Megaprojects, Digital Platforms, & Productivity: Evidence from the Human Brain Project

Ann-Christin Kreyer[1,2] Lucy Xiaolu Wang[3,1,4]

[1]Max Planck Institute for Innovation and Competition, Germany

[2]Munich Graduate School of Economics, LMU Munich, Germany

[3]University of Massachusetts Amherst, United States

[4]Canadian Centre for Health Economics, Canada

Junior Innovation Economics Conference, HBS 2023.5.11

Productivity slowdown, institutions, and AI in life sciences

- Declining productivity in (biomedical) science & challenging to reach the knowledge frontier (Jones 2009; Bloom et al. 2020)
- Important to create non-market incentives to accelerate long-term, large-scale R&D in multi-disciplinary research
- Particularly relevant & urgent for advancing neurosciences
  - Leading causes of death/disability, expensive, lack of cure
- AI showed potential in life science & gained rising interest
  - But the combination of AI and lab science is not always natural

This Paper: the Human Brain Project (HBP)

Can a ten-year megaproject spur research in AI-brain sciences?

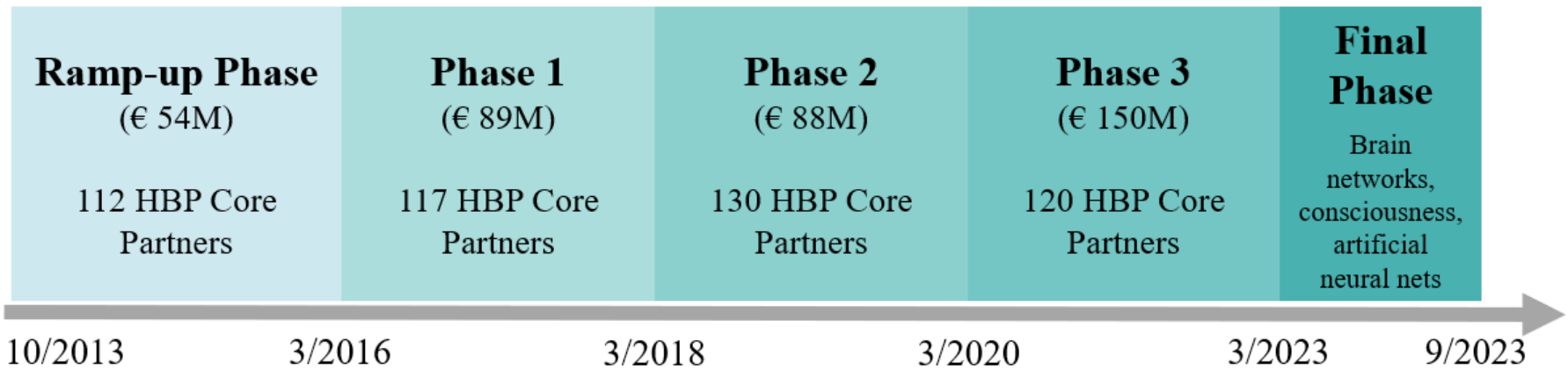
- EU's "Future & Emerging Technology Flagship" 2013-2023
  - 10-year, €1B megaproject, 3 main phases
- Current goal: advance brain science with AI platforms
  - build digital platforms & give grants; public/private access areas to resources within project infrastructure
- Broad areas: study the brain (empirical data, brain simulation, ML); apply the knowledge (medical applications); sharing data, tools & resources (build digital research infrastructure)

Human Brain Project

European Commission

HBP timeline & key events

- 2014-2015: HBP revamp
- 2016: HBP digital platforms started operating
- 2018: transition to EBRAINS (upgraded platform host), built on EU neuroscience supercomputing centers
- 12 subprojects by 2020.3, then reorg. to 9 work packages



Research Questions

Q: How does the HBP affect the rate and direction of R&D?

