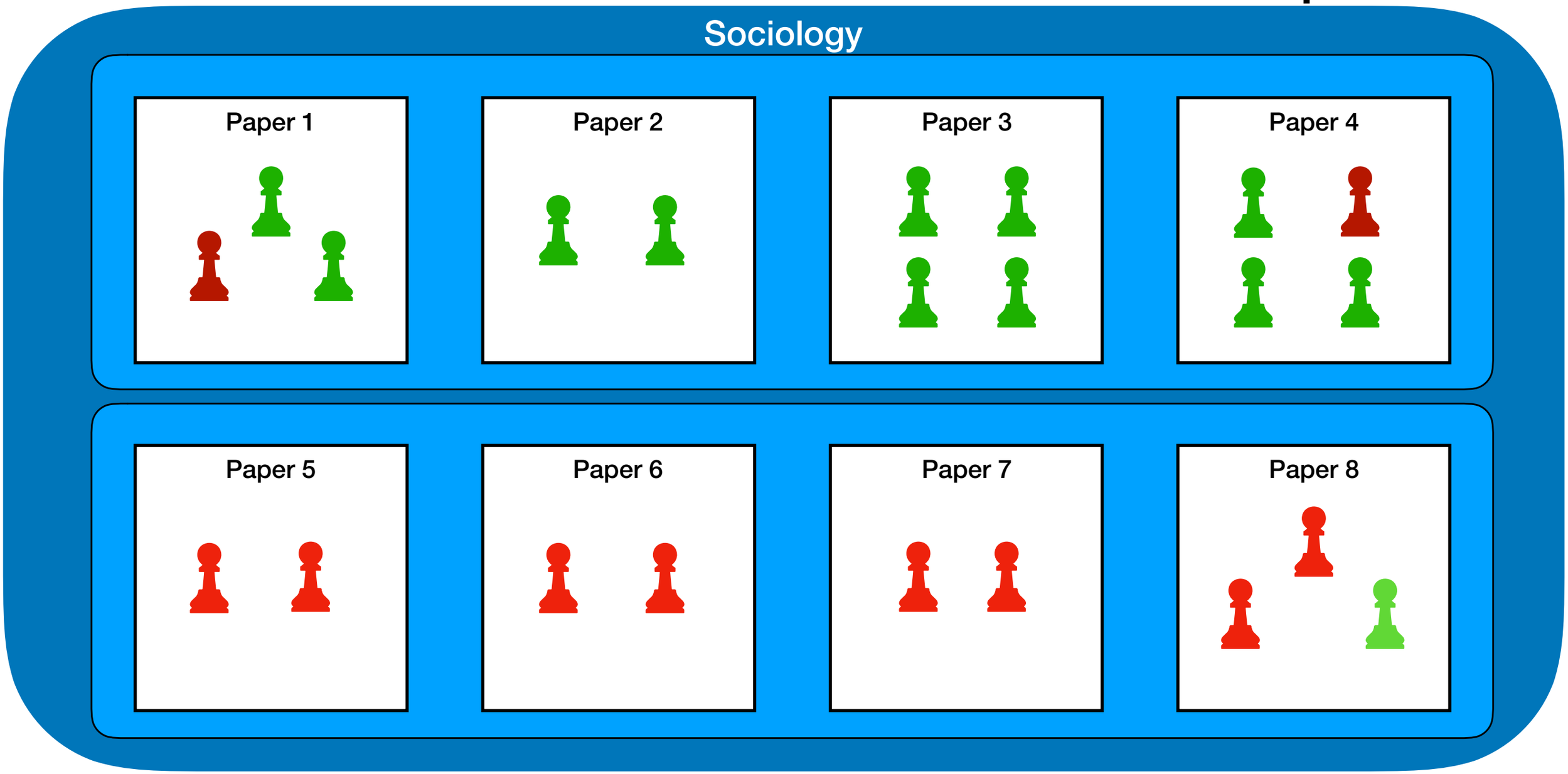
Sociology

2. Is this more often than we would otherwise expect?

Sociology

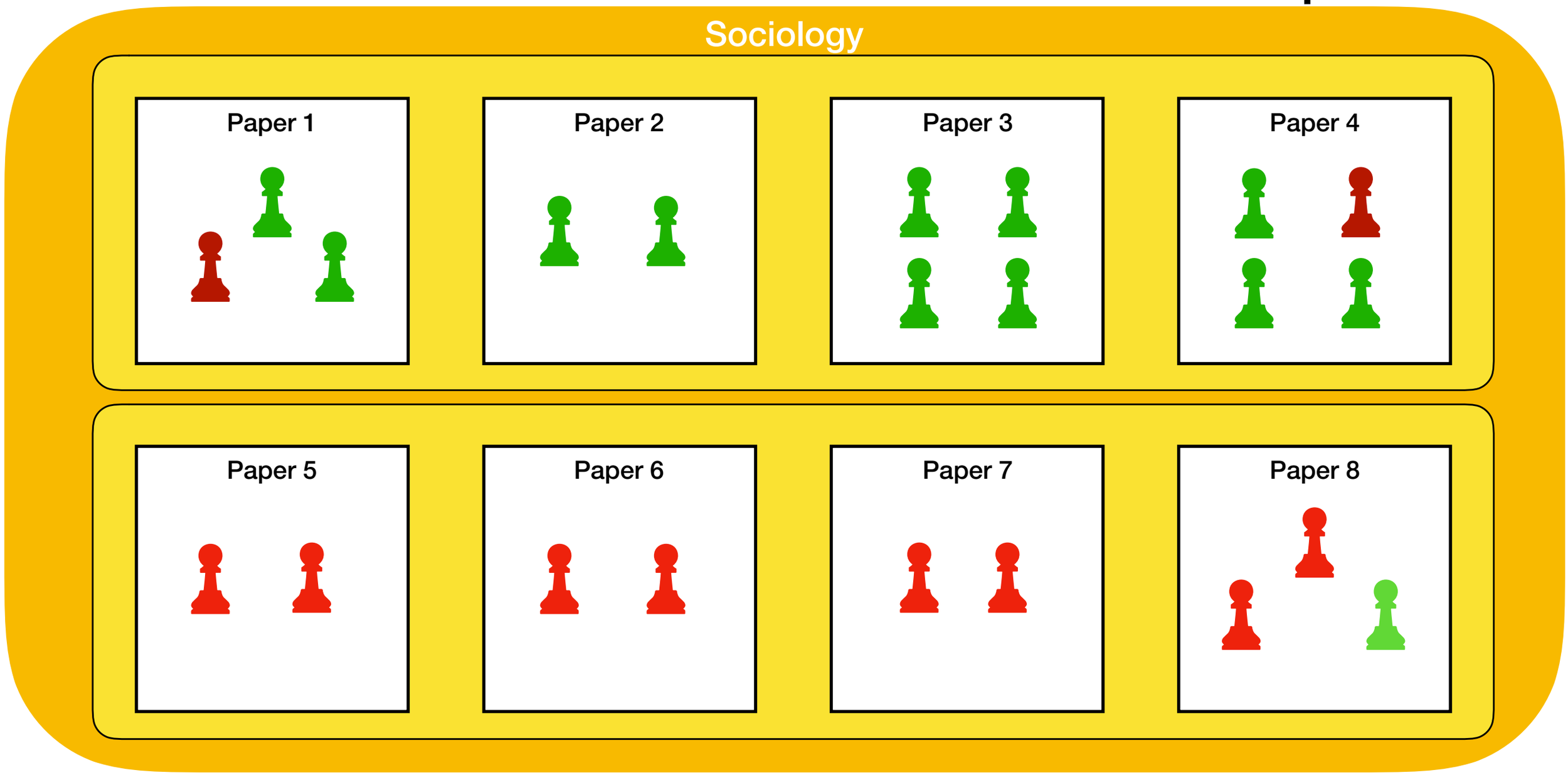


