Letter

Dynamics of social interaction: Modeling the genesis of scientific collaboration

Emma R. Zajdela,[1,*] Kimberly Huynh,[2] Andy T. Wen,[3] Andrew L. Feig,[2] Richard J. Wiener,[2] and Daniel M. Abrams[1,4,5,†]

[1]*Department of Engineering Sciences and Applied Mathematics, Northwestern University, Evanston, 60208 Illinois, USA*

[2]*Research Corporation for Science Advancement, Tucson, 85712 Arizona, USA*

[3]*Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, 60208 Illinois, USA*

[4]*Northwestern Institute for Complex Systems, Northwestern University, Evanston, 60208 Illinois, USA*

[5]*Department of Physics and Astronomy, Northwestern University, Evanston, 60208 Illinois, USA*

(Received 9 May 2022; accepted 20 July 2022; published 4 October 2022)

Collaboration plays a key role in physics and in the broader scientific enterprise. Here we develop a mathematical model for predicting new collaborations. We demonstrate that a simple ordinary differential equation model is a good fit to a data set that tracks collaborations resulting from four series of annual conferences on diverse scientific topics, 12 conferences in total over a period of five years. The model, inspired by the physics of catalysis, attempts to quantify the time-varying probability that any pair of individuals will initiate a new collaboration. It takes as input the pair's prior familiarity with one another as well as their pattern of interaction over time, and incorporates the effect of temporally decaying memory. This model accurately reproduces the collaborations formed across all first-year conferences in the four series and outperforms seven other candidate models. We also find evidence that prescribed interaction can lead to novel team formation, with observed collaboration probabilities increased by almost an order of magnitude. These results suggest that encounters among individual researchers at conferences, including encounters engineered by organizers, play an important role in shaping the future of science.

DOI: 10.1103/PhysRevResearch.4.L042001

I. INTRODUCTION

Physics-based models for social phenomena have had significant success in recent years (see, e.g., Refs. [1–6] and review papers [7–9]). In particular, the tools of statistical mechanics and nonlinear dynamics have been usefully applied to a wide range of phenomena from political polarization to team decision making to human activity prediction. In this work we attempt to develop a physical model for the catalysis of new collaboration[1] formation among scientists attending conferences.

The scientific enterprise has increasingly become a team effort [10–13]; the typical number of coauthors on physics papers has increased drastically since the 1960s [14]. Some evidence also suggests that multiauthor teams are more productive and do work that is more novel [15,16]. There are, however, intellectual, technical, and logistical obstacles, which impede the formation of new teams [17]. In particular, research has shown that geographical proximity is a factor in team assembly [18]. Conferences can help overcome these barriers and are one of the main catalysts for the formation of new scientific collaborations. However, convening conferences is expensive in terms of organizational, travel, environmental, and opportunity costs; the direct monetary cost for academic meetings alone is estimated at multiple 10^9 of US dollars each year [19,20].

Past research has mostly focused on measuring or modeling various aspects of scientific collaborations (understood as coauthorship on publications) once formed [21–23] and the makeup of successful teams (including, e.g., metrics such as the number of institutions present [24], team size [16], and team freshness [25]). There have been some efforts to study scientific team assembly (see, e.g., Refs. [12,17,26,27]), but little is still known about the impact of conferences on collaboration initiation. Some limited evidence, however, demonstrates that increased interaction among potential members raises the likelihood of team formation [28–30].

Here we present evidence that properly engineered interaction leads to collaboration, and we go beyond empirical observation by proposing a mathematical model for the origin of this phenomenon. Such a model has the potential to allow for optimization of conference design to promote collaboration.

II. NOVEL DATASET

We constructed a longitudinal data set derived from a diverse set of conferences known as "Scialogs" [31]. Organized by the nonprofit funding agency Research Corporation for

*emmazajdela@u.northwestern.edu

†dmabrams@northwestern.edu

[1]We use the terms "collaboration," "team formation," and "team assembly" interchangeably, as these terminologies can all be found in the literature designating the same phenomenon.

TABLE I. Conference statistics.

Conference	Year	Topic	Part.	Fellows	Pairs of Fellows
A	2015	Molecules Come to Life	64	52	1326
B	2015	Time Domain Astrophys.	59	49	1176
C	2017	Advanced Energy Storage	71	60	1170
D	2018	Chem. Machin. of the Cell	60	50	1225

Science Advancement (RCSA), these conferences seek to accelerate the work of science through research, intensive dialog, community building, and by catalyzing new scientific collaborations on challenges of global significance. Scialog conferences last three days and have an interactive format, with the participation of around 50 fellows, who are invited early-career scientists, and around ten facilitators, who are more senior scientists. Participants are assigned to small- (three to four people) and medium-sized (eight to ten people) sessions and self-assemble into teams at the end of the conference to submit proposals, five to eight of which are funded. For each conference, we have detailed records including: how well each participant knew the other participants before the conference (prior knowledge K_0), which sessions they attended [this defines the pairwise interaction profile $I(t)$], and with whom they wrote proposals (collaborated). Although Scialog conferences are multiyear initiatives, we focus here on the first year to omit effects of repeated participation. See Table I for conference statistics.

III. EMPIRICAL RESULTS

We first tested whether pairs who collaborated were different from pairs that did not in terms of total effective interaction I_{tot}, proportional to the time spent in a session and inversely proportional to its size (see Supplemental Material (SM) [32]). To do so, we employed the Mann-Whitney U test [33]; metrics were computed for each conference individually as well as aggregations of the conferences. All metrics indicate that collaborators have significantly more interaction than noncollaborators (see Fig. 1). The right panel of Fig. 1 shows that collaborators spent 63% more total effective time together than noncollaborators on average across all conferences. This is equivalent to being in a group of 12 people for an extra 45 m (60% of the duration of a topical discussion session) or being in a group of four people for an extra 15 m (50% of the duration of a small group session).

To disentangle causality from correlation for the effect of interaction on collaboration initiation, we performed a test based on 2500 counterfactual schedules for one of the conferences (see SM for details). For each counterfactual schedule, we computed the mean interaction[2] for all pairs that, in the actual conference, ended up collaborating, denoted $\bar{I}_{\text{CF}}$. We found that the mean interaction at the actual conference $\bar{I}_{\text{A}}$ was nearly always much greater; this was true for more than 99% of the counterfactual schedules. The only cases where $\bar{I}^{i}_{\text{CF}} > \bar{I}_{\text{A}}$ was observed corresponded to counterfactual scenarios sharing the same exact small-group session assignments but with variations in the larger topical discussion session assignments. Note that this method enabled us to blindly recover the small-group assignments knowing only which pairs ultimately collaborated, which strongly suggests a causal connection between intense interaction in a small-group setting and team formation. To quantify the statistical significance of this result, we performed a Wilcoxon signed-rank test [34]. The null hypothesis that the distribution of $\bar{I}_{\text{A}} - \bar{I}^{i}_{\text{CF}}$ has zero median is rejected at the 10^{-5} level of significance. See Fig. 2 for a graphical display of this result.