- Are researchers with earlier access to HBP resources produce more, higher quality, and/or more interdisciplinary work?
- Heterogeneity of the HBP impact by career stages of researchers and by sub-fields within neuroscience/CS?
- What are the relative merits of digital research infrastructure and grants for individual researchers & neuroscience the field?

Goal: to provide the 1st empirical analysis on a contemporary long-term science megaproject, with digital infrastructure & grants

Productivity slowdown, burden of knowledge, megaprojects

- Innovation gets harder with expanding knowledge frontier
  - Need large research teams w complementary skills (Jones 2009)
  - More resources and knowledge re-combination cross-fields
- But large teams can be inefficient & non-creative
  - Moral hazard w credit sharing (Che & Yoo 2001)
  - "Large teams develop & small teams disrupt" (Wu et al., 2019)
- Most innovation projects are risky, unpredictable, long-term
  - Grants/firms under-invest (Azoulay et al. 2011; Budish et al. 2015)
  - Large spillover effects from public R&D and long-term projects (Williams 2013; Myers & Lanahan 2022)
- Megaprojects are on the rise in numbers & controversy
  - ...often involve infrastructure building, large teams, long term
  - Many megaprojects under-perform (Denicol et al. 2020)

▸ literature & contribution

HBP overview: core & partnering project partners

HBP Core Project Partners:

- Sign a partnering agreement w/ EC (incl. grant) each phase
- Are responsible for coordinating & executing the research plan & infrastructure development

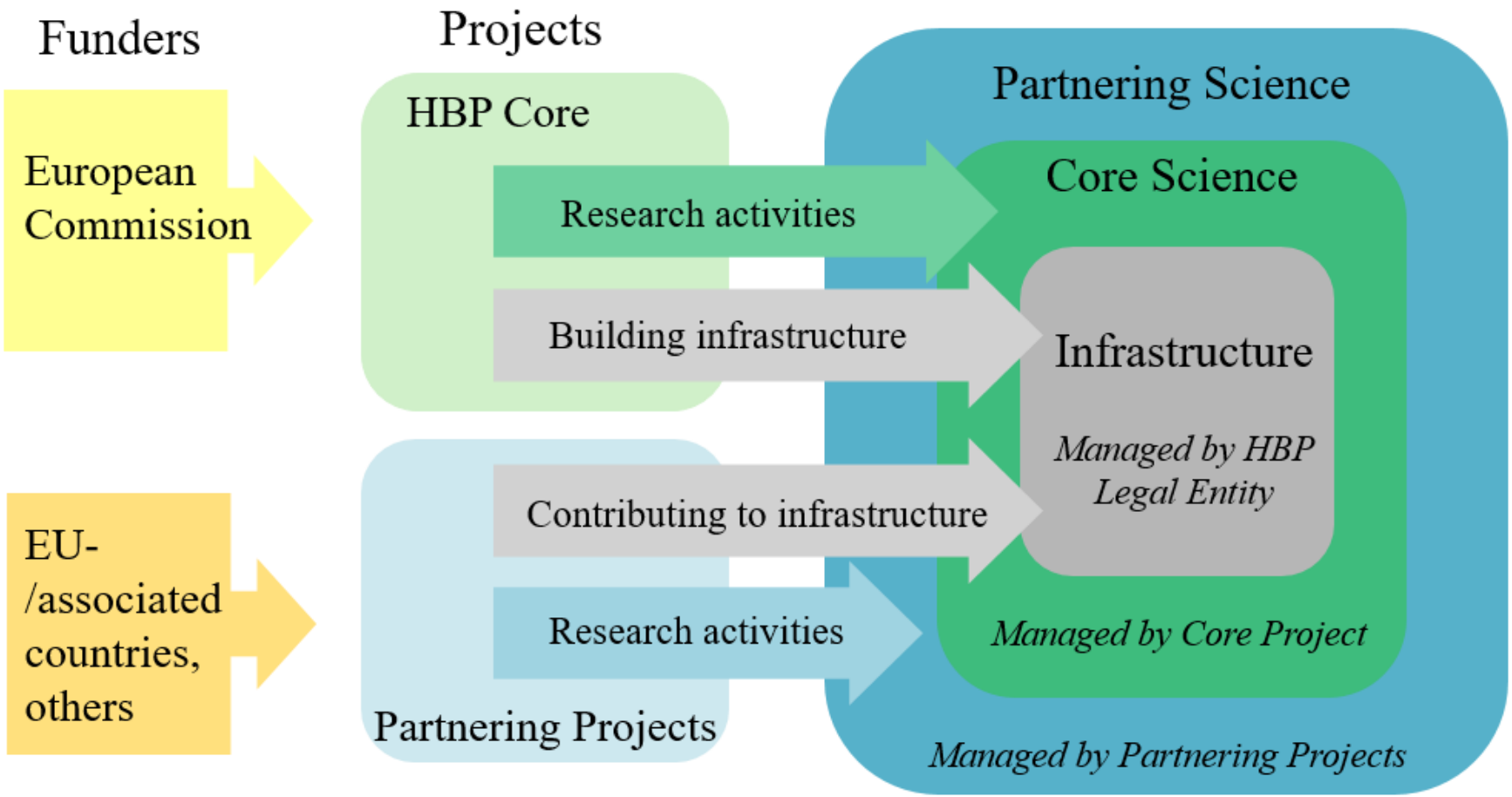
▸ core project partner application process