2. Is this more often than we would otherwise expect?



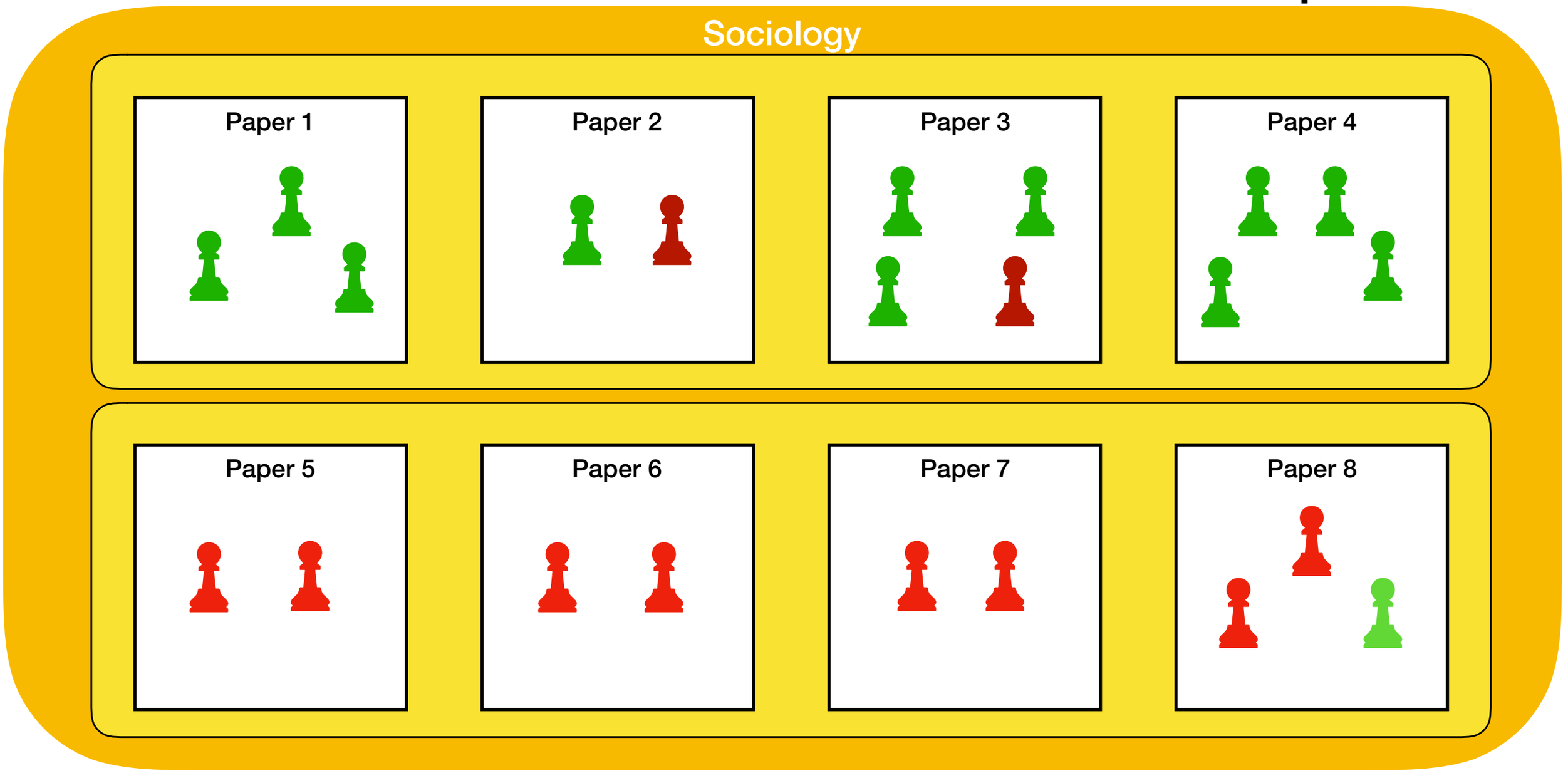
Each terminal field has different gender ratios, numbers of authorships, and numbers of authorships per paper.