In addition to showing that interaction has a statistically significant effect on collaboration probability, we also wish to know the size of the effect. To evaluate that, we restrict our data to pairs with prior knowledge $K_0 = 0$ ($N = 984$) and used bootstrap statistics to estimate the odds of collaboration for pairs who coattended one minisession (0.15, 95% CI [0.10 0.21]) and those who did not coattend any minisession, but could have in one of the 2500 counterfactual scenarios (0.017, 95% CI [0.0085 0.028]). In this case, coattending a minisession multiplied the chance of a pair collaborating by 8.7.

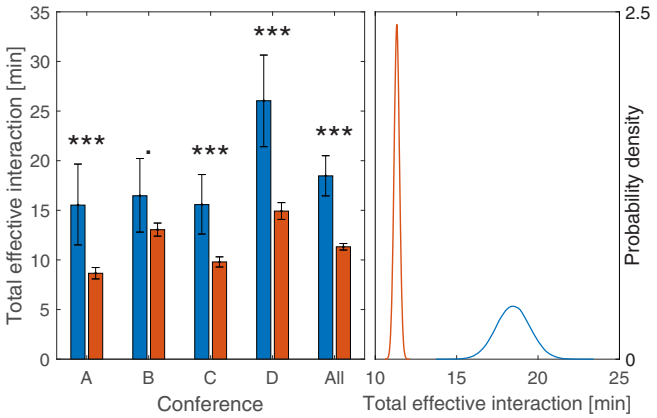

FIG. 1. Effect of interaction on collaboration. Left: Blue (left) and red (right) paired bars show bootstrap estimates for mean total effective interaction time for collaborators and noncollaborators, respectively. Paired bars are shown for each conference analyzed (A–D) as well as the combined data set of all conferences (All). p values of the Mann-Whitney U test for A: 6.0×10^{-4}, B: 6.3×10^{-2}, C: 1.4×10^{-4}, D: 7.7×10^{-6}, All: 1.0×10^{-12}. Error bars show mean values of the bootstrapped data with 95% confidence intervals. Right: kernel density estimates showing bootstrapped mean total effective interaction time distributions for collaborators (blue, right) and noncollaborators (red, left) for combined data set of all conferences.

[2]Here for brevity and clarity we use the word "interaction" to mean total effective interaction over the course of the conference.

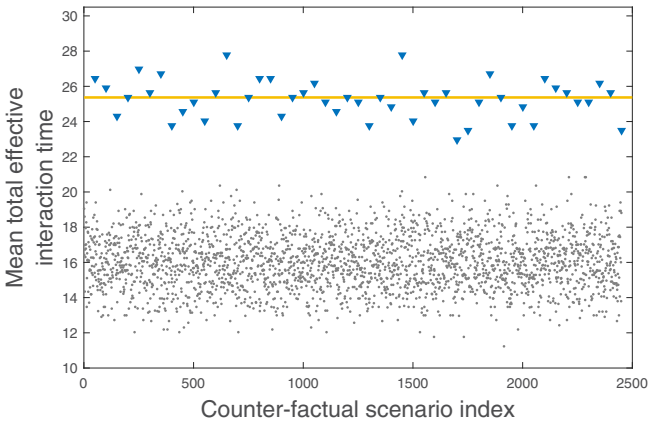

FIG. 2. Mean total effective interaction of collaborators. Yellow line represents the actual conference, blue inverted triangles are the counterfactual conference solutions where the collaborators were in the same small groups as in the actual conference. Gray points are the other counterfactual conference solutions.

IV. MODEL

Beyond empirical observations, we develop a mechanistic model for the dynamics of team formation at conferences. In 1996, the physicist Serge Galam wrote: "Do humans behave like atoms?" [35]. The model we present is based on the idea that scientists at a conference behave like molecules in a solution, where formation of a collaboration is analogous to undergoing a chemical reaction. The conference itself acts as a catalyst by lowering the barriers to collaboration and creating more productive collisions among the scientists who participate. The model takes as input the pairwise levels of interaction $I(t)$ among conference participants as well as their prior knowledge K_0 of one another, and estimates the probability $P(t)$ that any pair of participants will subsequently form a collaboration.

As a first simple model, consider a pair of attendees at a conference. We assume that collaboration probability $P(t)$ rises for nonzero interaction intensity I between participants, and when interaction ceases, probability of a collaboration forming decays. For simplicity we assume linear growth and decay processes, leading to the following ordinary differential equation (ODE) governing the change in collaboration probability over time:

$$\frac{dP}{dt} = \underbrace{S\frac{I}{I_{\max}}(1-P)}_{\text{strengthening}} - \underbrace{WP\left(1-\frac{I}{I_{\max}}\right)}_{\text{weakening}}. \tag{1}$$

To understand the changes in collaboration probabilities that may occur during the course of a conference, we focus on the case where $I = I(t)$ is not constant, representing the time-varying strength of interaction between two individuals. Figure 3(a) shows a realistic looking example for a two-day conference, with three sessions of different lengths and intensities, one small (four people) and two larger (12 people) in this example. Note that $I(t)$ is a dimensionless quantity; for more details on how $I(t)$ was constructed from data, see SM.

