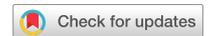




ARTICLE



<https://doi.org/10.1057/s41599-021-00869-9>

OPEN

Grand challenges and emergent modes of convergence science

Alexander M. Petersen¹[✉], Mohammed E. Ahmed²² & Ioannis Pavlidis²[✉]

To address complex problems, scholars are increasingly faced with challenges of integrating diverse domains. We analyzed the evolution of this convergence paradigm in the ecosystem of brain science, a research frontier that provides a contemporary testbed for evaluating two modes of cross-domain integration: (a) cross-disciplinary collaboration among experts from academic departments associated with disparate disciplines; and (b) cross-topic knowledge recombination across distinct subject areas. We show that research involving both modes features a 16% citation premium relative to a mono-domain baseline. We further show that the cross-disciplinary mode is essential for integrating across large epistemic distances. Yet we find research utilizing cross-topic exploration alone—a convergence shortcut—to be growing in prevalence at roughly 3% per year, significantly outpacing the more essential cross-disciplinary convergence mode. By measuring shifts in the prevalence and impact of different convergence modes in the 5-year intervals up to and after 2013, we find that shortcut patterns may relate to competitive pressures associated with Human Brain funding initiatives launched that year. Without policy adjustments, flagship funding programs may unintentionally incentivize suboptimal integration patterns, thereby undercutting convergence science's potential in tackling grand challenges.

¹Department of Management of Complex Systems, Ernest and Julio Gallo Management Program, School of Engineering, University of California, Merced, CA, USA. ²Computational Physiology Laboratory, University of Houston, Houston, TX, USA. ✉email: petersen.xander@gmail.com; ipavlidis@uh.edu

Introduction

The history of scientific development is characterized by a pattern of convergence-divergence cycles (Roco, 2013). In divergence, conflicting social forces can induce fragmentation (Baliotti et al., 2015), and disciplinary spin-offs occur as new techniques, tools, applications, and specialized sub-theories spawn from efforts to understand emergent complexity (Bonaccorsi, 2008). In classically construed convergence, originally distinct disciplines synergistically interact to accelerate breakthrough discovery in complex problems (National Research Council, 2014), thereby representing an intrepid form of interdisciplinarity in terms of the number, distance, and novelty of the disciplinary configurations entailed (Nissani, 1995).

Consequently, the cognitive and social dimensions of interdisciplinarity (Wagner, 2011) are also augmented in classic convergence. In addition to the cognitive challenges arising from integrating distinct knowledge domains and overcoming their different communication styles and ontologies, there are also bureaucratic and socio-political burdens associated with assembling and harnessing the expertise of scholars from different disciplines (Barry et al., 2008). Altogether, the widely documented tensions underlying interdisciplinarity are compounded in convergence science, owing to the intellectual and organizational challenges associated with the number of and distance between the disciplines being integrated (Bromham et al., 2016, Fealing, 2011, National Research Council, 2005). At the extreme, contemporary convergence meets large team science (Börner, 2010, Milojevic, 2014, Pavlidis et al., 2014, Wuchty et al., 2007), where collaboration on a grand scale across distinct academic cultures and units faces unique behavioral (Van Rijnsoever and Hessels, 2011), institutional, and cross-sectoral barriers (National Research Council, 2014).

Two early successful examples of classic convergence are worth mentioning to draw a comparative baseline. First, the Manhattan Project (MP), where physicists, chemists, and engineers successfully worked in the 1940s to control nuclear fission and produce the first atomic bomb, under a tightly run government program (Hughes and Hughes, 2003). A half-century later (the 1990s–2000s), the Human Genome Project (HGP) forged a multi-institutional bond integrating biologists and computer scientists, under an organizational model known as consortium science. In this model, teams of teams organize around a well-posed grand challenge (Helbing, 2012), with a common goal to share benefits equitably within and beyond institutional boundaries (Petersen, 2018). Within 10 years, the HGP led to the mapping and sequencing of the human genome, ushering civilization into the genomics era, with numerous other “omics” initiatives quickly following.

Brain science is presently supported by major funding programs that span the world over (Grillner, 2016). In late 2013, the United States launched the BRAIN Initiative® (Brain Research through Advancing Innovative Neurotechnologies), a public-private effort aimed at developing new experimental tools that will unlock the inner workings of brain circuits (Jorgenson, 2015). At the same time, the European Union launched the Human Brain Project (HBP), a 10-year funding program based on exascale computing approaches, which aims to build a collaborative infrastructure for advancing knowledge in the fields of neuroscience, brain medicine, and computing (Amunts, 2016). In 2014, Japan launched Brain Mapping by Integrated Neurotechnologies for Disease Studies (Brain/MINDS), a program to develop innovative technologies for elucidating primate neural circuit functions (Okano et al., 2015). China followed in 2016 with the China Brain Project (CBP), a 15-year program targeting the neural basis of human cognition (Poo, 2016). Canada (Jabalpurwala, 2016), South Korea (Jeong, 2016), and Australia

(Australian Brain Alliance Steering Committee, 2016) followed suit, launching their own brain programs in the late 2010s.

By nature and historical precedence, convergence tends to operate on the frontier of science. In the 2010s, brain science was declared the new research frontier (Quaglio, 2017) promising health and behavioral applications (Eyre, 2017). Intensification of brain research has been taking place against a backdrop of an increasingly globalized, interconnected, and online scientific commons. This stands in sharp contrast to the nationally unipolar and offline backdrop of the MP and even the HGP. Moreover, the brain funding programs were designed to act as behavioral incentives in a scientific marketplace aimed at efficiently matching funding, scholars, knowledge, and target problems (Stephan, 2012). However, despite being oriented around the compelling structure-function brain problem, there were few guidelines on how to configure scholarly expertise to address the brain challenge. As such, these characteristics render brain research a “live experiment” in the international evolution of the convergence paradigm.

Accordingly, here we apply data-driven methods to reconstruct the brain science ecosystem and to thereby capture the contemporary “pulse” of convergence, explored through a progressive series of research questions regarding its prevalence, anatomy, and scientific impact. Given the wave of human brain science (HBS) funding programs, we further analyze how the trajectory of HBS convergence has been impacted by the ramp-up of funding initiatives around the world. Our approach accounts for both the cognitive and social dimensions of cross-domain integration by measuring both the topical diversity of the article’s core knowledge and the disciplinary diversity of its research team. We leverage two comprehensive taxonomies to distinguish and quantify mono-domain versus cross-domain activity: (a) To classify research topics, we use the *Medical Subject Heading (MeSH)* ontology developed by the US National Library of Medicine (2021). MeSH imposes content standardization that facilitates our cross-temporal analysis in the PubMed® database. Specifically, MeSH features a high degree of concept orthogonality thanks to its tree structure, and its terms being assigned to articles by experts. MeSH’s relative orthogonality and objectivity offer significant classification advantages with respect to taxonomies based on author keywords, which suffer from extensive redundancies and bias. (b) To classify departmental affiliations, we use the *Classification of Instructional Programs (CIP)* scheme developed by the National Center for Education Statistics (2021). CIP was devised to serve accreditation and historical tracking of degree-granting programs in North America, and thus it is well-suited to our purpose. Accordingly, we classify HBS research into four types defined by the {mono–, cross–} × {topic, discipline} domain combinatorics.

In a highly competitive and open science system with multiple degrees of freedom, our motivating hypothesis is that more than one operational cross-domain integration mode is likely to emerge among scholars. With this in mind, we identify five research questions (RQ) addressed in each figure of the manuscript. The first (RQ1) regards how to define convergence, which we address by developing a typological framework, one that is generalizable to other frontiers of biomedical science and is also relevant to the evaluation of HBS funding projects around the world. The second (RQ2) regards the status and impact of brain science convergence: have HBS interfaces have developed to the point of sustaining fruitful cross-disciplinary knowledge exchange? Does the increasing prevalence of teams adopting convergent approaches correlate with higher scientific impact research? RQ3 addresses whether convergence is evenly distributed across HBS subdomains? Furthermore, what empirical

combinations of distinct subject areas (knowledge) and disciplinary expertise (people) are overrepresented in convergent research? RQ4 follows by seeking to identify whether convergence is evenly distributed over time and geographic region? Finally, RQ5 asks: Does the propensity and/or citation impact of convergence science, depend on the convergence mode? To address this question, we implement hierarchical regression models that differentiate between three convergence modes: research involving cross-subject-area exploration, cross-disciplinary collaboration, or both. Given the trend-setting role of major funding initiatives, we hypothesize that the ramp-up of HBS flagship programs correlates with shifts in the prevalence and relative impact of research adopting these different convergence modes.

Our results bear timely science policy implications. Given the contemporary emphasis on accelerating breakthrough discovery (Helbing, 2012) through strategic team configurations (Börner, 2010), convergence science originators called for cross-disciplinary approaches integrating distant disciplines (National Research Council, 2014). In what follows, we test if this vision has been realized by counting the prevalence and impact of research articles belonging to any of four {mono-, cross-} × {topic, discipline} integration types. Among the said types, classic or full convergence corresponds to research endeavors featuring both cross-topic integration and cross-disciplinary collaboration; it stands in contrast to the other three types: mono-topic/mono-discipline; mono-topic/cross-discipline; and cross-topic/mono-discipline. Our analysis reveals that recently HBS teams tend to integrate diverse topics without necessarily integrating appropriate disciplinary expertise—we identify this cross-topic/mono-discipline approach as a *convergence shortcut*.

Theoretical background

To locate convergence science within the spectrum of interdisciplinarity (Barry et al., 2008), we draw upon the classic descriptions by Nissani (1995), whereby interdisciplinary knowledge captures the "distinctive components of two or more disciplines"; and interdisciplinary research is "combining distinctive components of two or more disciplines while searching or creating new knowledge". Scholars can thus explore intermediate or hybrid subject areas by way of two modes: by autodidactic expansive learning outside of their original domain (Engestrom and Sannino, 2010); and/or by collaboration with scholars with expertise in a different discipline (i.e., cross-disciplinary collaboration). Such interdisciplinary activities take on the characteristics of full-fledged convergence science when the integrated domains have evolved from distant epistemic origins. For example, brain science features the convergence of physiology, medicine, and neuroscience, catalyzed by techno-informatics and computing capabilities (Yang et al., 2021). Another example is the integration of disciplines characterized by either "soft" or "hard" methodologies (Pedersen, 2016).

For a review of the rationale and socio-politics propelling interdisciplinary studies and the recasting of the issue as a spectrum rather than a single form, see Barry et al. (2008); whereas for a review of the metrics for measuring interdisciplinarity see Wagner (2011), which also documents the cavalier use of interdisciplinary terminology. Here, we distinguish between two modes of interdisciplinarity—one cognitive and the other social—and in each case, we make use of well-established ontologies in order to avoid operational ambiguity. In the cognitive mode, we use the MeSH keyword system to classify topics comprising the biomedical knowledge network (Yang et al., 2021). In the social mode, we tabulate cross-disciplinary collaborations per an institutional perspective, identifying disciplines according to the primary departmental affiliation of each team member.

This work contributes to several other literature streams, including the quantitative analysis of recombinant search and innovation (Fleming, 2001, Fleming and Sorenson, 2004, Youn, 2015); cross-domain integration of expertise (Fleming, 2004, Leahey and Moody, 2014, Petersen, 2018, Petersen et al., 2019); cross-disciplinarity as a strategic team configuration (Cummings and Kiesler, 2005, Petersen, 2018) facilitated by division of labor across teams of specialists and generalists (Haeussler and Sauermann, 2020, Melero and Palomeras, 2015, Rotolo and Messeni Petruzzelli, 2013, Teodoridis, 2018); and science of science Fortunato (2018) and science policy (Fealing, 2011) evaluation of convergence science (National Research Council, 2005, 2014, Roco, 2013).