2. Is this more often than we would otherwise expect?



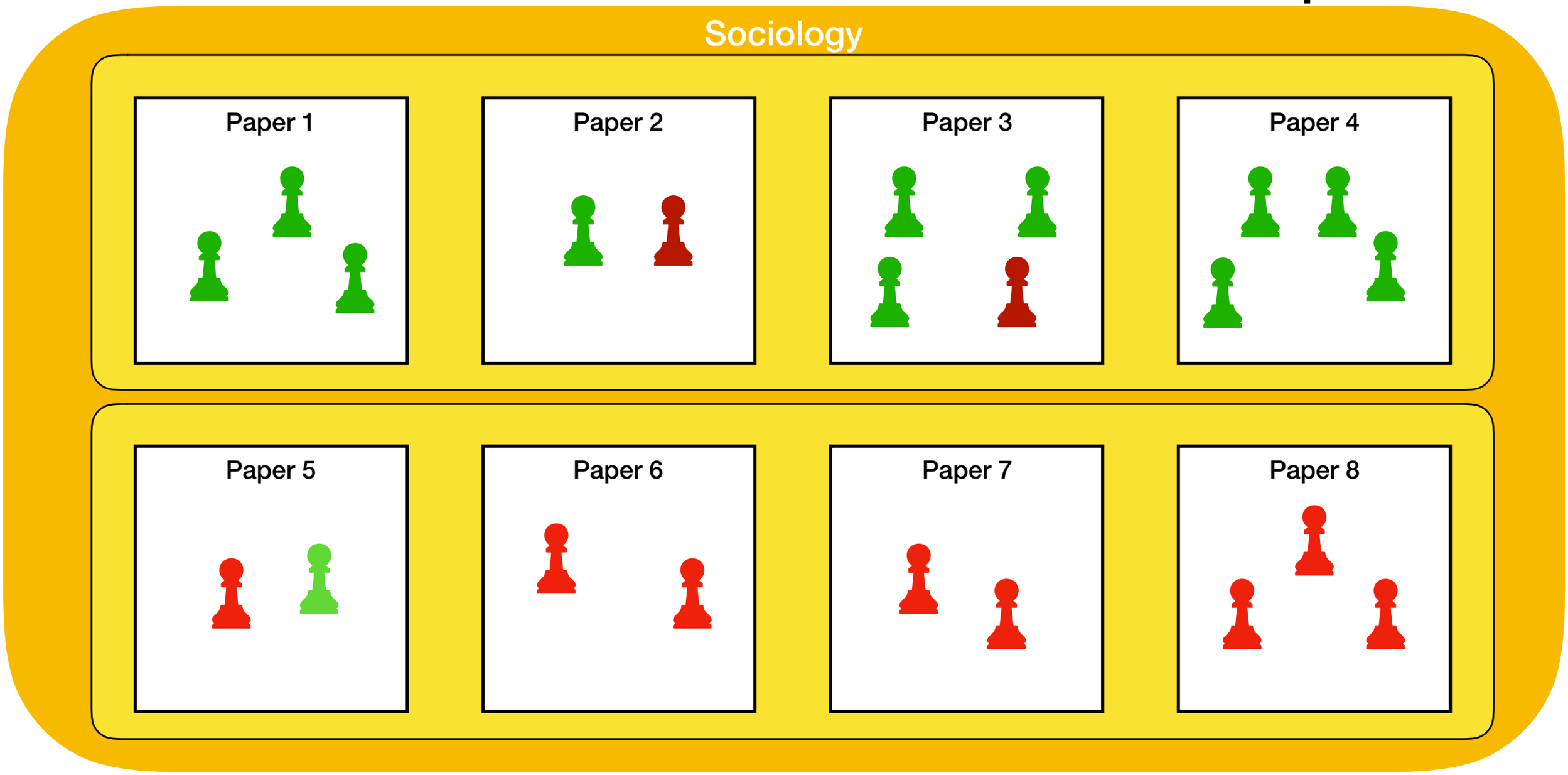
Fix gender ratio, number of authorships, number of authorships per paper *within each terminal field.*

2. Is this more often than we would otherwise expect?



Randomly reassign authorships *within* terminal fields.