HBP Partnering Projects Partners:

- Have existing projects w/ own funding
- Sign an agreement w/ Core Project Partners for selected tasks
- Form the creative and unrestrained component of the research and infrastructure-building process

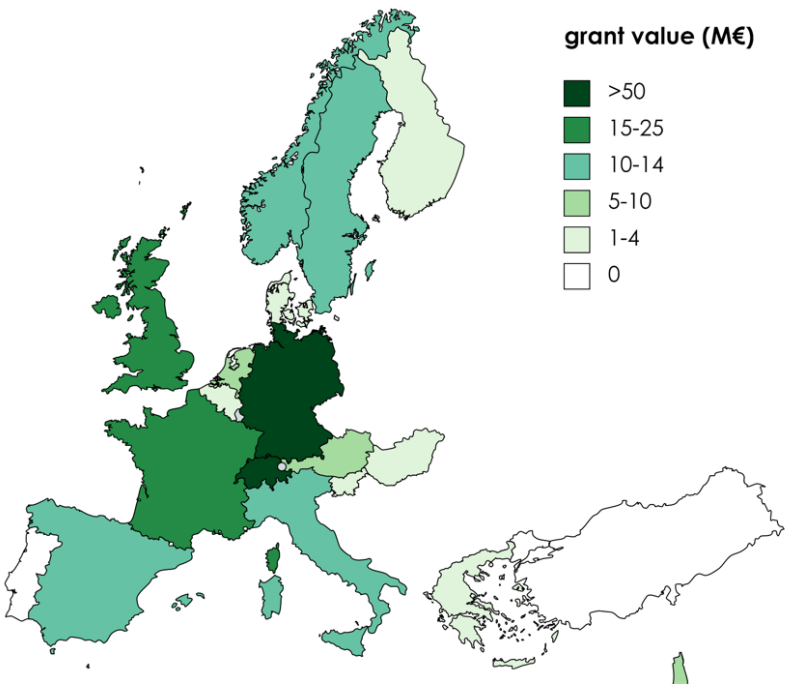
▸ partnering projects application process