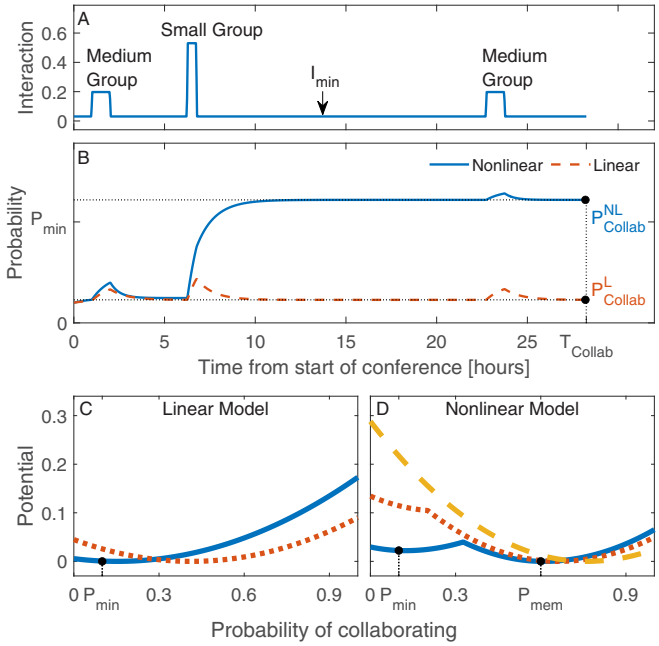

FIG. 3. Model examples. (a) Effective interaction $I(t)$ and (b) corresponding probability of collaboration as a function of time for a single pair of participants at a realistic conference. Example potential for (c) linear model, and (d) nonlinear catalysis model. In (b), solid blue curve shows the probability for the nonlinear catalysis model and dashed red curve shows the probability for the linear model. T_{Collab} is the time at which teams of participants are formed, $P^{\text{NL}}_{\text{Collab}}$ ($P^{\text{L}}_{\text{Collab}}$) is the probability at this time for the nonlinear (linear) model. In (c) and (d), solid blue curve shows potential for minimal interaction, dotted red curve shows potential for medium interaction (medium group), and in (d) dashed yellow curve shows potential for high interaction (small group). Parameter values are the same for linear and nonlinear models where applicable, selected for illustrative purposes ($a = 0.02$, $I_c = 0.1$; $I_{\max} = 0.6$; $P_{\min} = 0.1$; $P_{\text{mem}} = 0.6$; $P_{\max} = 0.9$; $W = 1$; $S = 0.5$.)

We can express the right-hand side of Eq. (1) as the derivative of a potential function $V(P)$:

$$\frac{dP}{dt} = -\frac{\partial V}{\partial P}, \tag{2}$$

where the minus sign implies that stable equilibria occur at potential minima. Since Eq. (1) is linear, the resulting potential is quadratic, with a local minimum somewhere in $0 \leqslant P \leqslant 1$ (depending on I). Figure 3(c) shows examples of such potentials.

We expect the linear model described above to capture some of the dynamics of formation of new scientific collaborations during conferences. One major limitation, however, is that the linear model relaxes to zero after interaction has ceased, which implies that participants completely forget one another. For a more realistic generalization, we wish to allow scientists who have interacted sufficiently to remember one another long after the interaction has ceased. To implement this, we modify the potential landscape for a new nonlinear model as shown in Fig. 3(d).

When interaction $I = 0$, there are two stable equilibria, one at the minimum probability $P_{\min}$, and the other at memory

TABLE II. Model selection.

Conf.	Nonlinear Model AIC	Next Best Model AIC	Next Best Model	Relative Likelihood
A	375.88	385.62	$aK_0 + bI_{\text{tot}} + c$	0.0077
B	337.02	340.79	$aK_0 + b$	0.15
C	522.91	541.64	$aK_0 + bI_{\text{tot}} + c$	8.6e–05
D	407.96	430.53	$aI_{\text{tot}} + b$	1.3e–05

state P_{mem}, with an unstable equilibrium in between. As the interaction increases, it acts as a catalyst by changing the shape of the potential function and reducing the barrier between the two stable states. At a critical value of the interaction I_c a bifurcation occurs and the barrier disappears, leaving only a single stable equilibrium. If the system gets sufficiently close to that new equilibrium before interaction ceases, the probability will remain permanently in the higher memory state P_{mem}. See SM for exact form.

V. MODEL SELECTION

We validate the nonlinear catalysis model by testing how well it explains which pairs of participants ultimately collaborated. The probability of collaborating is the output of the model at time $t = T_{\text{Collab}}$ (see Fig. 3), the start of the period allocated for team formation and proposal writing at the end of the conference. We compared the quality of the nonlinear catalysis model to seven null models by computing the Akaike information criterion (AIC) [36] and relative likelihood for each one. Table II shows that the AIC of the nonlinear catalysis model is lower compared to the next best model for conferences A, B, C, and D, indicating that it is the preferred model for all four conferences.

VI. DISCUSSION

Our analysis is predicated on the quality of interpersonal interactions as well as their quantity. Impromptu meetings around the coffee maker may differ from those at a conference where participants were encouraged to have a specific conversation and incentivized to form teams through a grant-awarding process, as was the case in the data set analyzed here. Another limitation of this proposed model is that we do not explicitly account for many issues likely affecting team assembly such as personality characteristics, homophily, and distance between research areas. However, our approach implicitly incorporates these to some extent through its probabilistic nature. They could also be explicitly incorporated in a more complex future model, but we see the success of the nonlinear catalysis model as remarkable precisely because of its simplicity.

The nonlinear catalysis model is not necessarily limited to scientific conferences and collaboration; we speculate that it may also have applicable extensions in other areas where matches between individuals within a network are sought. For example, in business settings, employers may wish to promote organic team building through prescribed sessions among employees. In romantic contexts, a model could inform online dating algorithms and approaches to social interaction. In pedagogical settings, educators might use in-class prescribed group exercises to promote formation of student study groups or teams for collaborative assignments.

VII. CONCLUSIONS

Scientific conferences play an important role in the diffusion of knowledge and generation of novel ideas. We have shown that properly engineered interaction at conferences induces the formation of new scientific collaborations. Our model helps to illuminate the mechanism by which this occurs, and we hope that it will play a role in designing more efficient future conferences to maximize their benefit for physics and science in general. A web application that allows users to interactively explore the model can be found at [37].

ACKNOWLEDGMENTS

The authors acknowledge the United States Department of Agriculture NACA 58-3022-0-005 and the Research Corporation for Science Advancement for providing data and assistance. E.R.Z thanks the National Science Foundation Graduate Research Fellowship Program DGE-184216 and the Northwestern Buffett Institute Global Impacts Graduate Research Fellowship for financial support and Maher Said for productive discussions.

[1] F. Baumann, P. Lorenz-Spreen, I. M. Sokolov, and M. Starnini, Modeling Echo Chambers and Polarization Dynamics in Social Networks, Phys. Rev. Lett. **124**, 048301 (2020).

[2] B. Karamched, M. Stickler, W. Ott, B. Lindner, Z. P. Kilpatrick, and K. c. v. Josić, Heterogeneity Improves Speed and Accuracy in Social Networks, Phys. Rev. Lett. **125**, 218302 (2020).

[3] P. J. Górski, K. Bochenina, J. A. Hołyst, and R. M. D'Souza, Homophily Based on Few Attributes Can Impede Structural Balance, Phys. Rev. Lett. **125**, 078302 (2020).