2. Is this more often than we would otherwise expect?

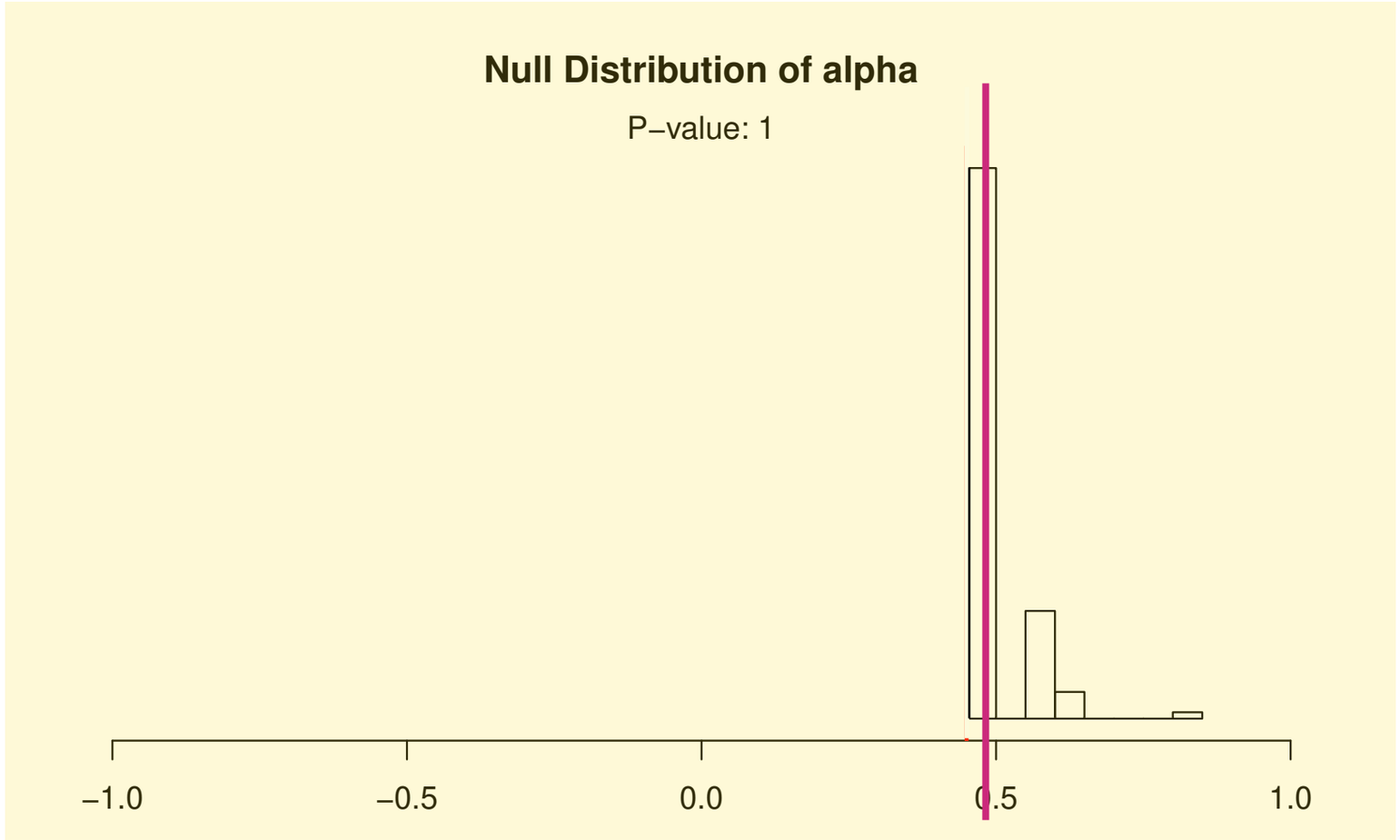


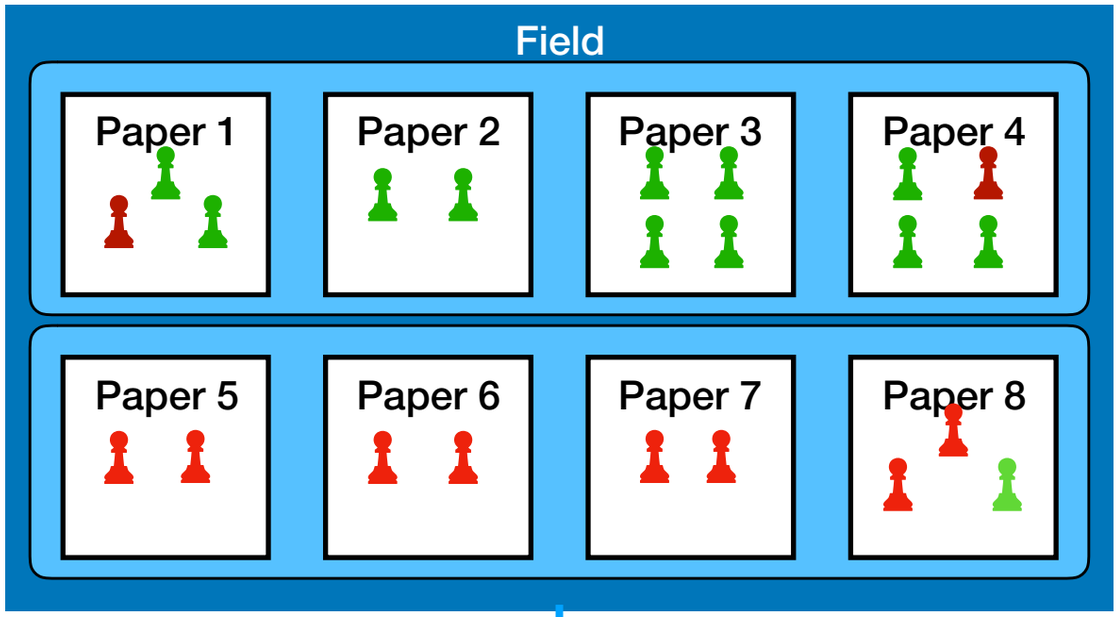
Randomly reassign authorships *within* terminal fields.

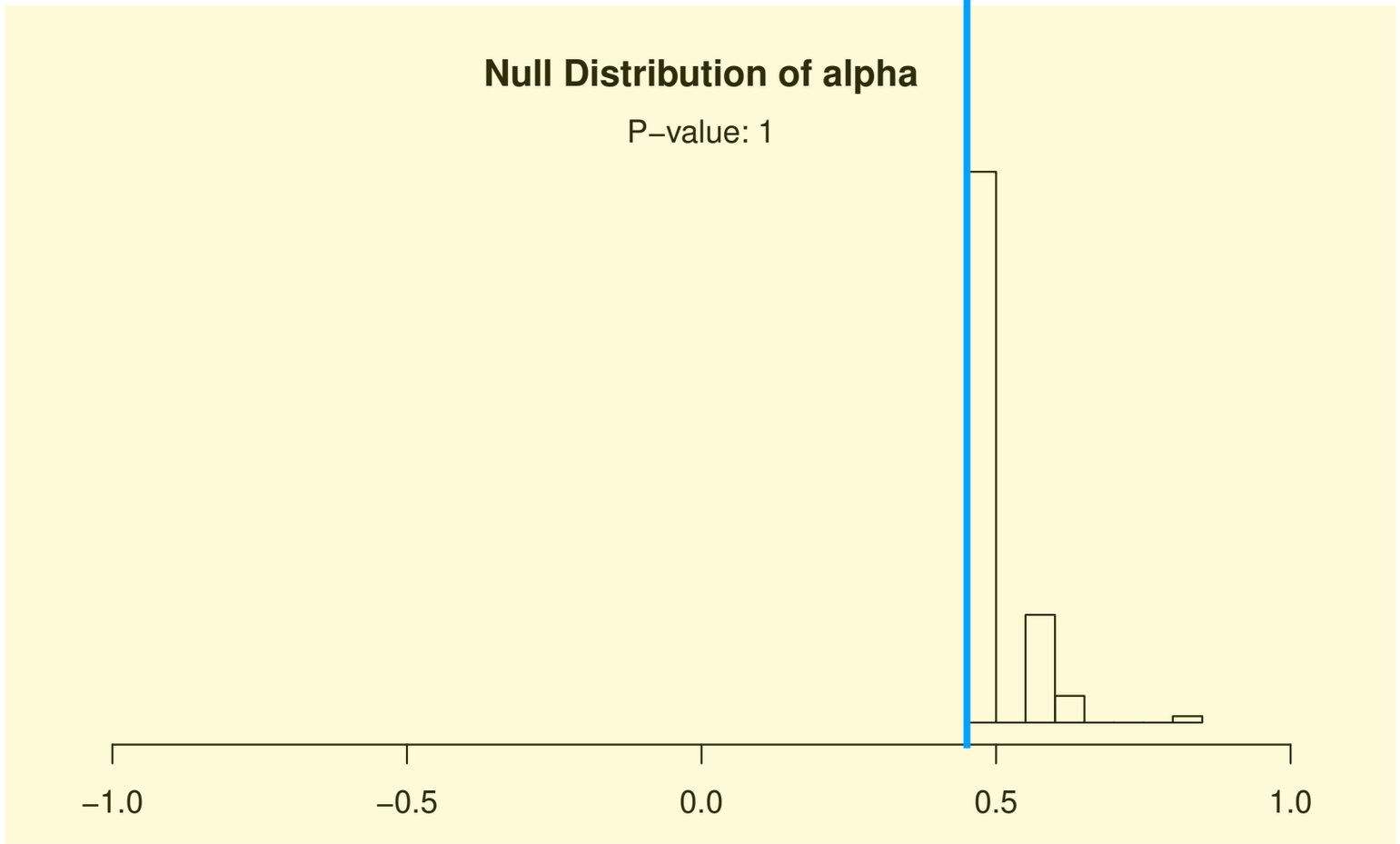
2. Is this more often than we would otherwise expect?