Efficient long-range epistemic exploration facilitated by multi-disciplinary teams is a defining value proposition of classic convergence (National Research Council, 2014) and increases the likelihood of large team science producing high-impact research (Wuchty et al., 2007). Hence, the emergence and densification of cross-domain interfaces are likely to increase the potential for breakthrough discovery by catalyzing recombinant innovation (Fleming, 2001). Following the triple-helix model of medical innovation (Petersen et al., 2016), recombinant innovation manifests from integrating expertise around the three dimensions of supply, demand, and technological capabilities, which for HBS are: (i) the biology domain that supplies a theoretical understanding of the anatomical structure-physiological function relation, (ii) the health domain that demands effective science-based solutions, and (iii) the techno-informatics domain which develops scalable products, processes, and services to facilitate matching supply from (i) with demand from (ii) (Yang et al., 2021).

In order to overcome the challenges of selecting new strategies from the vast number of possible combinations, prior research finds that innovators are more likely to succeed by way of exploiting their own local expertise (Fleming, 2001) rather than individually exploring distant configurations by way of internal expansive learning (Engestrom and Sannino, 2010). Extending this argument, exploration at uncharted multidisciplinary interfaces is likely to be more successful when integrating knowledge across a team of experts from different domains, thereby hedging against recombinant uncertainty underlying the exploration process (Fleming, 2004).

A complementary argument for classic convergence derives from the advantage of diversity to harness collective intelligence for identifying successful hybrid strategies (Page, 2008), while also avoiding misinterpretations and incomplete ontologies (Barry et al., 2008, Linkov et al., 2014). Recent work also provides support for the competitive advantage of diversity stemming from cross-border mobility (Petersen, 2018).

Methods

Data collection and notation. We integrated publication and author data from Scopus, PubMed, and the Scholar Plot web app (Majeti, 2020) (see Fig. 1 and Supplementary Information (SI) Appendix S1 for detailed description). In total, our data sample spans 1945–2018 and consists of 655,386 publications derived from 9,121 distinct Scopus Author profiles, to which we apply the following variable definitions and subscript conventions to capture both article-level and scholar-level information. At the article level, subscript p indicates publication-level information such as publication year, y_p ; the number of coauthors, k_p ; and the number of keywords, w_p . Regarding the temporal dimension, a superscript, $>$ (respectively, $<$) indicates data belonging to the 5-year "post" period 2014–2018 (5-year "pre" period 2009–2013), while $N(t)$ represents the total number of articles published in year t .

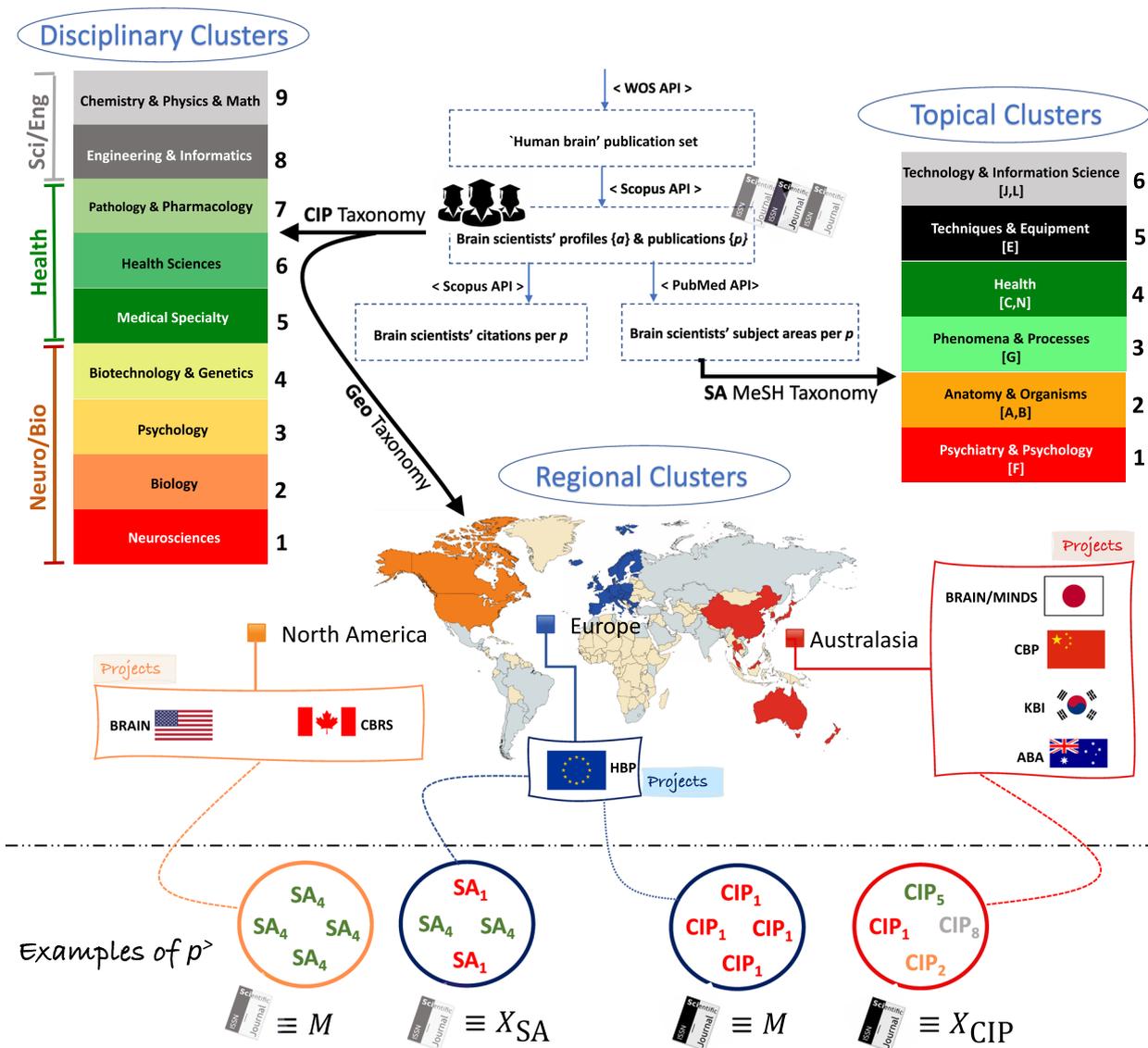


Fig. 1 Data collection and classification schemes. The upper part of the figure shows the data generation mechanism along with the resulting topical (SA) and disciplinary (CIP) clusters. The middle part of the figure shows on the world map regional clusters pertaining to three large HBS funding initiatives—North America (NA), Europe (EU), and Australasia (AA). The lower part of the figure shows an example of how all three categorizations are operationalized for analytic purposes. Circles represent four research articles with authorship from distinct regions. The articles feature different keyword (SA) or disciplinary (CIP) category mixtures assigned one of two diversity measures: mono-domain (M) and cross-domain (X).

Regarding proxies for scientific impact, we obtained the number of citations $c_{p,t}$ from Scopus, which are counted through late 2019. Since nominal citation counts suffer from systematic temporal bias, we use a normalized citation measure, denoted by z_p . Regarding author-level information, we use the index a —e.g., we denote the academic age measured in years since a scholar’s first publication by $\tau_{a,p}$.

To address RQ1, we first classified research according to three category systems indicative of topical, disciplinary, and regional clusters.

Topical classification using MeSH. The first category system captures research topic clusters grouped into Subject Areas (SAs). To operationalize SA categories we leverage the system of roughly 30,000 *Medical Subject Heading (MeSH)* article-level “keywords” that are organized in a comprehensive and standardized hierarchical ontology maintained by the US National Library of Medicine (2021). Each MeSH descriptor has a tree number that

identifies its location within one of 16 broad categorical branches within a prescribed knowledge network. Here we merged 9 of the science-oriented MeSH branches (A,B,C,E,F,G,J,L,N) into 6 Subject Area (SA) clusters (see Fig. 1). Figure S1 shows the 50 most prominent MeSH descriptors for each SA cluster. See Yang et al. (2021) for a complementary analysis of the entire PubMed database, in particular regarding the utility of MeSH for quantifying cross-SA convergence.

Hence, we take the set of MeSH for each p denoted by \vec{W}_p , and map these MeSH, using the generic operator O_{SA} , into their corresponding MeSH branch, thereby yielding a count vector with six elements: $O_{SA}(\vec{W}_p) = \vec{SA}_p$. Each SA component corresponds to one or two top-level MeSH categories that are defined by PubMed (indicated by the letters in brackets): (1) Psychiatry and Psychology [F], (2) Anatomy and Organisms [A, B], (3) Phenomena and Processes [G], (4) Health [C,N], (5) Techniques and Equipment [E], and (6) Technology and Information Science [J,L]. Notably, regarding the *structure-*

function problem that is a fundamental focus in much of biomedical science, SA category (2) represents the domain of structure while (3) represents function. The variable $N_{SA,p}$ counts the total number of SA categories present in a given article, with min value 1 and max value 6; 72% of articles have two or more SA; the mean (median) SA_p is 2.1 (2), with standard deviation 0.97.

Disciplinary classification using CIP. The second taxonomy identifies disciplinary clusters determined by author departmental affiliation, which we categorized according to *Classification of Instructional Programs* (CIP) codes maintained by the National Center for Education Statistics (2021). Article-level CIP category counts are represented by \overrightarrow{CIP}_p , and comprised of the following 9 elements: (1) Neurosciences, (2) Biology, (3) Psychology, (4) Biotech and Genetics, (5) Medical Specialty, (6) Health Sciences, (7) Pathology and Pharmacology, (8) Engineering and Informatics, and (9) Chemistry, Physics and Math. The variable $N_{CIP,p}$ counts the total number of CIP categories present in a given article, with min value 1 and max value 9; see SI Appendix S1 offers more details.

We obtained host department information from each scholar’s Scopus Profile. Based upon this information provided in the profile description, and in some cases using additional web search and data contained in the Scholar Plot web app (Majeti, 2020), we manually annotated each scholar’s home department name with its CIP code (National Center for Education Statistics, 2021). We then merged these CIP codes into 9 broad clusters and three super-clusters (Neuro/Biology, Health, and Science and Engineering, as indicated in Fig. 1); for a list of constituent CIP codes for each cluster see Fig. S1C. Analogous to the notation for assigning \overrightarrow{SA}_p , we take the set of authors for each p denoted by \vec{A}_p , and map their individual departmental affiliations to the corresponding CIP cluster (represented by the operator O_{CIP}), yielding a count vector with nine elements: $O_{CIP}(\vec{A}_p) = \overrightarrow{CIP}_p$.

Regional classification. The third taxonomy captures the broad geographic scope of each publication’s research team determined by each Scopus author’s affiliation location. We represent regions using a count vector \vec{R}_p , which has 4 elements representing North America, Europe, Australasia, and the Rest of the World.