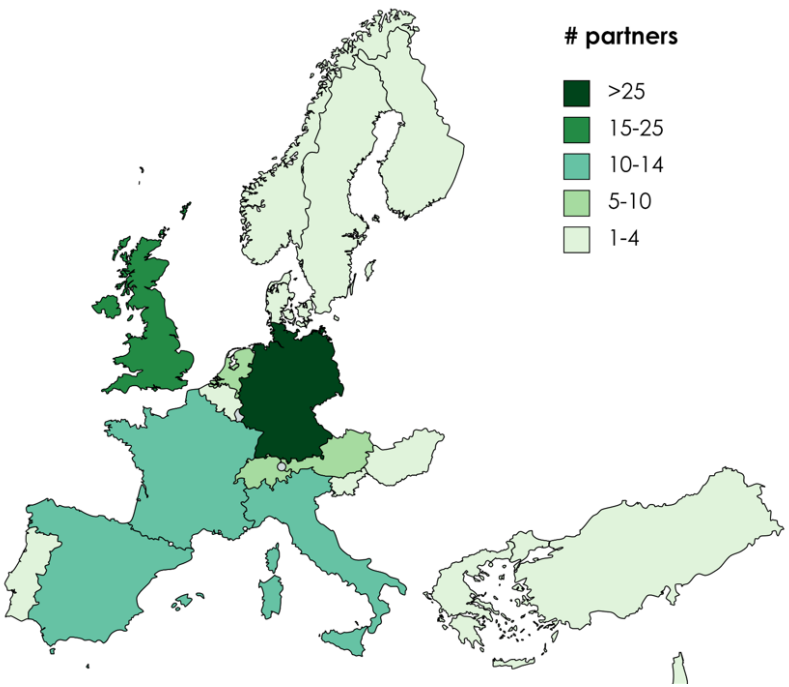
HBP structure: core & partnering projects



HBP overview: core project partners & grants

by the end of phase 2 (2020.3.31): 130 partners in 19 countries





Note: Authors' graph. Excluded are 5 countries (# partners, € received) only participated in phase 0: Canada (1), China (1, 187.5k€), Cyprus (1, 106.2k€), Japan (2), United States (4)

Data: Institutions, grants, individuals, and publications

- Details on 173 HBP Core Project partners by phase 2
  - Department/unit involved, timing of engagement, grants
- Phase-specific data
  - Funding amounts, reports, and deliverables
- Details on 797 active individuals (phases 0-2, +29 in phase 3)
  - HBP involvement (phase active, role), demographic info (gender, affiliation, seniority, research field), Scopus ID
- Publication & citation data: Scopus
  - Publication history of individuals, N=50,616 (2000-2023)
- Sources: HBP websites & YouTube channels, deliverables & reports, framework partnership agreements & amendments, CORDIS, HBP PLUS, search on GoogleScholar/LinkedIn/institution websites

Data collection in progress (which seems never end...)

- Hope to obtain EC core partners proposal evaluation data
- Original CfP for each phase:
  - Research direction & requirements
- Data on the 75 partnering projects
  - Contact persons, country of origin, and project aim
  - HBP affiliation & duration of affiliation
- The preparatory study report pre-HBP (conducted in 2012.4)
  - Potential participants (researchers & research institutions)
  - Estimated costs & resources, expected scientific impact
- NeuroFuture Letter (opposing HBP in 2014)
  - Names of 815 signatories, Scopus IDs, demographic data
- HBP final reports (to be released likely 2023.10)

Benchmark model: DiD w/ staggered timing

- Exploit the time-varying HBP resources (grant/platforms) access among researchers who ever actively participated
  - individuals in core partner institutions can access HBP platforms more easily (exogenous to most individual scientists)

$$y_{i(j)t} = \beta HBP_{it} + \delta_i + \delta_t + \epsilon_{it}$$

- $y_{i(j)t}$: # pubs & citation-weighted pubs (log, numbers, ihs)
- HBP_{it}: indicator whether an individual has access to HBP resources
- $\delta_i + \delta_t$: individual & year fixed effects; s.e. cluster at individual level
- $+\delta_j$ if power allows; identifying variation=job movers w diff HBP access
- apply new methods, e.g., Bacon decomp. 2021, C&S 2021, Roth 2022

Alt. (preferred): synthetic DiD (sDiD)