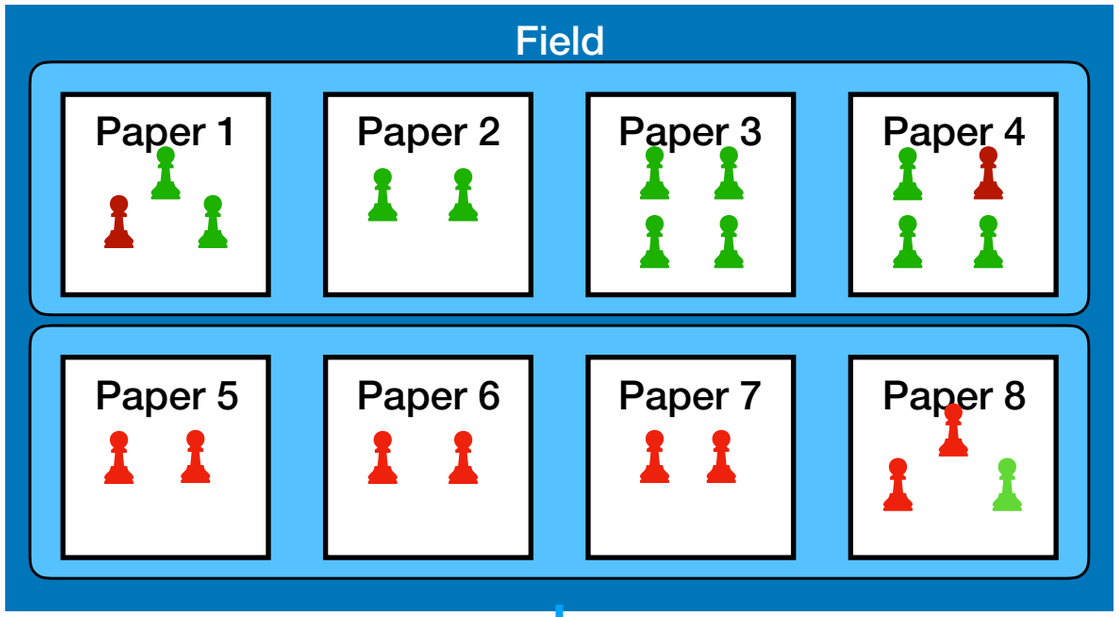
Compositional homophily =

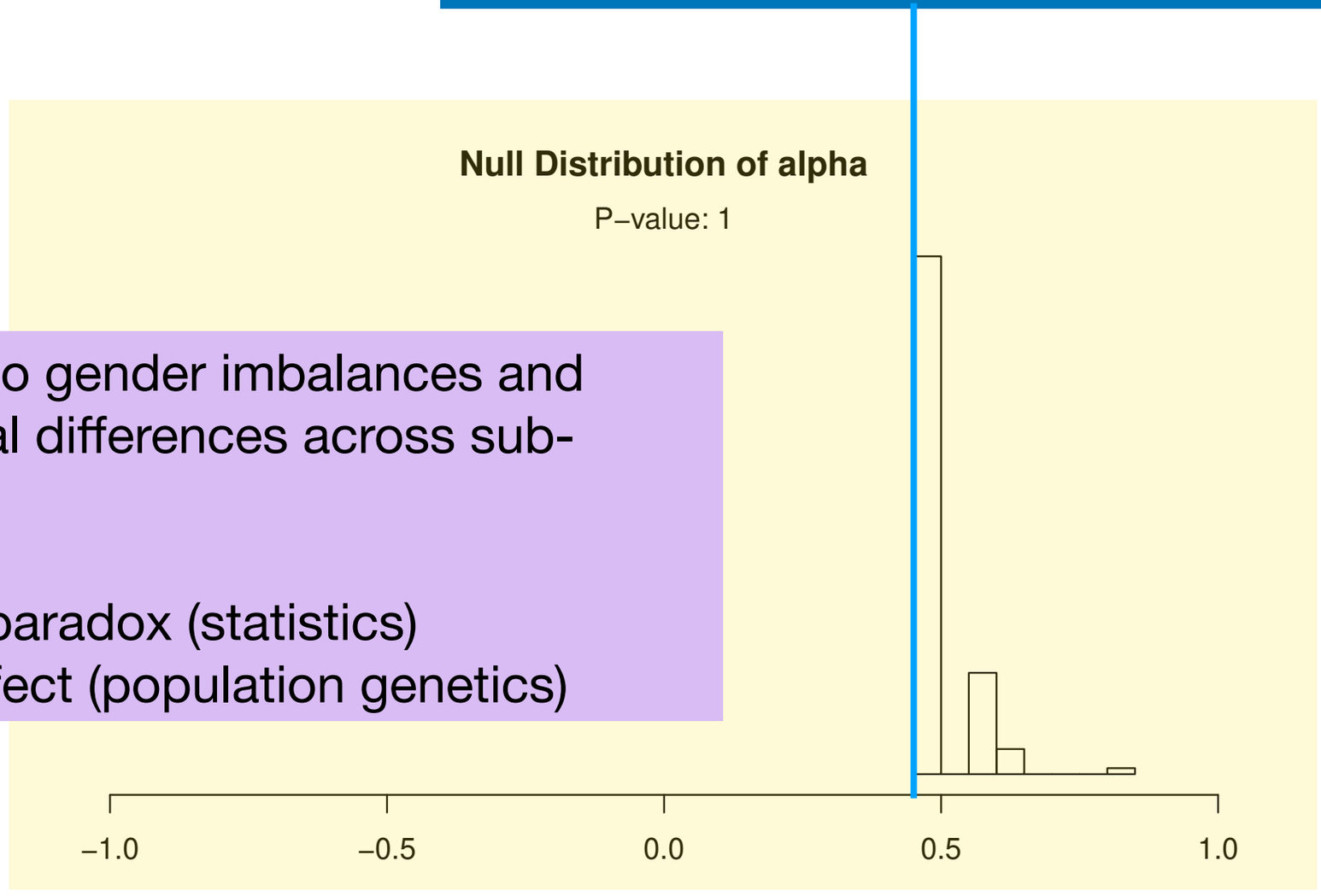
Deviation of alpha from the expected alpha under structural homophily that arises due to the tendency for authorships to co-author with those who are intellectually closer.