Interdisciplinary classification of articles as mono-domain or cross-domain. We distinguish between the cognitive and social interdisciplinary dimensions (Wagner, 2011) in the combined feature vector $\vec{F}_p \equiv \{\overrightarrow{SA}_p, \overrightarrow{CIP}_p\}$. As indicated in Fig. 1, based upon the distribution of types tabulated as counts across vector elements, an article is either cross-domain, representing a diverse mixture of types denoted by X ; or in the case that there is just a single type, the article is mono-domain, denoted by M . We use a generic operator notation to specify how articles are classified as X or M . The objective criteria of the feature operator O is specified by its subscript: for example $O_{SA}(\vec{F}_p)$ yields one of two values: X_{SA} or M ; similarly, $O_{CIP}(\vec{F}_p) = X_{CIP}$ or M . Note that all scholars map onto a single CIP, hence solo-authored research articles are by definition classified by O_{CIP} as M . While we acknowledge that is possible for a scholar to have significant expertise in two or more domains, we do not account for this duplicity, as it is likely to occur at the margins; hence, the home department CIP represents the scholar’s principle domain of expertise. We also classify articles featuring both X_{SA} and X_{CIP} as $O_{SA\&CIP}(\vec{F}_p) = X_{SA\&CIP}$ (and otherwise M).

See Fig. S1 for the composition of SA and CIP clusters, and SI Appendix S1 for additional description on how these classification systems are constructed. Figure S2 (Fig. S3) shows the frequency of each SA (CIP) category and the pairwise frequency of all $\{SA, SA\}(\{CIP, CIP\})$ combinations over the 10-year period centered on 2013, along with their relative changes after 2013; see SI Appendix S2–S3 for discussion of the relevant changes in SA and CIP categories after 2013.

Measuring cross-domain diversity. To complement these categorical measures, we also developed a scalar measure of an article’s cross-domain diversity. By way of example, consider the vector \overrightarrow{SA}_p (or \overrightarrow{CIP}_p) which tallies the SA (or CIP counts) for a given article p published in year t . We apply the outer tensor product $\overrightarrow{SA}_p \otimes \overrightarrow{SA}_p$ (or $\overrightarrow{CIP}_p \otimes \overrightarrow{CIP}_p$) to represent all pairwise co-occurrences in a weighted matrix $D_p(\vec{v}_p)$ (where \vec{v}_p represents a generic category vector; see SI Appendix S4 for examples of the outer tensor product). The sum of elements in this co-occurrence matrix is normalized to unity so that each $D_p(\vec{v}_p)$ contributes equally to averages computed across all articles from a given year or period. Since the off-diagonal elements represent cross-domain combinations, their relative weight given by $f_{D,p} = 1 - \text{Tr}(D_p) \in [0, 1)$ is a straightforward Blau-like measure of variation and disparity (Harrison and Klein, 2007).

In more detail, this measure of cross-domain diversity is defined according to categorical co-occurrence within individual research articles. Each article p has a count vector \vec{v}_p ; for discipline categories $\vec{v}_p \equiv \overrightarrow{CIP}_p$ and for topic categories $\vec{v}_p \equiv \overrightarrow{SA}_p$. We then measure article co-occurrence levels by way of the normalized outer-product

$$D_p(\vec{v}_p) \equiv \frac{U(\vec{v}_p \otimes \vec{v}_p)}{\|U(\vec{v}_p \otimes \vec{v}_p)\|}, \tag{1}$$

where \otimes is the outer tensor product, $U(\mathbf{G})$ is an operator yielding the upper-diagonal elements of the matrix \mathbf{G} (i.e., representing the undirected co-occurrence network among the categorical elements). In essence, $D_p(\vec{v}_p)$ captures a weighted combination of all category pairs. The resulting matrix represents dyadic combinations of categories as opposed to permutations (i.e., capturing the subtle difference between an undirected and directed network). While we did not explore it further, this matrix formulation may also give rise to higher-order measures of diversity associated with the eigenvalues of the outer-product matrix. The notation $\|...\|$ indicates the matrix normalization is implemented by summing all matrix elements. The objective of this normalization scheme is to control for the variation in \vec{v}_p in a systematic way. As such, this co-occurrence is an article-level measure of diversity that controls for variations in the total number of categories and fundamentally different count statistics for the CIP and SA category systems. Consequently, totaling $D_p(\vec{v}_p)$ across articles from a given publication year yields the total number of articles published in a given year, $\sum_{p|y_p \in t} \|D_{p,t}\| = N(t)$.

Following this standardization, $f_{D,p}$ is bounded in the range $[0, 1)$ and so the characteristic article diversity is well-defined and calculated as the average annual value $\langle f_D(t) \rangle$. In simple terms, articles featuring a single category have $f_{D,p} = 0$, whereas articles featuring multiple categories have $f_{D,p} > 0$. While the result of this approach is nearly identical to the Blau index (corresponding to $1 - \overrightarrow{SA}_p \cdot \overrightarrow{SA}_p / |SA_p|^2$, also referred to as the Gini-Simpson index),

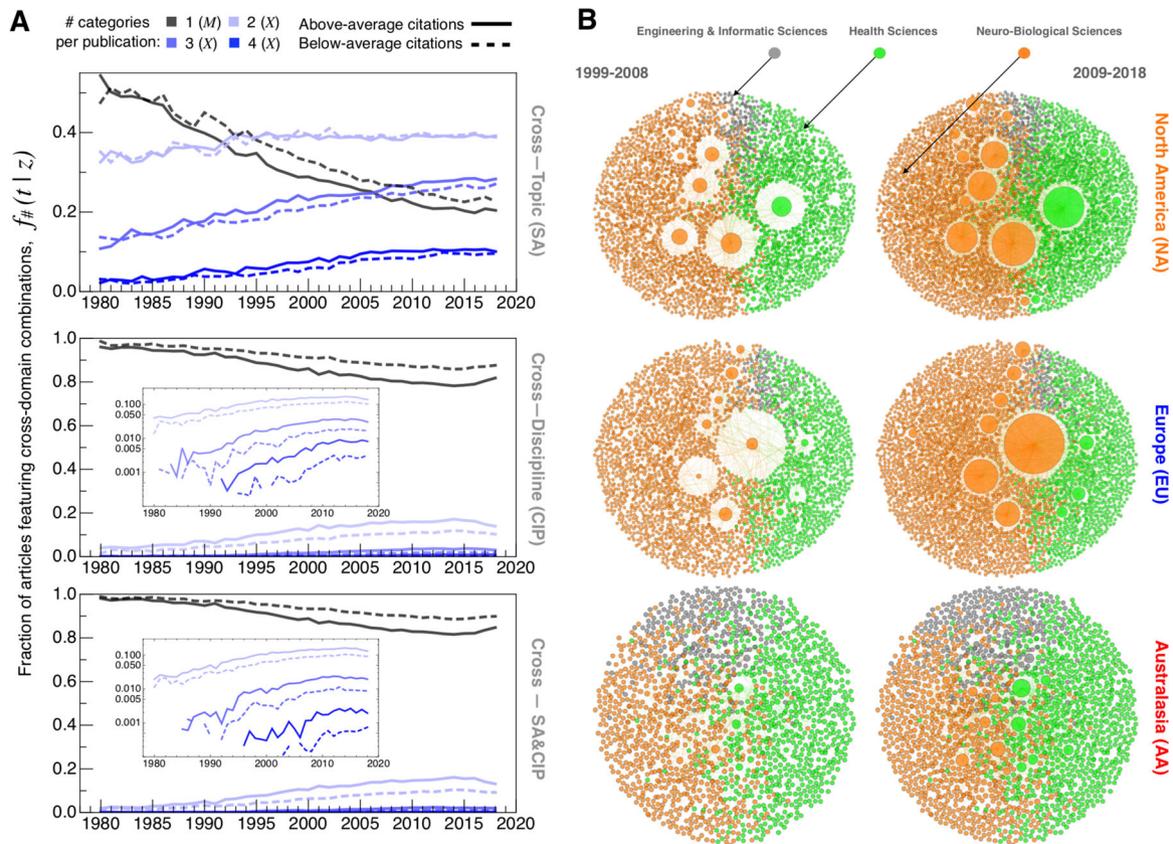


Fig. 2 Trends in cross-domain scholarship in Human Brain Science. **A** Fraction $f_{\#}(t|z)$ of articles published each year t that feature a particular number ($\#$) of categories. Articles are split into an above-average citation subset ($z_p > 0$) and below-average citation subset ($z_p < 0$). Upper panel: Articles categorized by SA. Middle panel: Articles categorized by CIP; subpanel shows data on the logarithmic y-axis; Lower panel: Articles categorized by both SA and CIP. Distinguishing frequencies by citation group indicates higher levels of cross-domain combinations among research articles with higher scientific impact—for both SA and CIP. However, cross-domain activity levels are visibly higher for SA than for CIP, indicating higher barriers to boundary-crossing arising from mixing different scholar expertise. **B** Snapshots of the collaboration network at 10-year intervals indicating researcher population sizes by region, and the densification of convergence science at cross-disciplinary interfaces. Nodes (researchers) are sized according to the number of collaborators (link degree) within each time window.

$f_{D,p}$ is motivated by way of dyadic co-occurrence rather than the standard formulation motivated around repeated sampling.

Results

Descriptive analysis

Increasing prevalence of cross-domain science. With the continuing shift towards large team science (Milojevic, 2014, Pavlidis et al., 2014, Petersen et al., 2014, Wuchty et al., 2007), one might expect a similar shift in the multiplicity of domains spanned by modern research teams—but to what degree? Figure 2A addresses RQ2 by showing the frequencies of mono-domain (M) research articles versus cross-domain articles (X) in our HBS sample. Articles were separated into above-average and below-average citation impact (z) for each year (t), and within each of these two subsets, we calculated the fraction $f_{\#}(t|z)$ of articles containing combinations across $\# = 1, 2, 3$, and 4 categories. The fraction of mono-domain articles are trending downward, which we observe for both research topics (SA) and authors' disciplinary affiliations (CIP). The decline is much steeper for SA than for CIP. Correspondingly, cross-domain articles have become increasingly prevalent, in particular for SA. For both SA and CIP the two-category mixtures dominate the three-category and four-category mixtures in frequency. Accordingly, in the sections that follow we do not distinguish between cross-domain articles with different $\#$.

As a first indication of the comparative advantage associated with X , we observe a robust inequality $f_{\#}(t|z > 0) > f_{\#}(t|z < 0)$ for

cross-domain research ($\# \geq 2$), meaning a higher frequency of cross-domain combinations is observed among articles with higher impact. Contrariwise, in the case of mono-domain research the opposite inequality persists, $f_1(t|z > 0) < f_1(t|z < 0)$. Taking into consideration temporal trends, these robust patterns indicate a faster depletion of impactful mono-domain articles, coincident with an increased prevalence of impactful research drawing upon integrative recombinant innovation.

Recombinant innovation at the convergence nexus. Comprehensive analysis of biomedical science indicates that convergence has largely been mediated around the integration of modern techninformatics capabilities (Yang et al., 2021). Classic convergence is typified by a high degree of diversity fostered through multi-party cognitive and institutional complementarity—two characteristics of the Bonaccorsi (2008) model of "New Science". Yet within any frontier domain, in particular HBS, the question remains as to the development of a functional nexus that sustains and possibly even accelerates high-impact discovery through the rich cross-disciplinary exchange of new knowledge and best practices. The robust inequality $f_{\#}(t|z > 0) > f_{\#}(t|z < 0)$ provides support at the aggregate level but does not lend any structural evidence.