- Re-weights and matches pre-period trends to weaken reliance on common trend assumption (Arkhangelsky et al., 2021).
- Unit and time weights to make the two-way fixed effect model "local" $\rightarrow$ more robust using only similar units & similar periods

$$\left(\hat{\beta}^{\text{sdid}}, \hat{\delta}, \hat{\gamma}\right) = \underset{\tau,\delta,\gamma}{\arg\min} \left\{ \sum_{i=1}^{N} \sum_{t=1}^{T} \left(Y_{it} - \beta HBP_{it} - \delta_i - \gamma_t\right)^2 \hat{\omega}_i^{\text{sdid}} \hat{\lambda}_t^{\text{sdid}} \right\}$$

- $\hat{\beta}^{\text{sdid}}$: average treatment effect of HBP
- $\hat{\omega}_i^{\text{sdid}}, \hat{\lambda}_t^{\text{sdid}}$: unit & time weights
- δ, γ: individual & year fixed effect vectors

▸ others: mDiD with entropy balancing

Alternative idea: use institutional HBP partnership status and network centrality to key HBP leaders as IVs for individual HBP participation

Text/citation analyses: impact, similarity, novelty

Presence of spillover effects and cross-subsidy in science makes it worthwhile to decompose the impact (à la Sampat & Lichtenberg 2011)

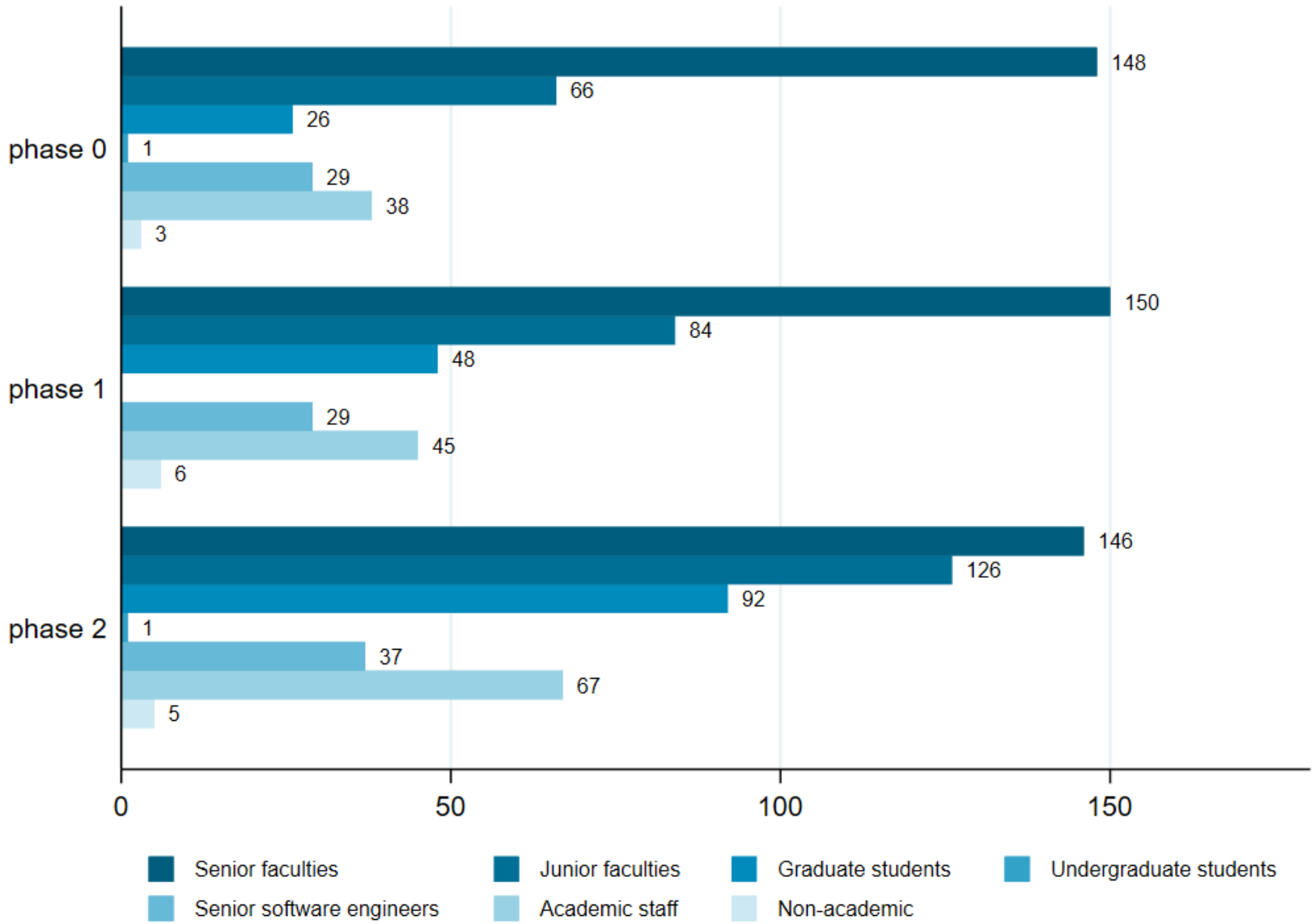
- Direct impact: pubs by HBP-ppl & pubs w/ HBP acknowledgement
- Indirect impact: pubs citing earlier pubs by HBP-ppl/w/ relevance