Field
Paper 1
Paper 2
Paper 3
Paper 4
Paper 5
Paper 6
Paper 7
Paper 8


Null Distribution of alpha
P−value: 1
−1.0
−0.5
0.0
0.5
1.0





Reversal due to gender imbalances and other structural differences across sub-populations:

- Simpson’s paradox (statistics)
- Wahlund effect (population genetics)

Testing procedure

- Generate counterfactuals:
    - Assume authorships in the same terminal field are exchangeable
    - Assume authorships in different terminal fields are less likely to swap and use citation flows to determine swap probabilities
    - Use a Markov Chain Monte Carlo Metropolis Hastings sampler to generate counterfactuals
    - 75,000 draws form null distribution after burn-in
- Compare counterfactual alpha values to observed alpha:
    - Hypothesis test for all levels (terminal fields, composite fields, major fields)
    - Use p-values adjusted by the Benjamini-Yuketieli procedure to control False Discovery Rate at 0.05

Results: Behavioral homophily

- 84% of top fields

Observed

Avg counterfactual

Field	α	P-value
JSTOR	.11/.05	.00
Mol/Cell Bio	.05/.01	.00
Eco/Evol	.06/.02	.00
Economics	.11/.02	.00
Sociology	.19/.07	.00
Prob/Stat	.09/.03	.00
Org/mkt	.16/.04	.00
Education	.16/.04	.00
Occ Health	.10/.02	.00
Anthro	.12/.03	.00
Law	.17/.08	.00
History	.16/.07	.00
Phys Anthro	.07/.01	.00
Intl Poli Sci	.09/.02	.03
US Poli Sci	.15/.07	.00
Philosophy	.10/.03	.03
Math	.04/.01	1.00
Vet Med	.09/.01	.00
Cog Sci	.18/.09	.00
Radiation	.09/.01	.00
Demography	.15/.06	.00
Classics	.07/.01	.27
Opr Res	.03/.00	.73
Plant Phys	.08/.02	.03
Mycology	.03/.01	1.00

Results: Behavioral homophily

- 29% of composite fields
- 8% of terminal fields

- Trade-off between increasing testing power by aggregating data versus controlling for confounders by analyzing at a fine-grain level.

Field	Significant / Total	
	Term	Comp
JSTOR	124/1450	82/280
Mol/Cell Bio	29/178	19/44
Eco/Evol	17/257	15/56
Economics	9/136	11/28
Sociology	13/94	12/21
Prob/Stat	1/90	2/23
Org/mkt	8/68	3/4
Education	12/42	6/10
Occ Health	12/24	1/1
Anthro	5/63	2/8
Law	0/98	1/16
History	0/49	1/6
Phys Anthro	1/32	2/10
Intl Poli Sci	0/34	0/2
US Poli Sci	2/37	1/6
Philosophy	0/45	0/8
Math	0/46	0/9
Vet Med	7/19	1/2
Cog Sci	4/14	3/3
Radiation	3/14	1/5
Demography	0/20	1/2
Classics	0/35	0/8
Opr Res	0/18	0/4
Plant Phys	1/21	0/3
Mycology	0/16	0/1

Visualization

http://eigenfactor.org/projects/gender_homophily/

How sensitive are our results to missing gender indicators?

- To evaluate, we impute gender for authorships under two scenarios:
    - Low homophily: Impute each missing gender indicator at random according to the proportions of assigned genders in the authorship's *original terminal field*. Assumes no behavioral homophily in the imputed data.
    - High homophily: Impute each missing gender indicator at random according to the proportions of assigned gender in the *original paper*. If a paper contains only unassigned authorships, impute a single gender for all according to the proportions of assigned gender authorships for the terminal field. Assumes behavioral homophily since, by construction, papers with one or no assigned authorships are always gender homophilous.

How sensitive are our results to missing gender indicators?

- For each scenario, carry out 10 imputations and then repeat the entire sampling and testing procedures used for the main analysis.

- Original results:

- Top fields: 84%
- Composite fields: 28%
- Terminal fields: 8%

	Terminal	Composite	Top
Main Analysis	0.09	0.29	0.83
Low Imputation 1	0.06	0.23	0.83
Low Imputation 2	0.07	0.25	0.75
Low Imputation 3	0.06	0.25	0.75
Low Imputation 4	0.08	0.25	0.75
Low Imputation 5	0.06	0.26	0.75
Low Imputation 6	0.07	0.28	0.75
Low Imputation 7	0.07	0.25	0.83
Low Imputation 8	0.07	0.26	0.79
Low Imputation 9	0.07	0.25	0.75
Low Imputation 10	0.06	0.26	0.79
Low Imputation Avg	0.07	0.25	0.78
High Imputation 1	0.54	0.82	1.00
High Imputation 2	0.53	0.82	1.00
High Imputation 3	0.53	0.82	1.00
High Imputation 4	0.53	0.82	1.00
High Imputation 5	0.54	0.82	1.00
High Imputation 6	0.53	0.81	1.00
High Imputation 7	0.54	0.82	1.00
High Imputation 8	0.54	0.83	1.00
High Imputation 9	0.54	0.83	1.00
High Imputation 10	0.54	0.82	1.00
High Imputation Avg	0.54	0.82	1.00

Secondary analysis

- Fit logistic regression for all terminal fields where the outcome is whether there is statistically significant behavioral homophily

- Result: behavioral homophily has a statistically significant positive association with the proportion of females and terminal field size

- In economics, behavioral homophily was also found to be more prevalent in sub-fields with a higher proportion of females

- Homophily: as representation of women increases, more likely that same-gender individuals who are sufficiently compatible along key dimensions become available as co-authors

- Caveat: Larger field size and balanced gender representation increase power of testing procedure (i.e., ability to detect behavioral gender homophily).

Boschini, Anne, and Anna Sjögren. "Is team formation gender neutral? Evidence from coauthorship patterns." *Journal of Labor Economics* 25, no. 2 (2007): 325-365.