To further address RQ2, Fig. 2B illustrates the composition of the HBS convergence nexus, showing the integration of cross-disciplinary expertise manifesting in collaborations across three broad yet distinct biomedical domains. Shown are the

populations of HBS researchers by region, represented as collaboration networks compared over two non-overlapping 10-year intervals to indicate dynamics. Each node represents a researcher, colored according to three disciplinary CIP super-clusters: (i) neuro-biological sciences (corresponding to CIP 1–4), (ii) health sciences (CIP 5–7), and (iii) engineering and information sciences (CIP 8–9). Node locations are fixed to facilitate visual representation of network densification. Inter-regional and cross-regional comparison alludes to the emergence and densification of cross-domain interfaces (see also Fig. S4). Because the network layout is determined by the underlying structure, there is a high degree of clustering by node color, emphasizing both the relative sizes of the subpopulations that are well-balanced across region and time and also the convergent interfaces where cross-disciplinary collaboration and knowledge exchange are likely to catalyze. As such, these communities of expertise conjure the image of a Pólya urn, whereby successful configurations reinforce the adoption of similar future configurations.

The links that span disciplinary boundaries are fundamental conduits across which scientists’ strategic affinity for exploration (Foster et al., 2015, Rotolo and Messeni Petruzzelli, 2013) is effected via cross-disciplinary collaboration that brings “together distinctive components of two or more disciplines” (Nissani, 1995, Petersen, 2018). Our analysis of cross-disciplinary collaboration indicates that the fraction of articles featuring convergent collaboration has continued to grow over the last two decades (see Fig. S4). In what follows we further distinguish between integration across neighboring (Leahey and Moody, 2014) and distant domains, with the latter appropriately representing classic convergence (National Research Council, 2005, 2014, Roco, 2013).

Cross-domain convergence of expertise (CIP) and knowledge (SA). In an institutional context, team assembly should be optimized by strategically matching scholarly expertise and research topics to address the demands of a particular challenge. The role of expertise is increasingly important (Petersen et al., 2019) when the phenomena are complex and the underlying problems cannot be represented as a well-posed or closed-form problem, e.g., those involving human behavior (Oreskes, 2021) or climate change (Arroyave, 2021, Bonaccorsi, 2008). Hence, with 9 different disciplinary (CIP) domains historically faced with a variety of challenges, RQ3 addresses to what degree these domains differ in terms of their composition of targeted SA. Fig. 3A illustrates the evolution of topical diversity within and across each CIP cluster, revealing several common patterns. First, nearly all domains show a reduction in research pertaining to structure (SA 2), with the exception of Biotechnology and Genetics, which had a balanced structure-function composition from the outset. As such, this domain features a steady mix among SA 2–5, while being an early adopter of techno-informatics concepts and methods (SA 6). This stable mirroring of the innovation triple-helix (Petersen et al., 2016) may explain to some degree the longstanding success of the genomics revolution, where the core disciplines of biology and computing were primed for a fruitful union (Petersen, 2018). Other HBS disciplinary clusters are also integrating techno-informatic capabilities, reflecting a widespread pattern observed across all of biomedical science (Yang et al., 2021).

Which CIP-SA combinations are overrepresented in boundary-crossing HBS research? Inasmuch as mono-domain articles identify the topical boundary closely associated with individual disciplines, cross-domain articles are useful for identifying otherwise obscured boundaries that call for both X_{CIP} and X_{SA} in combination. We identified these novels CIP-SA relations by collecting articles that are purely mono-domain for both CIP and

SA (i.e., those with $O_{CIP}(\vec{F}_p) = O_{SA}(\vec{F}_p) = M$) and a complementary non-overlapping subset of articles that are simultaneously cross-domain for both CIP and SA (i.e., $O_{SA\&CIP}(\vec{F}_p) = X_{SA\&CIP}$).

Starting with mono-domain articles, we identified the SA that is most frequently associated with each CIP category. Formally, this amounts to calculating the bi-partite network between CIP and SA, denoted by $M_{CIP} \rightleftharpoons M_{SA}$. These CIP-SA associations are calculated by averaging the \vec{SA}_p for mono-domain articles from each CIP category, given by $\langle \vec{SA} \rangle_{CIP}$. Figure 3B highlights the most prominent CIP-SA links (see SI Appendix S5 for more details). Likewise, we also calculated the bi-partite network $X_{CIP} \rightleftharpoons X_{SA}$ using the subset of $X_{SA\&CIP}$ articles.

To identify the cross-domain frontier, we calculate the network difference $\Delta_{XM} \equiv (X_{CIP} \rightleftharpoons X_{SA}) - (M_{CIP} \rightleftharpoons M_{SA})$ and plot the links with positive values—i.e., CIP-SA links that are over-represented in $X_{CIP} \rightleftharpoons X_{SA}$ relative to $M_{CIP} \rightleftharpoons M_{SA}$. Results identify SA that is reached by way of cross-disciplinary teams. SA 3 (Phenomena and Processes), representing the brain function problem, stands out as a potent convergence nexus attainable by highly convergent teams combining disciplines 1, 2, 4, and 9.

A related key insight concerns the relative increase in SA integration achieved by increased CIP diversity. Figure S5 compares the average number of SA integrated by teams with varying numbers of distinct CIP, $N_{CIP,p}$. On average, mono-disciplinary teams ($N_{CIP,p} = 1$) span 2.2 SA, whereas teams with $N_{CIP,p} = 3$ spans 19% more SA, confirming that cross-disciplinary configurations are more effective in achieving research breadth.

Trends in cross-domain activity. To address the temporal and geographic parity associated with RQ4, we define three types of cross-domain configurations—*Broad*, *Neighboring*, and *Distant*—defined according to a particular combination of SA and CIP categories featured by a given article. Based upon stated mission and vision statements, the BRAIN initiative (NA) aligns with *Neighboring* and the Human Brain Project (EU) aligns more closely with *Distant* configurations.

Broad is the most generic cross-domain configuration, based upon combinations of any two or more SA (or CIP) categories, and represented by our operator notation as $O_{SA}(\vec{F}_p) = X_{SA}$ (or $O_{CIP}(\vec{F}_p) = X_{CIP}$, respectively). *Neighboring* is the X configuration that captures the neuro-psychological \leftrightarrow bio-medical interface representing articles that contain MeSH from SA (1) and also from SA (2, 3, or 4), represented summarily as $[1] \times [2 - 4]$; and for CIP, combinations containing CIP (1 or 3) and (2, 4, 5, 6, or 7), represented as $[1, 3] \times [2, 4 - 7]$. Articles featuring these configurations are represented using our operator notation as $O_{Neighboring, SA}(\vec{F}_p) = X_{Neighboring, SA}$, $O_{Neighboring, CIP}(\vec{F}_p) = X_{Neighboring, CIP}$, or $O_{Neighboring, SA\&CIP}(\vec{F}_p) = X_{Neighboring, SA\&CIP}$; alternatively, articles not containing the specific category combinations are represented by M .

Distant is the X configuration that captures the neuro-psychomedical \leftrightarrow techno-informatic interface. The specific set of category combinations representing this configuration are SA $[1-4] \times [5,6]$; and for CIP, $[1,3,5] \times [4,8]$; as above, articles featuring (or not featuring) these categorical combinations are represented by $X_{Distant, SA}$ (otherwise, M), $X_{Distant, CIP}$ (otherwise, M), $X_{Distant, SA\&CIP}$ (otherwise, M). By way of example, the bottom of Figure 1 illustrates an article combining SA 1 and 4, which is thereby classified as both X_{SA} and $X_{Neighboring, SA}$; and, an article featuring CIP 1, 3, 5, 8, which is thereby both X_{CIP} and $X_{Distant, CIP}$.

To complement these categorical variables, we also developed a Blau-like measure of cross-domain diversity, given by $f_{D,p}$ (see

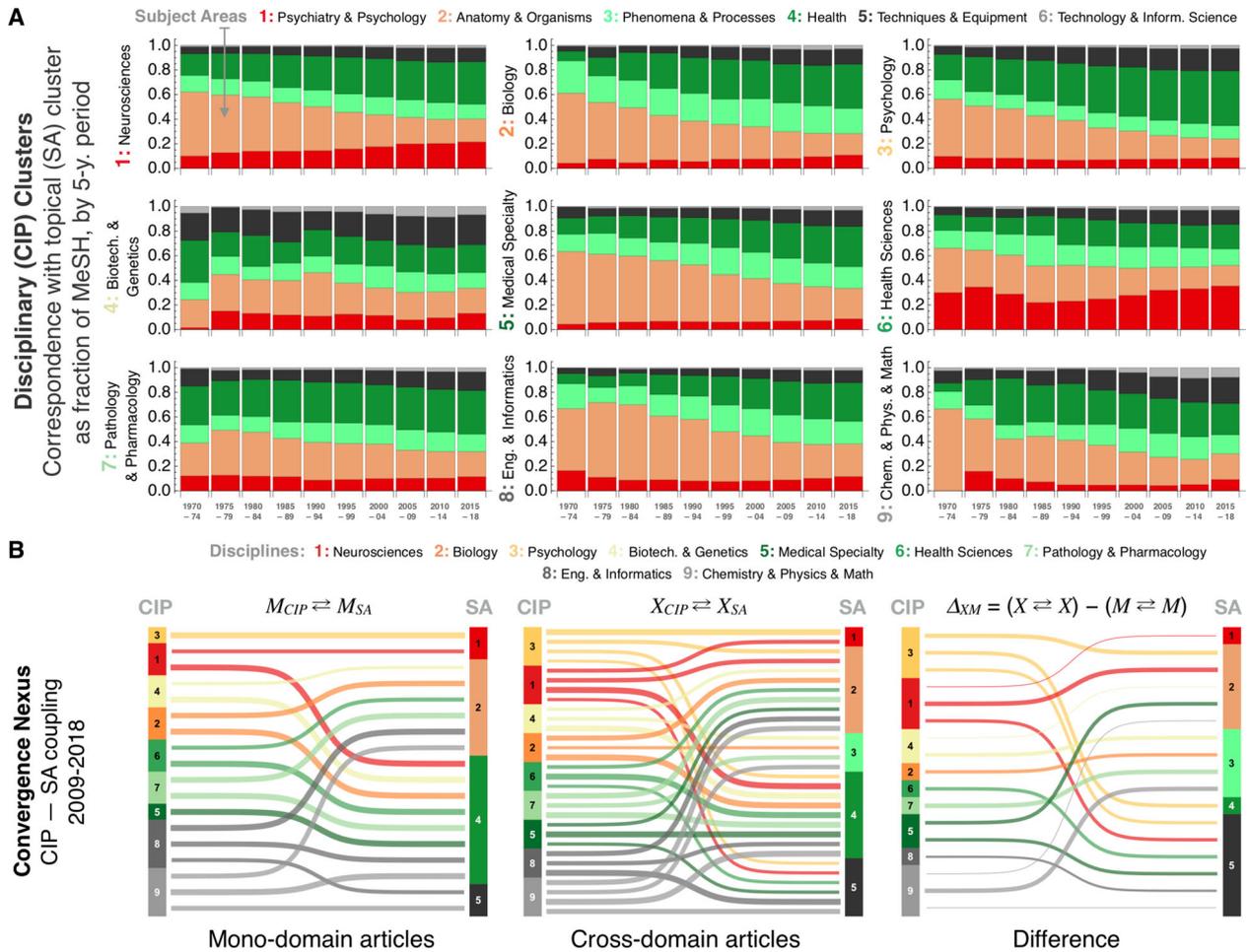


Fig. 3 Evolution of SA boundary-crossing within and across disciplinary clusters. **A** SA composition of HBS research within disciplinary (CIP) clusters. Each subpanel represents articles published by researchers from a given CIP cluster, showing the fraction of article-level MeSH belonging to each SA, shown over 5-year intervals across the period 1970–2018. The increasing prominence of SA 5 and 6 in nearly all disciplines, indicates the critical role of informatics capabilities in facilitating biomedical convergence (Yang et al., 2021). **B** Empirical CIP-SA association networks were calculated for non-overlapping sets of mono-domain ($M_{CIP} \rightleftharpoons M_{SA}$) and cross-domain ($X_{CIP} \rightleftharpoons X_{SA}$) articles and based upon the *Broad* configuration. The difference between these two bi-partite networks (Δ_{XM}) indicates the emergent research channels that are facilitated by simultaneous X_{CIP} and X_{SA} boundary crossing—in particular integrating SA 2 with 3 (i.e., the structure-function nexus) facilitated by teams combining disciplines 1, 2, 4, and 9.