Text-based novelty, disruptiveness, and HBP relevance measures using NLP: sentenceBERT (sBERT) (Reimers & Gurevych 2019)

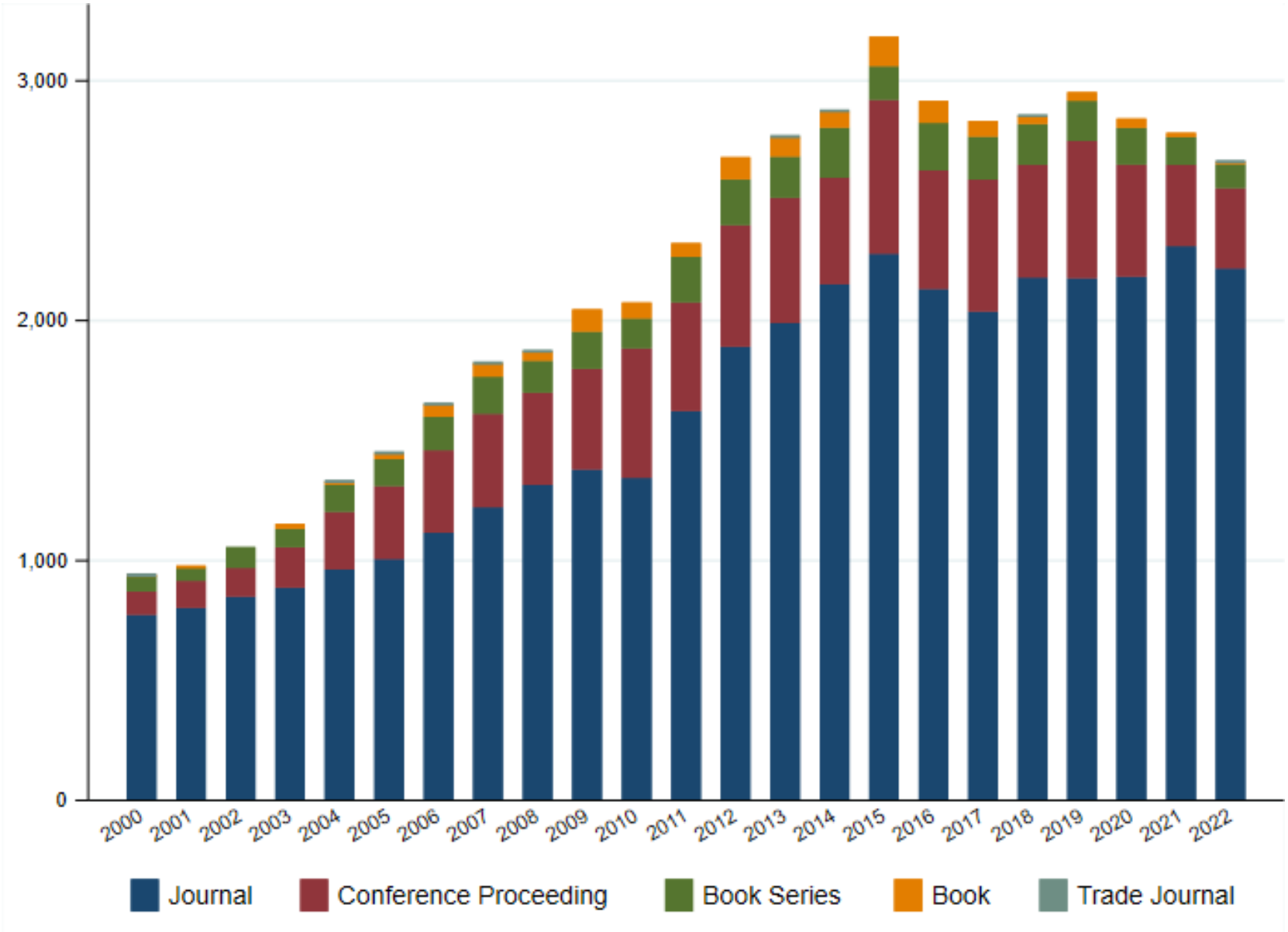
- Other options: Term Frequency-Inverse Document Frequency (TF-IDF), Universal Sentence Encoder (USE), doc2vec