[4] X. Wang, A. D. Sirianni, S. Tang, Z. Zheng, and F. Fu, Public Discourse and Social Network Echo Chambers Driven by Socio-Cognitive Biases, Phys. Rev. X **10**, 041042 (2020).

[5] C. W. Lynn, L. Papadopoulos, D. D. Lee, and D. S. Bassett, Surges of Collective Human Activity Emerge from Simple Pairwise Correlations, Phys. Rev. X **9**, 011022 (2019).

[6] H. Z. Brooks and M. A. Porter, A model for the influence of media on the ideology of content in online social networks, Phys. Rev. Res. **2**, 023041 (2020).

[7] C. Castellano, S. Fortunato, and V. Loreto, Statistical physics of social dynamics, Rev. Mod. Phys. **81**, 591 (2009).

[8] M. Jusup, P. Holme, K. Kanazawa, M. Takayasu, I. Romić, Z. Wang, S. Geček, T. Lipić, B. Podobnik, L. Wang, W. Luo, T. Klanjšček, J. Fan, S. Boccaletti, and M. Perc, Social physics, Phys. Rep. **948**, 1 (2022).

[9] M. Perc, The social physics collective, Sci. Rep. **9**, 16549 (2019).

[10] S. Wuchty, B. F. Jones, and B. Uzzi, The increasing dominance of teams in production of knowledge, Science **316**, 1036 (2007).

[11] S. Fortunato, C. T. Bergstrom, K. Börner, J. A. Evans, D. Helbing, S. Milojević, A. M. Petersen, F. Radicchi, R. Sinatra, B. Uzzi, A. Vespignani, L. Waltman, D. Wang, and A.-L. Barabási, Science of science, Science **359**, eaao0185 (2018).

[12] S. Milojević, Principles of scientific research team formation and evolution, Proc. Natl. Acad. Sci. USA **111**, 3984 (2014).

[13] M. Ahmadpoor and B. F. Jones, Decoding team and individual impact in science and invention, Proc. Natl. Acad. Sci. USA **116**, 13885 (2019).

[14] S. Mallapaty, Paper authorship goes hyper, Nature Index (January 30, 2018).

[15] B. Uzzi, S. Mukherjee, M. Stringer, and B. Jones, Atypical combinations and scientific impact, Science **342**, 468 (2013).

[16] L. Wu, D. Wang, and J. A. Evans, Large teams develop and small teams disrupt science and technology, Nature (London) **566**, 378 (2019).

[17] A. Lungeanu, Y. Huang, and N. S. Contractor, Understanding the assembly of interdisciplinary teams and its impact on performance, J. Informetric. **8**, 59 (2014).

[18] A. Wax, L. A. DeChurch, and N. S. Contractor, Self-organizing into winning teams: Understanding the mechanisms that drive successful collaborations, Small Group Res. **48**, 665 (2017).

[19] M. Klöwer, D. Hopkins, M. Allen, and J. Higham, An analysis of ways to decarbonize conference travel after COVID-19, Nature (London) **583**, 356 (2020).

[20] S. Sarabipour, A. Khan, and Y. Seah, Changing scientific meetings for the better, Nature Hum. Behav. **5**, 296–300 (2021).

[21] R. J. Funk and J. Owen-Smith, A dynamic network measure of technological change, Manag. Sci. **63**, 791 (2017).

[22] G. Siudem, B. Żogała Siudem, A. Cena, and M. Gagolewski, Three dimensions of scientific impact, Proc. Natl. Acad. Sci. USA **117**, 13896 (2020).

[23] E. Vasilyeva, A. Kozlov, K. Alfaro-Bittner, D. Musatov, A. M. Raigorodskii, M. Perc, and S. Boccaletti, Multilayer representation of collaboration networks with higher-order interactions, Sci. Rep. **11**, 5666 (2021).

[24] B. F. Jones, S. Wuchty, and B. Uzzi, Multi-university research teams: Shifting impact, geography, and stratification in science, Science **322**, 1259 (2008).

[25] A. Zeng, Y. Fan, Z. Di, Y. Wang, and S. Havlin, Fresh teams are associated with original and multidisciplinary research, Nature Hum. Behav. **5**, 1314 (2021).

[26] R. Guimerà, B. Uzzi, J. Spiro, and L. A. N. Amaral, Team assembly mechanisms determine collaboration network structure and team performance, Science **308**, 697 (2005).

[27] R. B. Freeman, I. Ganguli, and R. Murciano-Goroff, Why and wherefore of increased scientific collaboration, in *The Changing Frontier: Rethinking Science and Innovation Policy* (University of Chicago Press, Chicago, 2014), pp. 17–48.

[28] S. Chai and R. B. Freeman, Temporary colocation and collaborative discovery: Who confers at conferences, Strat Mgmt J **40**, 2138 (2019).

[29] R. Campos, F. Leon, and B. McQuillin, Lost In the Storm: The Academic Collaborations That Went Missing In Hurricane Issac, Econ. J. **128**, 995 (2018).

[30] K. J. Boudreau, T. Brady, I. Ganguli, P. Gaule, E. Guinan, A. Hollenberg, and K. R. Lakhani, A Field Experiment on Search Costs and the Formation of Scientific Collaborations, Rev. Econ. Stat. **99**, 565 (2017).

[31] R. J. Wiener and S. Ronco, Scialog: The catalysis of convergence, ACS Energy Lett. **4**, 1020 (2019).

[32] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevResearch.4.L042001 for details on the nonlinear catalysis model potential, pre-conference survey, group assignments, and example interaction patterns.

[33] H. B. Mann and D. R. Whitney, On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other, Ann. Math. Statist. **18**, 50 (1947).

[34] F. Wilcoxon, Individual comparisons by ranking methods, Biometrics Bull. **1**, 80 (1945).

[35] S. Galam, When humans interact like atoms, in *Understanding Group Behavior*, edited by I. E. Witte and J. Davis (Psychology Press, Milton Park, 1996), Vol. 1, pp. 293–312.

[36] H. Akaike, A new look at the statistical model identification, IEEE Trans. Autom. Control **19**, 716 (1974).

[37] E. R. Zajdela, Exploring the dynamics of team formation at conferences, shiny Web Application Framework for R (2022), https://mathemmatician.shinyapps.io/DynamicsOfConferences/.