“Measuring cross-domain diversity”). Figure 4 shows the trends in mean diversity ($f_D(t)$) for the *Broad*, *Neighboring*, and *Distant* configurations. For each configuration, we provide a schematic motif illustrating the combinations measured by $D_p(\vec{v}_p)$, with diagonal components representing mono-domain articles (indicated by 1 on the matrix diagonal) and upper-diagonal elements capturing cross-domain combinations (indicated by X). Comparing SA and CIP, there are higher diversity levels for SA, in addition to a prominent upward trend. In terms of CIP, Fig. 4A indicates a decline in *Broad* diversity in recent years, with North America (NA) showing higher levels than Europe (EU) and Australasia (AA); these general patterns are also evident for *Neighboring* diversity—see Fig. 4B. *Distant* CIP diversity, shown in Fig. 4C, indicates a recent decline for AA and NA, with NA peaking around 2009; contrariwise, EU shows a steady increase consistent with the computational framing of the HBP.

In contradistinction, all three regions show a steady increase irrespective of configuration in the case of SA diversity, consistent with scholars integrating topics without integrating scholarly expertise. For both *Broad* and *Neighboring* configurations, NA and EU show remarkably similar levels of SA diversity above AA; however, in the case of *Neighboring*, AA appears to be catching

up quickly since 2010—see Fig. 4D, E. In the case of *Distant*, all regions show a steady increase that appears to be in lockstep for the entire period—see Fig. 4F. See Figs. S6–S7 and SI Appendix Text S6 for trends in SA and CIP diversity across additional configurations.

Quantitative model

Regression model—propensity for and impact of X. To address RQ5, we construct article-level and author-level panel data to facilitate measuring factors relating to SA and CIP diversity, as well as shifts related to the ramp-up of HBS flagship projects circa 2013 around the globe. To address these two outcomes, we modeled two dependent variables separately; Fig. S8 shows the distribution, and Fig. S9 shows the covariation matrix among the principal variables of interest. In the first model, the dependent variable is the propensity for cross-domain research (indicated by X; depending on the focus around topics, disciplines, or both, then X is specified by X_{SA} , X_{CIP} , or $X_{SA\&CIP}$). We use a Logit specification to model the likelihood $P(X)$. In the second model, the dependent variable is the article’s scientific impact, proxied by c_p . Building on previous efforts (Petersen, 2018, Petersen et al., 2018), we apply a logarithmic transform to c_p that facilitates

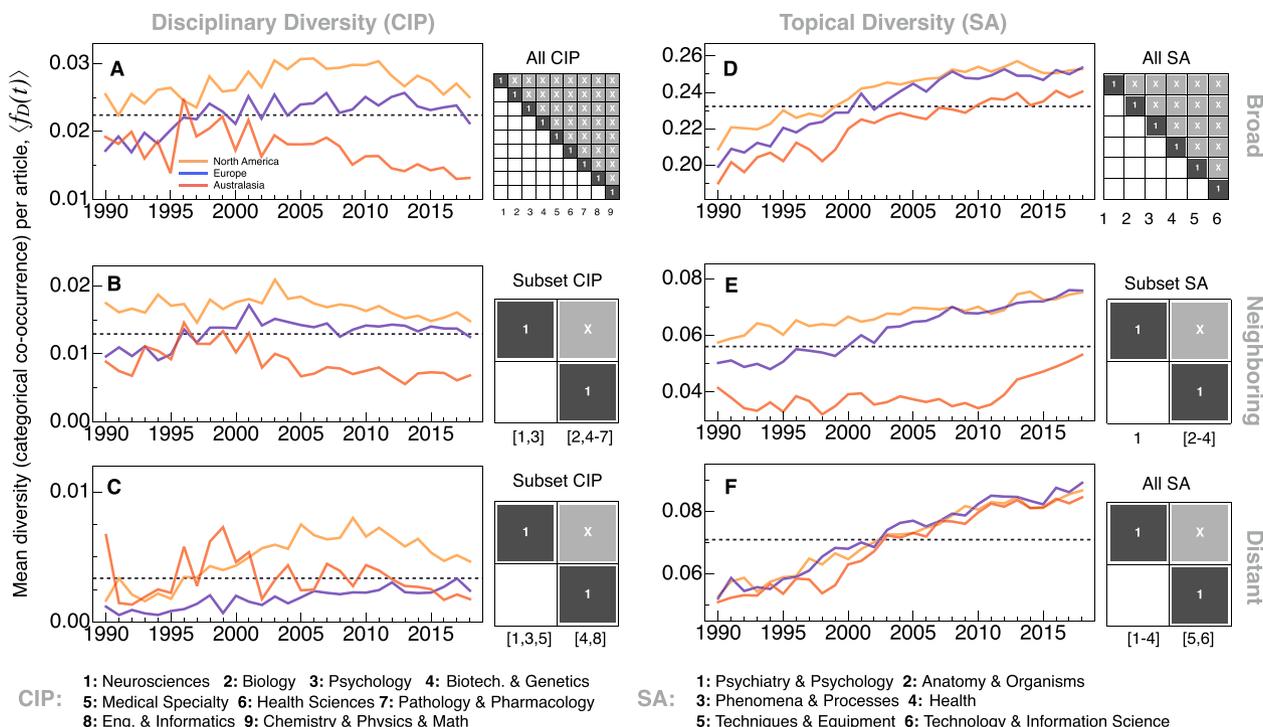


Fig. 4 Evolution of CIP and SA diversity in Human Brain Science research. A-F Each $\langle f_D(t) \rangle$ represents the average article diversity measured as categorical co-occurrence, by geographic region: North America (orange), Europe (blue), and Australasia (red). Each matrix motif indicates the set of CIP or SA categories used to compute D_p in Eq. (1); categories included in brackets are considered in the union. For example, panel (A) calculates $\langle f_{D,CIP}(t) \rangle$ across all 9 CIP categories; instead, panel (B) is based upon counts for two super-groups, the first consisting of the union of CIP counts for categories 1 and 3, and the second comprised of categories 2, 4, 5, 6 and 7. A, D Broad diversity is calculated using all categories considered as separate domains; (B, E) Neighboring represents the shorter-distance boundary across the neuro-psychological \leftrightarrow bio-medical interface; (C, F) Distant represents longer-distance convergence across the neuro-psycho-medical \leftrightarrow techno-informatic interface.

removing the time-dependent trend in the location and scale of the underlying log-normal citation distribution (Radicchi et al., 2008), as elaborated in what follows.

Normalization of citation impact. We normalized each Scopus citation count, $c_{p,t}$, by leveraging the well-known log-normal properties of citation distributions (Radicchi et al., 2008). To be specific, we grouped articles by publication year y_p , and removed the time-dependent trend in the location and scale of the underlying log-normal citation distribution. The normalized citation value is given by

$$z_p = (\ln(c_{p,t} + 1) - \mu_t) / \sigma_t, \tag{2}$$

where $\mu_t \equiv \langle \ln(c_t + 1) \rangle$ is the mean and $\sigma_t \equiv \sigma[\ln(c_t + 1)]$ is the standard deviation of the citation distribution for a given t ; we add 1 to $c_{p,t}$ to avoid the divergence of $\ln 0$ associated with uncited publications—a common method which does not alter the interpretation of results.

Figure S8G shows the probability distribution $P(z_p)$ calculated across all p within five-year non-overlapping time periods. The resulting normalized citation measure is well-fit by the Normal $N(0, 1)$ distribution, independent of t , and thus is a stationary measure across time. Publications with $z_p > 0$ are thus above the average log citation impact μ_t , and since they are measured in units of standard deviation σ_t , standard intuition and statistics of z -scores apply. The annual σ_t value is rather stable across time, with average and standard deviation $\langle \sigma \rangle \pm SD = 1.24 \pm 0.09$ over the 49-year period 1970–2018.

Model A: Quantifying the propensity for X and the role of funding initiatives. As defined, $O(\bar{F}_p) = X$ or M is a two-state outcome variable with complementary likelihoods, $P(X) + P(M) = 1$. Thus, we apply logistic regression to model the odds $Q \equiv \frac{P(X)}{P(M)}$, measuring the propensity to adopt cross-domain configurations. We then estimate the annual growth in $P(X)$ by modeling the odds as $\log(Q_p) = \beta_0 + \beta_y y_p + \vec{\beta} \cdot \vec{x}$, where \vec{x} represents the additional controls for confounding sources of variation, in particular increasing k_p , associated with the growth of team science (Milojevic, 2014, Wuchty et al., 2007). See SI Appendix Text S7, in particular Eqns. (S2)–(S4), for the full model specification; and, Tables S1–S3 for parameter estimates.

Summary results shown in Fig. 5A indicate a roughly 3% annual growth in $P(X_{SA})$, consistent with descriptive trends shown in Fig. 2. In contradistinction, growth rates for $P(X_{CIP})$ are generally smaller, indicative of the additional barriers to integrating individual expertise as opposed to just combining different research topics. In the case of $P(X_{SA\&CIP})$, the growth rate is higher for Distant, where the need for cross-disciplinary expertise cannot be short-circuited as easily as in Neighboring.

A component of RQ5 is how HBS projects have altered the propensity for X. Hence, we add an indicator variable I_{2014+} that takes the value 1 for articles with $y_p \geq 2014$ and 0 otherwise. Figure 5B indicates a significant decline in $P(X)$ for X_{CIP} and $X_{SA\&CIP}$ for each configuration on the order of $\sim 30\%$; this result is consistent with the recent increase in $f_1(t|z)$ visible in Fig. 2B.