Tag high-value studies w/ computational *and* clinical component

Descriptives: scientists by phase and seniority level



Histogram of publications by ever-HBP individuals



Preliminary Results

- The HBP has gained attraction over time
  - more individuals actively participate, more geographically diverse bases, and more junior faculties & graduate students
- Among ever-participants of HBP, the share of papers increased by HBP active participants & grantees
- HBP increases productivity (v prelim., country-level)
  - mainly through AI platforms rather than direct grants
- The effects are likely larger when account for spillover effects & partnering projects; analyses & heterogeneity in progress

Conclusion & Discussion Questions

- Challenges: many moving pieces, complicated institutions, huge data construction effort, and many possible directions
- Q1: What's the most interesting part of HBP in your view?
- Q2: What's the most convincing emp. strategy you'd push?
- Q3: What are other heterogeneity/mechanism tests to try?
- Anything else you'd like to share: other Qs/suggestions?

Thank you!

Email: xiaoluwang@umass.edu

Literature Review and Contribution

- **Institutions and long-term research:** institutions complement market incentives (Arrow 1962; Boudreau 2010; Furman & Stern 2011; Azoulay et al. 2011; Williams 2013; Budish et al. 2015; Murray et al. 2016; Huang et al. 2022; Myers & Lanahan 2022; Wang 2022)
- **Science of Science:** knowledge complementarity & creation within and cross-organizational teams (Nelson & Winter 1985; Cockburn & Henderson 1994; Hagedoorn 2002; Agrawal et al. 2006; Colombo et al. 2006; Melero & Palomeras 2015; Tortoriello et al. 2015)
- **Digitization in health care:** limited & mixed results, mostly on downstream health (Miller & Tucker 2011; McCullough et al. 2016; Agha 2014; Freedman et al. 2017; Goldfarb et al. 2020; Wang 2021)

Goal: 1st empirical analysis on a contemporary long-term science megaproject (in brain-AI) with digital infrastructure & grants

back

HBP core project (CP) partner application process

- Submit a proposal to the "Call for Expression of Interest"

  (based on Work Plan drawn up by current CP partners for the following phase)