Conclusion

- We detect behavioral gender homophily in 84% of top fields, 29% of composite fields, and 8% of terminal fields

- Since behavioral gender homophily is endemic to even some of the smallest intellectual communities, it might only be mitigated by changing the cultural norms and perceptions that drive behavioral gender homophily within intellectual communities

Future work: Strategic value of gender homophily?

- Short-term: does gender homophily increase retention, productivity, and impact of female authors?

    - Stereotype threat: presence of other women in male-stereotyped domains enhances confidence, performance, and motivation

- Long-term: do gender homophilous co-authorships lead to gender-homophilous intellectual communities? If so, does increasing the ratio of women in an intellectual community decrease its value/impact, just as increasing the ratio of women in an occupation decreases its prestige?

Murphy, Mary C., Claude M. Steele, and James J. Gross. "Signaling threat: How situational cues affect women in math, science, and engineering settings." *Psychological science* 18, no. 10 (2007): 879-885.

Stout, Jane G., Nilanjana Dasgupta, Matthew Hunsinger, and Melissa A. McManus. "STEMing the tide: using ingroup experts to inoculate women's self-concept in science, technology, engineering, and mathematics (STEM)." *Journal of personality and social psychology* 100, no. 2 (2011): 255.

Goldin, Claudia. "A pollution theory of discrimination: male and female differences in occupations and earnings." In *Human capital in history: The American record*, pp. 313-348. University of Chicago Press, 2014.

Future work: Fine-tuning methods

- Temporal aspects: Gender representation changes over time. Incorporate temporal info into the null distribution.

- Disambiguating authors: In terminal fields with few female authors, we may overestimate structural and compositional homophily (and underestimate behavioral homophily) by allowing multiple female authorships corresponding to the same author to be reassigned to the same paper.

Thank you

- For early discussions:
    - Jennifer Jacquet, Molly King, Shelley Correll, & Ted Bergstrom
- Funders:
    - Royalty Research Fund Grant #A118374, Elena Erosheva (PI) and Carole Lee (co-PI)
    - NSF Grant #1735194, Jevin West (co-PI)

For more information

- Link to paper: http://arxiv.org/abs/1909.01284
- Visualization: http://eigenfactor.org/projects/gender_homophily/
- Code for analysis and plots: https://github.com/ysamwang/genderHomophily
- Data: Under license by JSTOR to the authors. Requests for raw data should be made to JSTOR directly.

Size of each top level field

Label	Authors (Count)	Papers (Count)	Terminal Fields	Composite Fields
Anthropology	37588	30499	63	8
Classical studies	10596	9061	37	8
Cognitive science	15715	5553	14	3
Demography	9653	5509	20	2
Ecology and evolution	264853	116327	257	56
Economics	95934	59096	136	28
Education	40188	23065	42	10
History	26449	24043	49	6
Law	23974	19779	105	16
Mathematics	18348	14125	46	9
Molecular & Cell biology	382971	92528	178	44
Mycology	7469	3679	16	1
Operations research	13716	7780	18	4
Organizational and marketing	34254	17963	68	4
Philosophy	21738	19126	46	8
Physical anthropology	29693	16703	32	10
Plant physiology	9159	5436	21	3
Political science - international	15283	11835	34	2
Political science-US domestic	12581	7824	37	6
Pollution and occupational health	50967	12359	24	1
Probability and Statistics	37471	22094	90	23
Radiation damage	14118	4215	14	5
Sociology	57146	31662	94	21
Veterinary medicine	17756	4796	19	2

Structural characteristics of each field

	Prop Single Author		Single-Author			Multi-Author			
Label	Papers	Auth	% F	% M	% U	% F	% M	% U	ICC
Anthropology	0.86	0.70	0.27	0.63	0.10	0.28	0.61	0.12	0.10
Classical studies	0.93	0.79	0.22	0.70	0.08	0.27	0.65	0.08	0.05
Cognitive science	0.25	0.09	0.29	0.64	0.07	0.28	0.62	0.10	0.09
Demography	0.57	0.32	0.24	0.61	0.15	0.30	0.53	0.16	0.11
Ecology and evolution	0.37	0.16	0.14	0.79	0.08	0.20	0.70	0.11	0.10
Economics	0.55	0.34	0.08	0.81	0.11	0.11	0.77	0.12	0.10
Education	0.55	0.31	0.35	0.58	0.07	0.41	0.50	0.08	0.07
History	0.92	0.84	0.24	0.70	0.06	0.23	0.69	0.08	0.05
Law	0.85	0.70	0.17	0.78	0.06	0.22	0.71	0.06	0.05
Mathematics	0.75	0.58	0.06	0.76	0.18	0.06	0.73	0.21	0.19
Molecular & Cell biology	0.14	0.03	0.19	0.70	0.10	0.23	0.61	0.16	0.09
Mycology	0.45	0.22	0.20	0.71	0.09	0.22	0.65	0.13	0.08
Operations research	0.48	0.27	0.05	0.81	0.14	0.08	0.72	0.20	0.21
Organizational and marketing	0.40	0.21	0.18	0.72	0.10	0.19	0.68	0.12	0.12
Philosophy	0.89	0.78	0.09	0.82	0.08	0.10	0.78	0.11	0.12
Physical anthropology	0.66	0.37	0.22	0.72	0.06	0.22	0.68	0.09	0.07
Plant physiology	0.53	0.31	0.13	0.79	0.08	0.17	0.72	0.10	0.07
Political science - international	0.78	0.60	0.16	0.74	0.10	0.17	0.73	0.10	0.08
Political science-US domestic	0.57	0.35	0.17	0.76	0.07	0.18	0.75	0.07	0.05
Pollution and occupational health	0.22	0.05	0.24	0.65	0.11	0.31	0.53	0.17	0.19
Probability and Statistics	0.54	0.32	0.08	0.75	0.17	0.13	0.69	0.19	0.20
Radiation damage	0.23	0.07	0.22	0.66	0.11	0.22	0.62	0.16	0.14
Sociology	0.52	0.29	0.30	0.63	0.08	0.38	0.53	0.08	0.07
Veterinary medicine	0.26	0.07	0.27	0.60	0.12	0.25	0.59	0.15	0.22

Readjust the Benjamini-Yekutieli procedure to control the false discovery rate at 0.005 instead of 0.05