Model B: Quantifying the citation premium associated with X and funding initiatives. We model the normalized citation impact

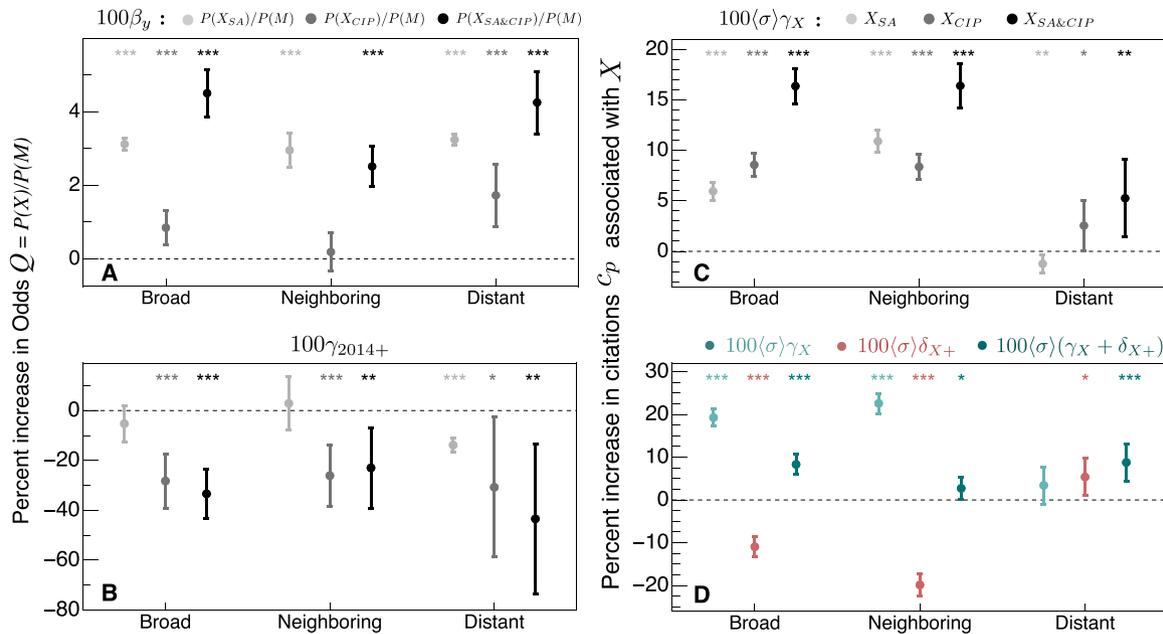


Fig. 5 Propensity for X and citation impact attributable to cross-domain activity at the article level. **A** The annual growth rate in the likelihood $P(X)$ of research having cross-domain attributes represented generically by X . **B** Decreased likelihood $P(X)$ after 2014. **C** Citation premium estimated as the percent increase in c_p attributable to cross-domain mixture X , measured relative to mono-domain (M) research articles representing the counterfactual baseline. Calculated using a researcher fixed-effect model specification which accounts for time-independent individual-specific factors; see Tables S4–S5 for full model estimates. **D** Difference-in-Difference (δ_{X+}) estimate of the "Flagship project effect" on the citation impact of cross-domain research. Shown are point estimates with a 95% confidence interval. Asterisks above each estimate indicate the associated p – value level: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

$z_p = \alpha_a + \gamma_{X_{SA}} I_{X_{SA,p}} + \gamma_{X_{CIP}} I_{X_{CIP,p}} + \vec{\beta} \cdot \vec{x}$, where \vec{x} represents the additional control variables and α_a represents an author fixed-effect to account for unobserved time-invariant factors specific to each researcher. The primary test variables are $I_{X_{SA,p}}$ and $I_{X_{CIP,p}}$, two binary factor variables with $I_{X_{SA,p}} = 1$ if $O_{SA}(\vec{F}_p) = X_{SA}$ and 0 if $O_{SA}(\vec{F}_p) = M$, defined similarly for CIP. To distinguish estimates by configuration, for *Neighboring* we specify $I_{X_{Neighboring,SA}}$ and $I_{X_{Neighboring,CIP}}$, with similar notation for *Distant*. Full model estimates are shown in Tables S4–S5.

Figure 5C summarizes the model estimates— $\gamma_{X_{SA}}$, $\gamma_{X_{CIP}}$ and $\gamma_{X_{SA\&CIP}}$ —quantifying the citation premium attributable to X . To translate the effect on z_p into the associated citation premium in c_p , we calculate the percent change $100\Delta c_p/c_p$ associated with a shift in $I_{X,p}$ from 0 to 1. Observing that $\sigma_t \approx \langle \sigma \rangle = 1.24$ is approximately constant over the period 1970–2018 and due to the property of logs, the citation percent change is given by $100\Delta c_p/c_p \approx 100\langle \sigma \rangle \gamma_X$ (see SI Appendix S7B).

Our results indicate a robust statistically significant positive relationship between cross-disciplinarity (X_{CIP}) and citation impact, consistent with the effect size in a different case study of the genomics revolution (Petersen, 2018), which supports the generalizability of our findings to other convergence frontiers. To be specific, we calculate an 8.6% citation premium for the *Broad* configuration ($\gamma_{X_{CIP}} = 0.07$; $p < 0.001$), meaning that the average cross-disciplinary publication is more highly cited than the average mono-disciplinary publication. We calculate a smaller 5.9% citation premium associated with X_{SA} ($\gamma_{X_{SA}} = 0.05$; $p < 0.001$). Yet the effect associated with articles featuring X_{CIP} and X_{SA} simultaneously is considerably larger (16% citation premium; $\gamma_{X_{SA\&CIP}} = 0.13$; $p < 0.001$), suggesting an additive effect.

Comparing results for the *Neighboring* configuration to the baseline estimates for *Broad*, the citation premium is relatively larger for X_{SA} (11% citation premium; $\gamma_{X_{Neighboring,SA}} = 0.088$; $p < 0.001$) and roughly the same for X_{CIP} and $X_{SA\&CIP}$. This result reinforces our findings regarding the *convergence short-cut* (when X_{CIP} is absent), indicating that this approach is more successful when integrating domain knowledge across shorter distances, consistent with innovation theory (Fleming, 2001).

The configuration most representative of classically construed convergence is *Distant*, which compared to *Broad* and *Neighboring* features smaller effect size for $X_{SA\&CIP}$ (5.2% citation premium; $\gamma_{X_{Distant,SA\&CIP}} = 0.04$; $p < 0.001$). The reduction in $\gamma_{X_{Distant,SA\&CIP}}$ relative to values for *Broad* and *Neighboring* configurations likely reflects the challenges bridging communication, methodological and theoretical gaps across the *Distant* neuro-psycho-medical ↔ techno-informatic interface. More interestingly, this configuration is distinguished by a negative X_{SA} estimate, indicating that the convergence shortcut yields less impactful research than mono-domain research. Nevertheless, it is notable that for this convergent configuration, there is a clear hierarchy indicating the superiority of cross-disciplinary collaboration approaches to integrating research across distant domains.

As in the Article-level model, we also tested for shifts in the citation premium attributable to the advent of Flagship HBS project funding using a similar difference-in-difference (DiD) approach. Figure 5D shows the citation premium $\gamma_{X_{SA\&CIP}}$ for articles published prior to 2014, and the difference δ_{X+} corresponding to the added effect for articles published after 2013. For *Broad* and *Distant* we observe $\delta_{X+} < 0$, indicating a reduced citation premium for post-2013 research. By way of example for the *Broad* configuration: whereas cross-domain articles published prior to 2014 show a 19% citation premium

($\gamma_{X_{SA\&CIP}} = 0.15$; $p < 0.001$), those published after 2013 have just a 19–11% = 8% citation premium ($\delta_{X_{SA\&CIP+}} = -0.09$; $p < 0.001$). The reduction of the citation premium is even larger for *Neighboring* ($\delta_{Neighboring, X_{SA\&CIP+}} = -0.16$; $p < 0.001$). Yet for *Distant*, we observe a different trend—research combining both X_{SA} and X_{CIP} simultaneously has an advantage over those with just X_{CIP} or X_{SA} , in that order ($\delta_{Distant, X_{SA\&CIP+}} = 0.04$; $p = 0.016$; 95% CI = [.01, .08]).

We briefly summarize coefficient estimates for the other control variables. Consistent with prior research on cross-disciplinarity (Petersen, 2018), we observe a positive relationship between team-size and citation impact ($\beta_k = 0.415$; $p < 0.001$), which translates to a $\langle \sigma \rangle \beta_k \approx 0.5\%$ increase in citations associated with a 1% increase in team size (since k_p enters in log in our specification). We also observe a positive relationship for topical breadth ($\beta_w = 0.03$; $p < 0.001$), which translates to a much smaller $\langle \sigma \rangle \beta_w \approx 0.04\%$ increase in citations associated with a 1% increase in the number of major MeSH headings. And finally, regarding the career life-cycle, we observe a negative relationship with increasing career age ($\beta_\tau = -0.011$; $p < 0.001$) consistent with prior studies (Petersen, 2018), translating to a $100\langle \sigma \rangle \beta_\tau \approx -1.3\%$ decrease in c_p associated with every additional career year. See Tables S4–S5 for the full set of model parameter estimates.

Behind the numbers. Qualitative inspection of prominent research articles in the $X_{Distant, SA\{\&\}CIP}$ configuration identifies four key convergence themes associated with past or developing breakthroughs:

Magnetic resonance imaging (MRI). MRI technology has been instrumental in identifying structure-function relations in brain networks and has reshaped brain research since the 1990s. As a method that involves both sophisticated technology and core brain expertise, MRI has been a focal point for $X_{Distant, SA\&CIP}$ scholarship. For example, ref. (Van Dijk et al., 2012) addresses the problem of motion, a pernicious confounding factor that can invalidate MR brain results. Hence, this research article exemplifies how a fundamental problem threatening an entire line of research acts as an attractor of distant cross-disciplinary collaborations with an all-encompassing theme, including authors from CIP 5 (medical specialists) and CIP 8 (engineers and computer scientists), while thematically spans four topical domains: SA 2 (Anatomy and Organisms), SA 3 (Phenomena and Processes), SA 5 (Techniques and Equipment), and SA 6 (Technology and Information Science).

Genomics. Following the completion of the Human Genome Project (HGP) in the early 2000s, genomics, and biotechnology methods have established a foothold in brain research. This convergent frontier made headway in solving long-standing morbidity riddles and formulating novel therapies, e.g., providing a deeper understanding of the genetic basis of developmental delay (Cooper, 2011) and developing a treatment for glioblastoma using a recombinant poliovirus (Desjardins, 2018). Both of these articles include authors from CIP 4 and 5, while thematically cast a wide net, with the former spanning SA 1, 3, 4, and 5, whereas the latter SA 2, SA 4, and SA 5.

Robotics. In the early 2010s, neurally controlled robotic prostheses reached fruition by way of collaboration between neuroscientists (CIP 1) and biotechnologists (CIP 4). A prime example of this emerging bio-mechanics frontier is research on robotic arms for tetraplegics (Hochberg, 2012), which thematically covers all SA 1–6.

Artificial Intelligence (AI) and big data. Following developments in machine learning capabilities (ML), deep AI methods were brought to bear on MR data, pushing brain imaging towards more quantitative, accurate, and automated diagnostic methods. Research on brain lesion segmentation using Convolutional Neural Networks (CNN) (Kamnitsas, 2017) is an apt example produced by a collaboration between medical specialists (CIP 5) and engineers (CIP 8), spanning thematically SA 2–4 and SA 6. Simultaneously, massive brain datasets combined with powerful AI engines made their appearance along with methods to control noise and ensure their validity, as exemplified by ref. (Alfaro-Almagro, 2018) produced by neuroscientists (CIP 1), health scientists (CIP 6), and engineers (CIP 8), and also featuring a nearly exhaustive topical scope (SA 2–6).

All together, case analysis indicates $X_{Distant, SA\&CIP}$ products are typically characterized by significant SA integration, typically including 3–4 non-technical SA plus 1–2 technical SA. This thematic coverage exceeds the disciplinary bounds implied by the CIP set of the authors, which typically includes one non-technical CIP plus one technical CIP.

Discussion

In a highly competitive and open science system with multiple degrees of freedom, more than one operational mode is likely to emerge. To assess the different configurations that exist, we developed an {author discipline \times research topic} classification that enables examination of several operational modes and their relative scientific impact.