- Eligibility: submit complete, relevant proposals by deadlines
  - Institution of EU-member states/country eligible for EU-grant
  - HBP specific participation criteria (e.g. unique research, unlikely to receive funding from other sources & tight integration across disciplines)

- Evaluation: by the European Commission (EC) & $\geq$3 ext. experts
  - Each evaluation criterion of the call is scored (0-5, 5=excellent)
  - Experts' joint decision is sent to the EC for reviews
  - EC ranks the proposals, # ranked proposals depends on budget
  - Funding decisions are based on ranking: reject vs admit

back

Sources: framework partnership agreements and amendments, EC online portal for funding and tender opportunities

HBP partnering projects application process

- Eligibility: can apply at any time
  - Candidates can be nominated by anyone (e.g., EC, HBP CP partners, national/regional funding agencies, private sector, self)
  - Candidate projects need to have their own funding,
  - and should contribute significantly to the Core Project
- Evaluation: by HBP core project (CP) partners
  - CP partners review the proposals and the fit with the respective sub-project
  - Final decision by the HBP Science & Infrastructure Board
  - Memoranda of Understanding btw HBP & partnering project

back

Sources: framework partnership agreements and amendments, EC online portal for funding and tender opportunities

Alt.: matched-based DiD (mDiD) w/ entropy balancing

- Employ mDiD methods using propensity score matching to create control group (Hainmueller,2012; Hainmueller & Xu, 2013)
- Entropy-balancing to adjust for remaining differences in distribution of covariates btw the groups (Zhao & Percival, 2016)
- List of observable characteristics (preliminary):
  - Gender, age
  - Publication count, field of research
  - Employed institution type, country
- Additionally: sosia-identified Scopus-"doppelganger" (with vs without US-based researchers)

back to sDiD

HBP organization: 12 subprojects

- 12 subprojects with distinct focus areas
  - 1 Mouse Brain Organisation, 2 Human Brain Organisation
  - 3 Systems and Cognitive Neuroscience
  - 4 Theoretical Neuroscience
  - 5-10 the six platforms (Neuroinformatics, Brain Simulation, High-Performance Analytics & Computing, Medical Informatics, Neuromorphic Computing, Neurorobotics)
  - 11 Management & Coordination, 12 Ethics & Society
- Phase 0-2 (2013.10-2020.3): researchers worked in subprojects and/or co-design projects (focus on aligned research infrastructure)
- Phase 3 (2020.4-2023): restructuring to 9 work packages:

Variation in grant & platform access (timing, revolution, rule changes)

- Timevarying access to grant ($0 \leftrightarrow 1$;) & platform ($0 \rightarrow 1 \rightarrow N$)
  - outcomes: scientific deliverables (paper, report, data,...), platform usage (knowledge sharing, user interaction, learning), & link of the two

HBP grants \ HBP Platforms	0	1
0	Others (e.g., Canada, China, Cyprus, Japan, USA; the 5 countries in phase 0 but were excluded after)	Partnership institutions or case-by-case basis (live in 2015; upgraded 2019 and with a legal entity for expanded access)
1	EU countries + Switzerland & Israel (often contract based; China & Cyprus received in phase 0)	Grantees have easy access to the platforms early on (e.g., institution emails are pre-qualified)

▸ back to bibliometrics ▸ back to background

Panel regression: pub-year-country level

$$y_{ct} = \beta HBP_{ct} + \lambda X_{ct} + \delta_c + \delta_t + \epsilon_{ct}$$

	(1)	(2)	(3)	(4)	(5)
HBP access (0/1)	153.9***				51.37***
	(29.56)				(15.59)
HBP grant (M€)		25.92***		9.292*	10.77***
		(2.122)		(5.197)	(3.782)
# HBP partner			22.03***	15.58***	10.58***
			(1.797)	(3.806)	(3.255)
Observations	3,312	3,312	3,312	3,312	3,312
R-squared	0.336	0.417	0.448	0.457	0.469
# countries	138	138	138	138	138
Country FE	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES

Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1