SUPPLEMENTAL MATERIAL FOR

Dynamics of social interaction: Modeling the genesis of scientific collaboration

Emma R. Zajdela,[1,*] Kimberly Huynh,[2] Andy T. Wen,[3] Andrew L. Feig,[2] Richard J. Wiener,[2] and Daniel M. Abrams[1,4,5,†]

[1] *Department of Engineering Sciences and Applied Mathematics, Northwestern University, Evanston, IL, USA*
[2] *Research Corporation for Science Advancement, Tucson, AZ, USA*
[3] *Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL, USA*
[4] *Northwestern Institute for Complex Systems, Northwestern University, Evanston, IL, USA*
[5] *Department of Physics and Astronomy, Northwestern University, Evanston, IL, USA*

This file contains:

- Supplemental text
- Supplemental Figures S1 to S3

S1. NONLINEAR CATALYSIS MODEL POTENTIAL

We wish to construct a potential function with the following key properties:

1. Two local minima with a barrier between them that lowers, then disappears, as interaction intensity I increases.
2. When $I = 0$, the two local minima should be located at $P_{\min}$ and P_{mem}.
3. Locally quadratic in P so that the corresponding ODE system ($dP/dt = -\nabla V$) is locally linear (at least in the neighborhood of each minimum).
4. Asymmetric curvature about each local minimum corresponding to asymmetric strengthening and weakening rate constants S and W.

These properties allow for a range of possible potential functions. We make several additional choices with the goal of producing a simple and tractable function:

1. Match form of nonlinear catalysis model with form of linear model for low interaction intensity ($I \ll 1$) and low collaboration probability ($P \ll 1$).
2. Locate the potential barrier at the midpoint between the two potential minima (i.e., at $P = \frac{1}{2}(P_{\min} + P_{\text{mem}})$) when $I = 0$.
3. As interaction intensity I increases, deflect both minima linearly and equally in the $+P$ direction such that the upper minimum moves from P_{mem} to $P_{\max}$ as I goes from 0 to $I_{\max}$.
4. Introduce a tunable parameter I_{c} such that the barrier exists for $I < I_{\text{c}}$ and disappears for $I \geq I_{\text{c}}$. To accomplish this, as interaction intensity I increases, deflect the barrier location (local maximum where branches of piecewise function meet) linearly in the $-P$ direction so that the barrier meets the lower potential minimum at interaction intensity $I = I_{\text{c}}$ (and thus only one equilibrium exists for $I \geq I_{\text{c}}$).

See Fig. S1 for an illustration of how the potential minima and the potential barrier depend on interaction intensity. See Fig. S2 for an illustration of the potential landscapes for varying interaction intensity in the nonlinear catalysis model and each of the below simplifications.

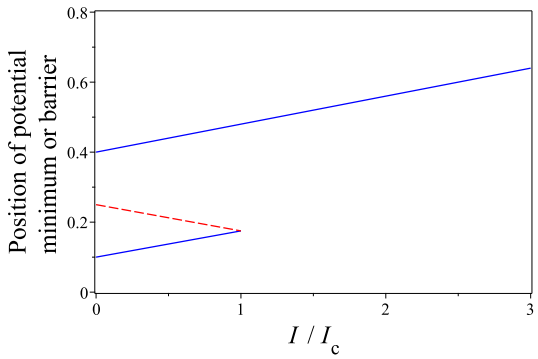

FIG. S1. **Example positions of minima and barrier.** Blue solid lines: Lower and upper potential minima, corresponding to $P_{\min}$ and P_{mem} when $I = 0$. Red dashed line: Potential barrier position ($\frac{1}{2}(P_{\min} + P_{\text{mem}})$ when $I = 0$). Potential barrier disappears when $I = I_{\text{c}}$. Constants were chosen for illustration purposes only (here set to $P_{\min} = 0.1$, $P_{\text{mem}} = 0.4$, $P_{\max} = 0.8$, $I_{\text{c}} = 1$, $I_{\max} = 5$).

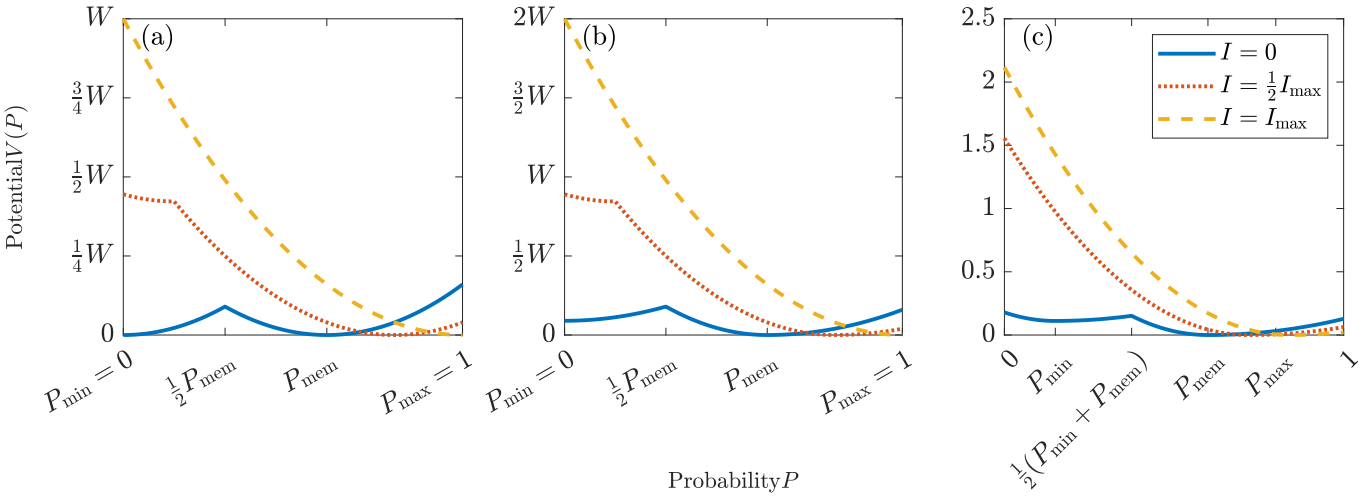

FIG. S2. **Potential functions for increasing interaction**. Panel (a) is the simplest version with parameters $P_{\min} = 0, P_{\max} = 1, I_{\max} = 1, I_c = 1/2, S = W, W = 1, P_{\mathrm{mem}} = 0.6$. Panel (b) is the simplified version with $P_{\min} = 0, P_{\max} = 1, I_{\max} = 1, I_c = 1/2, S = 2W, W = 1, P_{\mathrm{mem}} = 0.6$. Panel (c) is the full version with $P_{\min} = 0.15, P_{\max} = 0.8, I_{\max} = 1.2, I_c = 1/2, S = 3, W = 0.8, P_{\mathrm{mem}} = 0.6$. In each case, the solid blue curve shows the potential for interaction $I = 0$, the dotted red curve shows the potential for $I = \frac{1}{2}I_{\max}$ and the dashed yellow curve shows the potential for $I = I_{\max}$. Note that the vertical axis changes in each panel.