Competing convergence modes. Our key result regards the identification and assessment of a prevalent convergence shortcut characterized by research combining different SA (X_{SA}) but not integrating cross-disciplinary expertise (M_{CIP}). Assuming the HBS ecosystem to be representative of other competitive science frontiers, our results suggest that the two operational modes of convergence evolve as substitutes rather than complements. Trends from the last five years indicate an increasing tendency for scholars to shortcut cross-disciplinary approaches, and instead integrate by way of expansive learning. This appears to be in tension with the intended mission of flagship HBS programs. In fact, our analysis provides strong evidence that the rise of expedient convergence tactics may be an unintended consequence of the race among teams to secure funding.

In order to provide a timely assessment of convergence science, we addressed RQ1—how to measure convergence?—by developing a generalizable framework that differentiates between diversity in team expertise and research topics. While it is true that a widespread paradigm shift towards increasing team size has transformed the scientific landscape (Milojevic, 2014; Wuchty et al., 2007), this work challenges the prevalent view that larger teams are innately more adept at performing cross-domain research. Indeed, convergence does not only depend on team size but also on its composition. In reality, however, research teams targeting the class of hard problems calling for convergent approaches are faced with coordination costs and other constraints associated with crossing disciplinary and organizational boundaries (Cummings and Kiesler, 2005, 2008; Van Rijnsoever and Hessels, 2011). Consequently, teams are likely to economize in disciplinary expertise, and instead integrate cross-domain knowledge in part (or in whole) by way of polymathic generalists comfortable with the expansive learning approach. As a result, a team's composite disciplinary pedigree tends to be a subset of the topical dimensions of the problem under investigation.

As a consistency check, we also find this convergence shortcut to be more widespread in research involving topics that are epistemically close, as represented by the *Neighboring* configuration we analyzed. Contrariwise, in the neuro-psycho-medical ↔ techno-informatic interface, belonging to the *Distant* configuration, convergent cross-disciplinary collaboration runs strong. Perhaps not by serendipity, the mixed analysis further indicates that this is exactly the configuration where transformative science has long been occurring.

Arguably, a certain degree of expansive learning is needed for multidisciplinary teams to operate in harmony. For example, in the case of a psychologist collaborating with a medical specialist, it would be ideal if each one knew a little bit about the other's field so that they establish effective knowledge and communication bridges. After all, this is what transforms a multidisciplinary team into a cross-disciplinary team, such that convergence becomes operative. However, this approach is not the dominant trend in HBS (see section "Article level Model"), and is possibly a response to the broad and longstanding promotion of unqualified interdisciplinarity (Barry et al., 2008, Nissani, 1995). Again using our simple example, it may be that the medical specialist prefers not to partner at all with psychologists in the prosecution of bi-domain research, i.e., opting for the streamlined substitutive strategy of total replacement over the strategy of partial redundancy, which comes with the risks associated with cross-disciplinary coordination.

In confronting a rich set of questions we were faced with well-documented limitations to interdisciplinary studies, namely inconsistencies in defining and operationalizing metrics for interdisciplinarity, as exemplified by the inexact yet common method of assigning a discipline to a research article based upon the research area classification of the parent journal (Wagner, 2011). In this work, we address these issues by instead assigning disciplinary categories to researchers based upon their primary departmental affiliation, and to research topics based upon the location of the article's keywords within a comprehensive biomedical ontology. Moreover, a true multi-domain ecosystem like HBS provides an ideal testbed for evaluating the prominence, interactions, and impact of the said constitutional aspects of convergence (Eyre, 2017, Grillner, 2016, Jorgenson, 2015, Quaglio, 2017). A persisting limitation, however, is that we do not specify what task (e.g., analysis, conceptualization, writing) a given domain expert performed, and hence we do not account for the division of labor in the teams here analyzed. Indeed, recent work provides evidence that larger teams tend to have higher levels of task specialization (Haeussler and Sauermann, 2020), which thereby provides a promising avenue for future investigation, i.e., to provide additional clarity on how bureaucratization (Walsh and Lee, 2015) offsets the recombinant uncertainty (Fleming, 2001) associated with cross-disciplinary exploration. Another limitation regards the nuances of HBS programs that we do not account for, e.g., different grand objectives, funding levels, and disciplinary framing which varies across flagships.

Our results also provide clarity regarding recent efforts to evaluate the role of cross-disciplinarity in the domain of genomics (Petersen, 2018), where we used a similar scholar-oriented framework that did not incorporate the SA dimensions. One could argue that the cross-disciplinary citation premium reported in the genomics revolution arises simply from the genomics domain being primed for success. Indeed, Fig. 3A shows that HBS scholars in the domain of Biotech and Genetics discipline maintained high levels of SA diversity extending back to the 1970s. We do not observe similar patterns for other HBS sub-disciplines. Yet, our measurement of a ~16% citation premium for research featuring both modes ($X_{SA\&CIP}$) are remarkably similar in magnitude to the analog measurement of a ~20%

citation premium reported in (Petersen, 2018). In other recent work operationalizing STEM disciplines by way of WOS Research Area (SU) categories, which are journal specific (as opposed to article-level keywords and author-level CIP affiliations), (Okamura, 2019) estimates the analog interdisciplinary citation premium of roughly 20% for a set of 2560 "highly cited papers".

Econometric analysis. In order to measure shifts in the prevalence and impact of cross-domain integration, in addition to how they depend on the convergence mode, we employed an econometric regression specification that leverages author fixed-effects and accounts for research team size, in addition to a battery of other CIP and SA controls. Regarding the growth rate of HBS convergence science, Fig. 5A indicates that research integrating topics and disciplinary expertise is growing between 2 and 4% annually, relatively to the mono-disciplinary baseline; however, this upward trend reversed after the ramp-up of HBS flagships, as indicated in Fig. 5B. Our results also indicate that the citation impact of publications from polymathic teams ($X_{Neighboring,SA}$ and $X_{Distant,SA}$) is significantly lower than the impact of publications from more balanced cross-disciplinary teams ($X_{Neighboring,SA\&CIP}$ and $X_{Distant,SA\&CIP}$), see Fig. 5C. On a positive note, a DiD strategy provides support that HBS research featuring the $X_{Distant,SA\&CIP}$ configuration has increased in citation impact following the ramp-up of HBS flagships, see Fig. 5D. There are various possible explanations to consider, most prominent of which is that the cognitive and resource demands required to address grand scientific challenges have outgrown the capacity of even mono-disciplinary teams, let alone solo genius (Simonton, 2013).

Reflecting upon these results together, it is somewhat troubling that the polymathic trend proliferates and competes with the gold standard, that is, configurations featuring a balance of cross-disciplinary teams and diverse topics ($X_{SA\&CIP}$). Counterproductively, flagship HBS projects appear to have incentivized expansive research strategies manifest in a relative shift towards X_{SA} since the ramp-up of flagship projects in 2014. This trend may depend upon the particular flagship's objective framing. Take for instance the US BRAIN Initiative, with the expressed aim to support multi-disciplinary mapping and investigation of dynamic brain networks. As such, its corresponding research goals promote the integration of *Neighboring* topics, where scientists with polymathic tendencies may feel more emboldened to short-circuit expertise. In addition, there are practical pressures associated with proposal calls. Another possible explanation regarding team formation is that it may be easier and faster for researchers to find collaborators from their own discipline when faced with the pressure to meet proposal deadlines. In addition, funding levels are not unlimited, and bringing additional reputable specialists into the team comes with great financial consideration. Hence, a natural avenue for future investigation is to test whether other convergence-oriented funding initiatives also unwittingly amplify such suboptimal teaming strategies.

Theoretical insights—autodidactic expansive learning. Indeed, the polymathic trend described here pre-existed the flagship HBS projects, and so must have deeper roots. One hypothesis is that this trend represents an emergent scholarly behavior owing to efficient 21st century means to pursue new topics by way of expansive learning (Engstrom and Sannino, 2010) since the learning costs associated with certain tasks characterized by explicit knowledge have markedly decreased with the advent of the Internet and other means of rapid high-fidelity communication. Indeed, many of the activity signals brought to the fore by this study bear the hallmarks of expansive learning. Perhaps the

most telling signal is the propensity towards topically diverse publications—Fig. 4D–F, which largely stems from horizontal movements in the research focus of individual scientists rather than vertical integration among experts from different disciplines—Fig. 4A–C. The scientific system is increasingly interconnected, as evident from the densification of collaboration networks and emergent cross-disciplinary interfaces—Fig. 2B. These interfaces satisfy the conditions that are conducive to boundary-crossing, especially with respect to research topics, which can act as structures facilitating “minimum energy” expansion (Toiviainen, 2007). To this point, we also assessed whether the relationship between CIP diversity and SA integration depends on whether the configuration represents neighboring or distant domains. Analyzing the set of $X_{SA\&CIP}$ articles, we find that expansive integration is consistently most effective in *Distant* configurations, e.g., teams with $N_{CIP,p} = 3$ spans roughly 32% more SA than their mono-disciplinary counterparts—Fig. S5B.

Policy implications. Consistent also with other studies in expansive learning, actions taken by participants do not necessarily correspond to the intentions of the interventionists (Rasmussen and Ludvigsen, 2009). Here, the participants are brain scientists, and the interventionists are the funding agencies. While the latter aim to promote research powered by true multi-disciplinary teams, the former prefer the shortcut.

Policymakers and other decision-makers within the scientific commons are faced with the persistent challenge of efficient resource allocation, especially in the case of grand scientific challenges that foster aggressive timelines (Stephan, 2012). The implicit uncertainty and risk associated with such endeavors are bound to affect reactive scholar strategies, and this interplay between incentives and behavior is just one source of complexity among many that underly the scientific system (Fealing, 2011).

To begin to address this issue, policies addressing the challenges of historical fragmentation in Europe offer guidance. European Research Council (ERC) funding programs have been powerful vehicles for integrating national innovation systems by way of supporting cross-border collaboration, brain-circulation, and knowledge diffusion—yet with unintended outcomes that increase the burden of the challenge (Doria Arrieta et al., 2017). To address this geographic fragmentation, many major ERC collaborative programs require multi-national partnerships as an explicit funding criterion. Motivated by the effectiveness of this straightforward integration strategy, convergence programs could include analog cross-disciplinary criteria or review assessments to address the convergence shortcut. Such guidelines could help to align polymathic versus cross-disciplinary pathways towards more effective cross-domain integration. This type of alignment is critical to achieving the explicit objective delineated by the originators of convergence science (National Research Council, 2014), which is the emergence of “new sciences” characterized by a deep understanding of causality and complexity associated with multi-component multi-level systems (Bonaccorsi, 2008). Much like the vision for brain science—towards a more complete understanding of the emergent structure-function relation in an adaptive complex system—a better understanding of cross-disciplinary team assembly, among other team science considerations (Börner, 2010), will be essential in other challenging frontiers calling on convergence.

Conclusion

We introduced a new method to quantify convergence that accounts for both the topical diversity of an article’s content, as well as the disciplinary diversity of the team that produced it. The method uniquely captures the tension between the

institutionalized distribution of expertise and the freely evolving distribution of scientific topics. In this new framework, classic convergence—defined as both cross-topic and cross-discipline—is only one among three possibilities; the other two are mono-topic/cross-discipline and cross-topic/mono-discipline. The latter type is rather unsettling because it represents a bold bypass of institutionalized expertise; for this reason, we call it *shortcut convergence*. Applying our method of analysis to a comprehensive bibliographic dataset in human brain science, we found that shortcut convergence is crowding out classic convergence, despite the latter’s impact. Furthermore, we found that this trend is unintentionally aided by funding initiatives, partly because there has been no methodological framework thus far to monitor convergence nuances.