	BY; Rate $= .005$		
Field	P-value	Term	Comp
JSTOR	.00	68/1450	65/280
Mol/Cell Bio	.00	16/178	18/44
Eco/Evol	.00	7/257	13/56
Economics	.00	4/136	8/28
Sociology	.00	9/94	9/21
Prob/Stat	.00	0/90	1/23
Org/mkt	.00	5/68	3/4
Education	.00	6/42	4/10
Occ Health	.00	8/24	1/1
Anthro	.00	2/63	1/8
Law	.00	0/98	0/16
History	.00	0/49	0/6
Phys Anthro	.00	1/32	2/10
Intl Poli Sci	.03	0/34	0/2
US Poli Sci	.00	0/37	0/6
Philosophy	.03	0/45	0/8
Math	1.00	0/46	0/9
Vet Med	.00	5/19	1/2
Cog Sci	.00	2/14	3/3
Radiation	.00	3/14	1/5
Demography	.00	0/20	0/2
Classics	.27	0/35	0/8
Opr Res	.73	0/18	0/4
Plant Phys	.03	0/21	0/3
Mycology	1.00	0/16	0/1

Previous work

- Field-level:
    - Economics: study of publications from cohort of 178 PhDs found women $\geq$ 5x more likely than men to have female co-authors
    - A study of 1,045,401 multi-authored papers finds gender homophily at what we'd characterize as the field level
- Sub-field studies:
    - Economics: Study of 3,090 articles in the top three journals (1991-2002) found gender homophily at the sub-field level

McDowell, John M., and Janet Kiholm Smith. "The effect of gender–sorting on propensity to coauthor: Implications for academic promotion." *Economic Inquiry* 30, no. 1 (1992): 68-82.

AlShebli, Bedoor K., Talal Rahwan, and Wei Lee Woon. "The preeminence of ethnic diversity in scientific collaboration." *Nature communications* 9, no. 1 (2018): 5163.

Boschini, Anne, and Anna Sjögren. "Is team formation gender neutral? Evidence from coauthorship patterns." *Journal of Labor Economics* 25, no. 2 (2007): 325-365.

Structural versus compositional homophily

	α			FM			P-values		Signif Comp	
Field	Obs	Exp	Struct	Obs	Exp	Struct	Main	Struct	Main	Struct
JSTOR	.11	.05	.00	38.6	41.1	43.3	.00	.00	82/280	157/280
Mol/Cell Bio	.05	.01	.00	38.2	39.8	40.2	.00	.00	19/44	35/44
Eco/Evol	.06	.02	.00	31.9	33.3	34.0	.00	.00	15/56	33/56
Economics	.11	.02	.00	18.7	20.8	21.1	.00	.00	11/28	18/28
Sociology	.19	.07	.00	38.5	44.2	47.4	.00	.00	12/21	19/21
Prob/Stat	.09	.03	.00	26.0	27.8	28.6	.00	.00	2/23	12/23
Org/mkt	.16	.04	.00	29.0	33.2	34.7	.00	.00	3/4	4/4
Education	.16	.04	.00	41.2	47.2	49.2	.00	.00	6/10	9/10
Occ Health	.10	.02	.00	41.7	45.5	46.3	.00	.00	1/1	1/1
Anthro	.12	.03	.00	38.5	42.0	43.5	.00	.00	2/8	4/8
Law	.17	.08	.00	29.7	32.9	35.7	.00	.00	1/16	4/16
History	.16	.07	.00	32.9	36.5	39.1	.00	.00	1/6	2/6
Phys Anthro	.07	.01	.00	34.7	36.8	37.1	.00	.00	2/10	2/10
Intl Poli Sci	.09	.02	.00	27.3	29.1	29.8	.03	.00	0/2	1/2
US Poli Sci	.15	.07	.00	25.2	27.4	29.6	.00	.00	1/6	2/6
Philosophy	.10	.03	.00	18.7	20.3	20.8	.03	.00	0/8	0/8
Math	.04	.01	.00	14.3	14.7	14.9	1.00	.12	0/9	0/9
Vet Med	.09	.01	.00	38.3	41.6	42.0	.00	.00	1/2	2/2
Cog Sci	.18	.09	.00	35.7	39.4	43.3	.00	.00	3/3	3/3
Radiation	.09	.01	.00	34.0	36.8	37.3	.00	.00	1/5	4/5
Demography	.15	.06	.00	40.3	44.4	47.3	.00	.00	1/2	1/2
Classics	.07	.01	.00	38.7	41.2	41.7	.27	.01	0/8	0/8
Opr Res	.03	.00	.00	16.6	17.1	17.1	.73	.33	0/4	0/4
Plant Phys	.08	.02	.00	29.1	31.0	31.8	.03	.00	0/3	1/3
Mycology	.03	.01	.00	36.8	37.5	37.9	1.00	.60	0/1	0/1

Data reduction due to unimputed gender

Label	Prop Authors with Unimputed Gender	Authors		Papers	
		Remaining	Prop Lost	Remaining	Prop Lost
Anthropology	0.12	9326	0.17	3466	0.15
Classical studies	0.08	1976	0.11	610	0.09
Cognitive science	0.10	12510	0.13	3814	0.07
Demography	0.16	5069	0.22	1930	0.17
Ecology and evolution	0.11	192091	0.13	66152	0.08
Economics	0.12	51691	0.19	22178	0.15
Education	0.08	24356	0.12	9396	0.09
History	0.08	3699	0.12	1596	0.11
Law	0.06	6526	0.10	2765	0.09
Mathematics	0.21	5319	0.31	2459	0.25
Molecular & Cell biology	0.16	303761	0.18	73357	0.07
Mycology	0.13	4828	0.17	1759	0.11
Operations research	0.20	7217	0.28	3025	0.21
Organizational and marketing	0.12	22299	0.18	9137	0.13
Philosophy	0.11	3897	0.18	1770	0.15
Physical anthropology	0.09	16463	0.12	5175	0.07
Plant physiology	0.10	5388	0.14	2287	0.10
Political science - international	0.10	5128	0.16	2247	0.14
Political science-US domestic	0.07	7269	0.11	3068	0.09
Pollution and occupational health	0.17	39703	0.18	8845	0.06
Probability and Statistics	0.19	18763	0.27	7600	0.21
Radiation damage	0.16	10710	0.18	2902	0.09
Sociology	0.08	35858	0.12	13600	0.09
Veterinary medicine	0.15	13741	0.17	3275	0.06
Total	0.14	807588	0.16	252413	0.11