We write the potential function for the nonlinear catalysis model as a piecewise function of P dependent on the interaction intensity I. Although conceptually straightforward, its algebraic representation appears complicated because of its piecewise nature. Because of that, we first present two simplified special cases to illustrate its structure.

A. Simplest version where $P_{\min} = 0, P_{\max} = 1, I_{\max} = 1, I_c = 1/2, S = W$ (symmetric strengthening and weakening rates)

$$V(P) = \begin{cases} W\left[(P - \frac{1}{2}IP_{\mathrm{mem}})^2 - \frac{1}{4}I(P_{\mathrm{mem}} + 2)(3IP_{\mathrm{mem}} - 2I - 2P_{\mathrm{mem}})\right] & P < \frac{1}{2}P_{\mathrm{mem}}(1 - I) \\ W\left[P - I - (1 - I)P_{\mathrm{mem}}\right]^2 & P \geq \frac{1}{2}P_{\mathrm{mem}}(1 - I) \end{cases} \tag{1}$$

B. Simplified version where $P_{\min} = 0, P_{\max} = 1, I_{\max} = 1, I_c = 1/2, S = 2W$ (asymmetric strengthening and weakening rates)

First case: $0 \leq I \leq I_c$

$$V_{lowI}(P) = \begin{cases} 2W\Big[(P - \frac{1}{2}IP_{\mathrm{mem}})^2 & P \leq \frac{1}{2}IP_{\mathrm{mem}} \\ \quad +(-\frac{1}{4}P_{\mathrm{mem}}^2 - P_{\mathrm{mem}} + 1)I^2 + IP_{\mathrm{mem}} + \frac{1}{8}P_{\mathrm{mem}}^2\Big] & \\ W\Big[(P - \frac{1}{2}IP_{\mathrm{mem}})^2 & \frac{1}{2}IP_{\mathrm{mem}} < P \leq \frac{1}{2}(1 - I)P_{\mathrm{mem}} \\ \quad +(-\frac{1}{2}P_{\mathrm{mem}}^2 - 2P_{\mathrm{mem}} + 2)I^2 + 2IP_{\mathrm{mem}} + \frac{1}{4}P_{\mathrm{mem}}^2\Big] & \\ 2W\left[P - I - (1 - I)P_{\mathrm{mem}}\right]^2 & \frac{1}{2}(1 - I)P_{\mathrm{mem}} < P \leq P_{\mathrm{mem}} + (1 - P_{\mathrm{mem}})I \\ W\left[P - I - (1 - I)P_{\mathrm{mem}}\right]^2 & P_{\mathrm{mem}} + (1 - P_{\mathrm{mem}})I < P \end{cases} \tag{2}$$

Second case: $I_{\rm c} < I \leq I_{\rm cint}$

$$V_{medI}(P) = \begin{cases} 2W\left[(P - \frac{1}{2}IP_{\rm mem})^2 + (-\frac{1}{4}P_{\rm mem}^2 - P_{\rm mem} + 1)I^2 + IP_{\rm mem} + \frac{1}{8}P_{\rm mem}^2\right] & P \leq P_{\rm int} \\ 2W\left[P - I - (1-I)P_{\rm mem}\right]^2 & P_{\rm int} < P \leq P_{\rm mem} + (1-P_{\rm mem})I \\ W\left[P - I - (1-I)P_{\rm mem}\right]^2 & P_{\rm mem} + (1-P_{\rm mem})I < P \end{cases} \tag{3}$$

with

$$P_{\rm int} = -\frac{P_{\rm mem}\left[(I^2 - 2I + \frac{7}{8})P_{\rm mem} - I^2 + I\right]}{3IP_{\rm mem} - 2I - 2P_{\rm mem}} \tag{4}$$

Third case: $I > I_{\rm cint}$

$$V_{highI}(P) = \begin{cases} 2W\left[P - I - (1-I)P_{\rm mem}\right]^2 & P \leq P_{\rm mem} + (1-P_{\rm mem})I \\ W\left[P - I - (1-I)P_{\rm mem}\right]^2 & P_{\rm mem} + (1-P_{\rm mem})I < P \end{cases} \tag{5}$$

C. Full version ($P_{\rm min} \neq 0$, arbitrary constants)

First case: $0 \leq I \leq I_{\rm c}$

$$V_{lowI}(P) = \frac{1}{I_{\rm max}^2}\begin{cases} V_1(P) & P \leq P_{\rm low} \\ V_2(P) & P_{\rm low} < P \leq P_{\rm med} \\ S[(P_{\rm mem} - P)I_{\rm max} + (P_{\rm max} - P_{\rm mem})I]^2 & P_{\rm med} < P_{\rm high} \\ W[(P_{\rm mem} - P)I_{\rm max} + (P_{\rm max} - P_{\rm mem})I]^2 & P_{\rm high} < P \end{cases} \tag{6}$$

with

$$P_{\rm low} = P_{\rm min} + \frac{(P_{\rm mem} - P_{\rm min})I}{4I_{\rm c}} \tag{7}$$

$$P_{\rm med} = \frac{P_{\rm min}(I + 2I_{\rm c}) - P_{\rm mem}(I - 2I_{\rm c})}{4I_{\rm c}} \tag{8}$$

$$P_{\rm high} = P_{\rm mem} + \frac{(P_{\rm max} - P_{\rm mem})I}{I_{\rm max}} \tag{9}$$

$$\begin{aligned} V_1(P) = \frac{1}{8I_{\rm c}^2}\Big(&\{[(8P^2 - 16PP_{\rm min} + 2P_{\rm mem}^2 - 4P_{\rm mem}P_{\rm min} + 10P_{\rm min}^2)S - 2(P_{\rm mem} - P_{\rm min})^2W]I_{\rm c}^2 \\ &-4(P_{\rm mem} - P_{\rm min})I[(-P_{\rm min}/2 - P_{\rm mem}/2 + P)S - W(P_{\rm mem} - P_{\rm min})]I_{\rm c} + I^2(P_{\rm mem} - P_{\rm min})^2(S - 2W)\}I_{\rm max}^2 \\ &+4ISI_{\rm c}(P_{\rm mem} - P_{\rm min})(P_{\rm max} - P_{\rm mem})(I + 2I_{\rm c})I_{\rm max} + 8I^2SI_{\rm c}^2(P_{\rm max} - P_{\rm mem})^2\Big) \end{aligned} \tag{10}$$