One might ask if the convergence mechanism we discovered to be humming under the surface is active in fields other than brain science? We believe this is likely, because brain science is associated with a multi-domain ecosystem primed for convergence, and this is a highly representative case. Nevertheless, repeating the type of research we described here in more scientific fields is the only way to ascertain the universality of the convergence mechanism we unearthed. In this direction, our general methods and open code offer a valuable tool to those in the researchers interested to take up this task.

Data availability

The curated dataset used in the analysis described herein can be accessed through the Open Science Framework repository <https://osf.io/d97eu>. The R code used to produce some of the figures and tables reported in the paper is accessible through GitHub at <https://github.com/UH-CPL/Emergent-Modes-of-Convergence-Science>.

Received: 8 March 2021; Accepted: 30 June 2021;

Published online: 04 August 2021

References

- Alfaro-Almagro F, Jenkinson M, Bangnerter NK, Andersson JL, Griffanti L, Douaud G, Sotiropoulos SN, Jbabdi S, Hernandez-Fernandez M, Vallee E, Vidaurre D (2018) Image processing and quality control for the first 10,000 brain imaging datasets from UK Biobank. *Neuroimage* 166:400–424
- Amunts K et al. (2016) The Human Brain Project: creating a European research infrastructure to decode the human brain. *Neuron* 92(3):574–581
- Arroyave F, Goyeneche OY, Gore M, Heimeliks G, Jenkins J, Petersen A (2021) On the social and cognitive dimensions of wicked environmental problems characterized by conceptual and solution uncertainty. <https://arxiv.org/abs/2104.10279>
- Australian Brain Alliance Steering Committee and others (2016) Australian Brain Alliance. *Neuron* 92(3):597–600
- Baliotti S, Mäs M, Helbing D (2015) On disciplinary fragmentation and scientific progress. *PLoS ONE* 10(3):e0118747
- Barry A, Born G, Weszkalnys G (2008) Logics of interdisciplinarity. *Econ Soc* 37(1):20–49
- Bonaccorsi A (2008) Search regimes and the industrial dynamics of science. *Minerva* 46(3):285–315
- Börner K et al. (2010) A multi-level systems perspective for the science of team science. *Sci Transl Med* 2(49):49cm24–49cm24
- Bromham L, Dinnage R, Hua X (2016) Interdisciplinary research has consistently lower funding success. *Nature* 534(7609):684–687
- Cooper GM et al. (2011) A copy number variation morbidity map of developmental delay. *Nat Genet* 43(9):838
- Cummings JN, Kiesler S (2005) Collaborative research across disciplinary and organizational boundaries. *Soc Stud Sci* 35(5):703–722
- Cummings JN, Kiesler S (2008) Who collaborates successfully? Prior experience reduces collaboration barriers in distributed interdisciplinary research. In: *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work*. ACM, pp. 437–446
- Desjardins A et al. (2018) Recurrent glioblastoma treated with recombinant poliovirus. *New Engl J Med* 379(2):150–161

- DoriaArrieta OA, Pammolli F, Petersen AM (2017) Quantifying the negative impact of brain drain on the integration of European science. *Sci Adv* 3:e1602232
- Engeström Y, Sannino A (2010) Studies of expansive learning: foundations, findings and future challenges. *Educ Res Rev* 5(1):1–24
- Eyre HA et al. (2017) Convergence science arrives: how does it relate to psychiatry? *Acad Psychiatry* 41(1):91–99
- Fealing KH (eds) (2011) *The Science of Science Policy: A Handbook*. Stanford Business Books, Stanford
- Fleming L (2001) Recombinant uncertainty in technological search. *Manag Sci* 47(1):117–132
- Fleming L (2004) Perfecting cross-pollination. *Harvard Business Rev* 82(9):22–24
- Fleming L, Sorenson O (2004) Science as a map in technological search. *Strategic Manag J* 25(8–9):909–928
- Fortunato S et al. (2018) Science of science. *Science* 359(6379):eaao0185
- Foster JG, Rzhetsky A, Evans JA (2015) Tradition and innovation in scientists' research strategies. *Am Sociol Rev* 80(5):875–908
- Grillner S et al. (2016) Worldwide initiatives to advance brain research. *Nat Neurosci* 19(9):1118–1122
- Haeussler C, Sauermann H (2020) Division of labor in collaborative knowledge production: the role of team size and interdisciplinarity. *Res Policy* 49(6):103987
- Harrison DA, Klein KJ (2007) What's the difference? Diversity constructs as separation, variety, or disparity in organizations. *Acad Manag Rev* 32(4):1199–1228
- Helbing D (2012) Accelerating scientific discovery by formulating grand scientific challenges. *Eur Phys J Special Topics* 214(1):41–48
- Hochberg LR et al. (2012) Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. *Nature* 485(7398):372–375
- Hughes JA, Hughes J (2003) *The Manhattan Project: Big Science and the Atom Bomb*. Columbia University Press
- Jabalpurwala I (2016) Brain Canada: one brain one community. *Neuron* 92(3):601–606
- Jeong SJ et al. (2016) Korea brain initiative: integration and control of brain functions. *Neuron* 92(3):607–611
- Jorgenson LA et al. (2015) The BRAIN Initiative: developing technology to catalyze neuroscience discovery. *Philos Trans Royal Soc B* 370(1668):20140164
- Kamnitsas K et al. (2017) Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med Image Anal* 36:61–78
- Leahey E, Moody J (2014) Sociological innovation through subfield integration. *Soc Curr* 1(3):228–256
- Linkov I, Wood M, Bates M (2014) Scientific convergence: dealing with the elephant in the room. *Environ Sci Technol* 48(18):10539–10540
- Majeti D et al. (2020) Scholar Plot: design and evaluation of an information interface for faculty research performance. *Front Res Metrics Anal* 4:6
- Melero E, Palomeras N (2015) The Renaissance Man is not dead! The role of generalists in teams of inventors. *Res Policy* 44(1):154–167
- Milojevic S (2014) Principles of scientific research team formation and evolution. *Proc Natl Acad Sci USA* 111(11):3984–3989
- National Center for Education Statistics (2021) *The Classification of Instructional Programs (CIP)*. <https://nces.ed.gov/ipeds/cipcode>, Accessed 11 June 2021.
- National Research Council (2005) *Facilitating Interdisciplinary Research*. National Academies Press, Washington, D.C.
- National Research Council (2014) *Convergence: Facilitating Transdisciplinary Integration of Life Sciences, Physical Sciences, Engineering, and Beyond*. National Academies Press, Washington, D.C.
- Nissani M (1995) Fruits, salads, and smoothies: a working definition of interdisciplinarity. *J Educ Thought* 29:119–126
- Okamura K (2019) Interdisciplinarity revisited: evidence for research impact and dynamism. *Pal Commun* 5(1):1–9
- Okano H, Miyawaki A, Kasai K (2015) Brain/MINDS: brain-mapping project in Japan. *Philos Trans R Soc B* 370(1668):20140310
- Oreskes N (2021) What makes a problem 'Hard'? *Sci Am* 324(5):77
- Page SE (2008) *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*. Princeton University Press
- Pavlidis I, Petersen AM, Semendeferi I (2014) Together we stand. *Nat Phys* 10(10):700–702
- Pedersen DB (2016) Integrating social sciences and humanities in interdisciplinary research. *Pal Commun* 2(1):1–7
- Petersen AM (2018) Multiscale impact of researcher mobility. *J Royal Soc Interface* 15(146):20180580
- Petersen AM et al. (2018) Cross-disciplinary evolution of the genomics revolution. *Sci Adv* 4(8):eaat4211
- Petersen AM, Pavlidis I, Semendeferi I (2014) A quantitative perspective on ethics in large team science. *Sci Eng Ethics* 20:923–945
- Petersen AM, Rotolo D, Leydesdorff L (2016) A triple helix model of medical innovation: supply, demand, and technological capabilities in terms of Medical Subject Headings. *Res Policy* 45(3):666–681
- Petersen AM, Vincent EM, Westerling A (2019) Discrepancy in scientific authority and media visibility of climate change scientists and contrarians. *Nat Commun* 10:3502
- Poo MM et al. (2016) China Brain Project: basic neuroscience, brain diseases, and brain-inspired computing. *Neuron* 92(3):591–596
- Quaglio G et al. (2017) Understanding the brain through large, multidisciplinary research initiatives. *Lancet Neurol* 16(3):183–184
- Radicchi F, Fortunato S, Castellano C (2008) Universality of citation distributions: toward an objective measure of scientific impact. *Proc Natl Acad Sci USA* 105:17268–17272
- Rasmussen I, Ludvigsen S (2009) The hedgehog and the fox: a discussion of the approaches to the analysis of ICT reforms in teacher education of Larry Cuban and Yrjö Engeström. *Mind Culture Activity* 16(1):83–104
- Roco M et al. (2013) *Converging knowledge, technology, and society: beyond convergence of nano-bio-info-cognitive technologies*. Springer, New York
- Rotolo D, MesseniPetruzzelli A (2013) When does centrality matter? Scientific productivity and the moderating role of research specialization and cross-community ties. *J Organizational Behavior* 34(5):648–670
- Simonton DK (2013) After Einstein: scientific genius is extinct. *Nature* 493(7434):602–602
- Stephan P (2012) *How economics shapes science*. Harvard University Press, Cambridge
- Teodoridis F (2018) Understanding team knowledge production: the interrelated roles of technology and expertise. *Manag Sci* 64(8):3625–3648
- Toivainen H (2007) Inter-organizational learning across levels: an object-oriented approach. *J Workplace Learn* 19(6):343–358
- US National Library of Medicine (2021) *Medical Subject Headings* 2021. <https://meshb.nlm.nih.gov>, Accessed 6 Nov 2021
- Van Dijk KR, Sabuncu MR, Buckner RL (2012) The influence of head motion on intrinsic functional connectivity MRI. *Neuroimage* 59(1):431–438
- Van Rijnsvoever FJ, Hessels LK (2011) Factors associated with disciplinary and interdisciplinary research collaboration. *Res Policy* 40(3):463–472
- Wagner CS et al. (2011) Approaches to understanding and measuring interdisciplinary scientific research (IDR): a review of the literature. *J Informetr* 5(1):14–26
- Walsh JP, Lee YN (2015) The bureaucratization of science. *Res Policy* 44(8):1584–1600
- Wuchty S, Jones BF, Uzzi B (2007) The increasing dominance of teams in production of knowledge. *Science* 316(5827):1036–1039
- Yang D, Pavlidis I, Petersen AM (2021) Biomedical convergence facilitated by the emergence of technological and informatic capabilities. <https://arxiv.org/abs/2103.10641>
- Youn H et al. (2015) Invention as a combinatorial process: evidence from US patents. *J Royal Soc Interface* 12(106):20150272

Acknowledgements

AMP and IP acknowledge partial support from NSF award # 1738163 entitled "EAGER-From Genomics to Brain Science: What Makes Researchers Tick in Transdisciplinary Initiatives". IP acknowledges additional support from the Eckhard-Pfeiffer Distinguished Professorship Fund. AMP acknowledges additional support from a Hellman Fellow award. Any opinions, findings, and conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the funding agencies. The authors would like to thank Elsevier for providing them with enhanced Application Programming Interface (API) privileges to the Scopus database for collecting part of the study's data.

Author contributions

AMP performed the research, participated in the writing of the manuscript, collected, analyzed, and visualized the data; MA developed software to collect, analyze, and visualize the data; and IP designed the research, performed the research, and participated in the writing of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1057/s41599-021-00869-9>.

Correspondence and requests for materials should be addressed to A.M.P. or I.P.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021