and

$$\begin{aligned} V_2(P) = \frac{1}{16I_{\rm c}^2}\Big(&\{[(4S - 4W)P_{\rm mem}^2 - 8P_{\rm min}(S - W)P_{\rm mem} + 4SP_{\rm min}^2 + 16(P - P_{\rm min}/2)(P - (3P_{\rm min})/2)W]I_{\rm c}^2 \\ &-8(P_{\rm mem} - P_{\rm min})((-S/2 - W)P_{\rm mem} + PW + P_{\rm min}S/2)II_{\rm c} + I^2(P_{\rm mem} - P_{\rm min})^2(S - 3W)\}I_{\rm max}^2 \\ &+8ISI_{\rm c}(P_{\rm mem} - P_{\rm min})(P_{\rm max} - P_{\rm mem})(I + 2I_{\rm c})I_{\rm max} + 16I^2SI_{\rm c}^2(P_{\rm max} - P_{\rm mem})^2\Big) \end{aligned} \tag{11}$$

Second case: $I_{\rm c} < I \leq I_{\rm cint}$

$$V_{medI}(P) = \begin{cases} V_3(P) & P \leq P_{\rm int} \\ \dfrac{S[(P_{\rm mem}-P)I_{\rm max} + (P_{\rm max}-P_{\rm mem})I]^2}{I_{\rm max}^2} & P_{\rm int} < P \leq P_{\rm mem} + \dfrac{(P_{\rm max}-P_{\rm mem})I}{I_{\rm max}} \\ \dfrac{W[(P_{\rm mem}-P)I_{\rm max} + (P_{\rm max}-P_{\rm mem})I]^2}{I_{\rm max}^2} & P_{\rm mem} + \dfrac{(P_{\rm max}-P_{\rm mem})I}{I_{\rm max}} \leq P \end{cases} \tag{12}$$

with

$$\begin{aligned} P_{\rm int} = \frac{1}{16SI_{\rm c}\{[(P_{\rm max}-P_{\rm mem})I + I_{\rm max}(P_{\rm mem}-P_{\rm min})]I_{\rm c} - \dfrac{II_{\rm max}(P_{\rm mem}-P_{\rm min})}{4}\})} \{[8S(P_{\rm mem}+P_{\rm min})(P_{\rm max}-P_{\rm mem})I \\ +6((S+\frac{W}{3})P_{\rm mem} + 5(S-\frac{W}{5})\frac{P_{\rm min}}{3})(P_{\rm mem}-P_{\rm min})I_{\rm max}]I_{\rm c}^2 - 4(P_{\rm mem}-P_{\rm min})[S(P_{\rm max}-P_{\rm mem})I \\ +(((S+2W)P_{\rm mem} + P_{\rm min}(S-2W))I_{\rm max})/2]II_{\rm c} - I^2I_{\rm max}(P_{\rm mem}-P_{\rm min})^2(S-2W)\} \end{aligned} \tag{13}$$

and

$$\begin{aligned} V_3(P) = \frac{1}{8I_{\rm max}^2I_{\rm c}^2}\Big(\{[(8P^2 - 16PP_{\rm min} + 2P_{\rm mem}^2 - 4P_{\rm mem}P_{\rm min} + 10P_{\rm min}^2)S - 2(P_{\rm mem}-P_{\rm min})^2W]I_{\rm c}^2 \\ -4(P_{\rm mem}-P_{\rm min})I((-\frac{P_{\rm min}}{2} - \frac{P_{\rm mem}}{2} + P)S - W(P_{\rm mem}-P_{\rm min}))I_{\rm c} + I^2(P_{\rm mem}-P_{\rm min})^2(S-2W)\}I_{\rm max}^2 \\ +4ISI_{\rm c}(P_{\rm mem}-P_{\rm min})(P_{\rm max}-P_{\rm mem})(I+2I_{\rm c})I_{\rm max} + 8I^2SI_{\rm c}^2(P_{\rm max}-P_{\rm mem})^2\Big) \end{aligned} \tag{14}$$

Third case: $I > I_{\rm cint}$

$$V_{highI}(P) = \begin{cases} \dfrac{S[(P_{\rm mem}-P)I_{\rm max} + (P_{\rm max}-P_{\rm mem})I]^2}{I_{\rm max}^2} & P \leq P_{\rm mem} + \dfrac{(P_{\rm max}-P_{\rm mem})I}{I_{\rm max}} \\ \dfrac{W[(P_{\rm mem}-P)I_{\rm max} + (P_{\rm max}-P_{\rm mem})I]^2}{I_{\rm max}^2} & P_{\rm mem} + \dfrac{(P_{\rm max}-P_{\rm mem})I}{I_{\rm max}} \leq P \end{cases} \tag{15}$$

Value of $I_{\rm cint}$
$I_1 = I$ such that $P_{\rm int}(I) = 0$, $I \in \mathbb{R}$, $0 \leq I \leq I_{\rm max}$,
$I_2 = I$ such that $P_{\rm int}(I) = 1$, $I \in \mathbb{R}$, $0 \leq I \leq I_{\rm max}$,
$I_{\rm cint} = \min(I_1, I_2)$.

S2. PRE-CONFERENCE SURVEY

Before attending the conference, participants and fellows were asked to complete the following survey.
100% of participants completed the pre-conference survey for conferences A, C and D and 98% for Conference B.

Prior Knowledge

For each name please choose one answer that best describes your relationship with that person prior to this Scialog meeting. There are four categories to choose from:

Unfamiliar: You are not aware of the research of the person.

Awareness: Choose this option if you are aware of the research of the person. Examples of "awareness" would be knowing the person's specific area of expertise or knowing details of a recent publication.

Discussion: Choose this option if you have had a substantive discussion about research with this person, through face-to-face conversation, email correspondence, or other means. Please do not select this choice if you have talked with this person and exchanged only basic information about the areas you work in. This level of relationship is meant to be higher than the previous level of "awareness" and presupposes awareness.

Collaborator: Choose this option if you have ever worked on a project or written a paper together, or formally collaborated with this person on or toward a tangible research output. Please do not select this choice if you have only technically "collaborated" but have never had a substantive research discussion with this person (e.g., coauthored a paper with 100 authors but never interacted). This level of relationship is meant to be higher than the previous level of "discussion" and presupposes awareness and discussion.

Names are listed alphabetically.

Surveys are customized to each respondent. Your name will not appear on the list.

Interest in Discussion Topics

Please choose your interest level for the proposed discussion topics below. Your input will be used to select the topics for discussion groups at the conference and help us choose which groups you'll be in. Click the "details" button to see more information.

These topics are based on suggestions made by Scialog Fellows, including you, in the conference registration form. Our hope is you will be able to indicate at least a few, and perhaps many, that you are "really into" or would "chime in."

The order of topics is randomized.
Respondents are asked to rate each topic on a 5-point scale: No way - Might nap - Would listen - Would chime in - Really into it.

Nominating critical discussion participants

Listed below are the topics you expressed interest in. For each topic, if you think another Scialog Fellow is an essential person to have in a discussion on that topic, please indicate them below. You may select up to two for each topic but aren't required to select any. Click on the box and start typing or scroll to select a fellow.

The pre-conference survey results were incorporated into the interaction function as "prior knowledge" K_0 for each pair of fellows (A,B) where K_0 is the sum of prior knowledge reported by A about B and B about A. Thus K_0 for each pair ranges from 0–6 where 0 represents both fellows being unfamiliar with each other and 6 represents both fellows reporting having previously collaborated.

The rules of the Scialog conferences do not allow for participants who have previously collaborated (i.e. pairs with $K_0 \geq 5$) to be on the same proposal submission team. Therefore, when fitting the model to data, we eliminated pairs with $K_0 \geq 5$ (2.1% of pairs at Conference A, 11.7% of pairs at Conference B, 3.1% of pairs at Conference C, 1.5% of pairs at Conference D).

S3. GROUP ASSIGNMENTS

Group assignments were determined prior to the conference with the goal of creating diverse groups for the topical and small group discussion sessions. For the topical discussion groups most of the fellows who were placed in a group had rated their interest in the topic as a 4 or 5 (on a 1-5 scale with 5 indicating the most interest). Fellows were not placed in a group if they rated the topic under 3. These assignments were accomplished while maintaining diversity in the groups in terms of academic disciplines, research methodologies (e.g. theoretical vs experimental methods), and gender.

For the small groups, nearly all fellows in a group had no previous awareness of the others' research and none had previously engaged in scientific discussions with the other group members. Participants were mixed so that most small groups included fellows with different disciplines and methodologies.

A simulated annealing algorithm was used to provide candidate groupings based on these criteria. Group assignments for all topical sessions were optimized simultaneously to minimize the same fellows having repeated assignments together in different sessions. Similarly, all the small group sessions were optimized simultaneously so that no fellow was ever placed in a small group with another specific fellow more than once. The algorithm typically returns several solutions with the same or similar energy levels, especially in the case of the small group sessions, which were less constrained. The organizers made the final selection of the group assignments from among the several best solutions.

Participants who had previously collaborated were not allowed to submit a proposal together, and we therefore eliminated these pairs when fitting models to data. The median percentage of pairs omitted for this reason was 4.0%.

S4. EXAMPLE INTERACTION PATTERNS

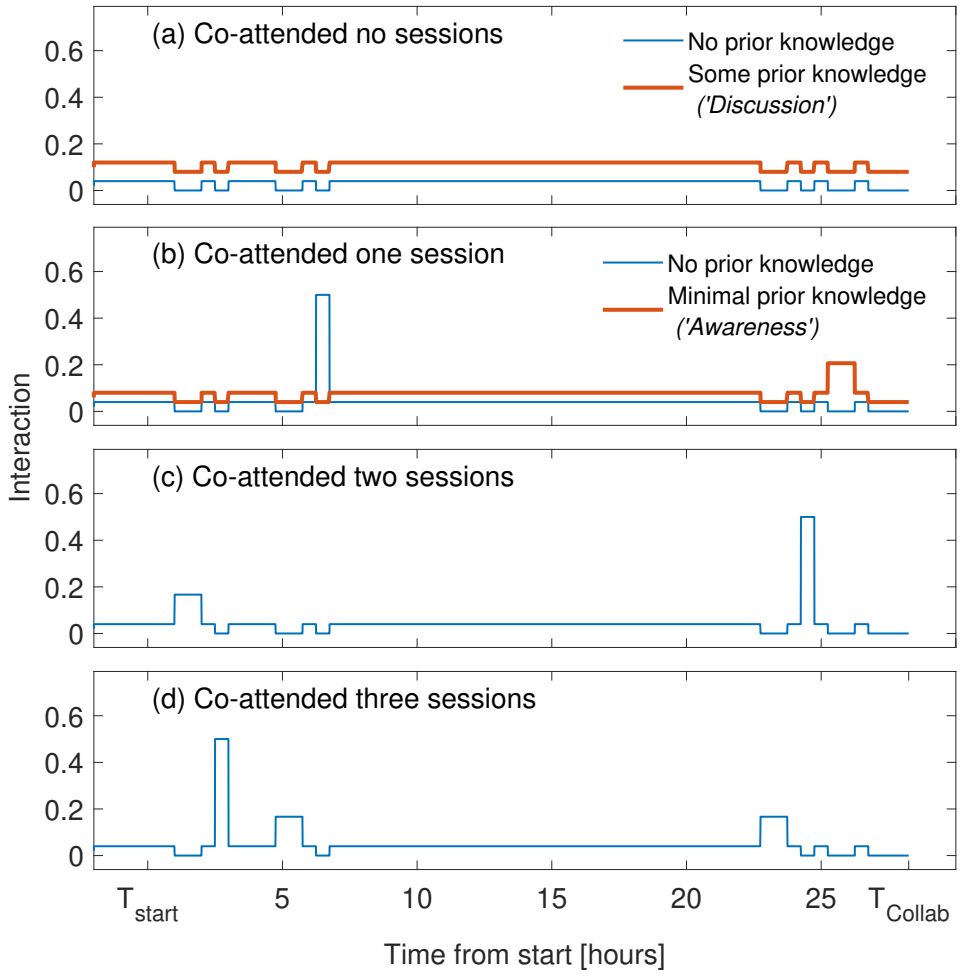

FIG. S3. **Example interactions between pairs of participants** from an hour before the first session T_{START} to the time when proposal writing teams are formed T_{COLLAB}. Higher tophat functions correspond to small group sessions (3-4 people) and medium tophat functions correspond to topical discussion sessions (around 12 people). Drops to zero occur when participants are in different simultaneous sessions. In panel (b), participants with minimal prior knowledge ($K_0 \leq 2$) may co-attend larger topical sessions but not small group sessions, but in panel (a), participants with some prior knowledge ($K_0 \leq 4$) may not co-attend and sessions.

* emmazajdela@u.northwestern.edu
† dmabrams@northwestern.edu