Supporting Information

Bateman and Hess 10.1073/pnas.1421286112

Study 1 Methods

We sent requests to 5,750 scientists who had published a paper in diabetes in the last 5 y. We received surveys from 1,183 respondents, and 748 were complete and usable. Of these, 63% of the respondents were male, the average age was 48, and 51% identified their nationality as the United States.

Respondents considered the following scenario and two hypothetical research projects, counterbalanced across the sample:

> As a scientific researcher, you are faced with choices about which research projects to pursue. The following questions refer to two hypothetical research projects. As "hypotheticals", please assume a perfect world, with all else being equal and funding available for both projects.
>
> Project A (Deep Project): Imagine that you are considering a focused and specialized project that fits your particular interests and leverages your deep expertise in a specific area.
>
> Project B (Broad Project): Imagine that you are considering a broad project that spans several topical domains, including at least one that coincides with your area(s) of expertise and interest.

Several questions asked participants to compare the two projects. The first three questions requested responses using sliding scales similar to Fig. 1. These questions were "To what extent. . ."

1. Would you describe Project A (B) as being a significant opportunity?
2. Would you describe Project A (B) as potentially very important?
3. Would you describe Project A (B) as highly risky?

The remaining questions were:

4. Up to 100%, what is the likelihood that you would pursue project A (B)?
5. If you were forced to choose just one of the two projects, would you choose Project A or Project B?

Study 2 Methods

To ensure that the use of email addresses was not biasing our sample, we performed *t* tests between authors who provided an email address in their publications (most often the corresponding author) and those who did not. We found no significant differences between the two groups of scientists (email provided vs. no email) in terms of the number of publications and the number of coauthors. We then randomly selected and sent surveys to 3,000 of the 5,000 available email addresses. Of the 3,000 emails in our sample frame, 900 were rejected or returned as unusable or invalid. We received 478 responses to our on-line survey request, 12 of which were incomplete, resulting in a sample of 466 completed surveys and a response rate (of 2,100 valid email addresses) of 22%. This response rate is consistent with similar studies of scientists (1).

An independent company that specialized at the time in identifying and collecting information about scientists' backgrounds and performance provided scientific profiles of all respondents. These profiles included demographic information, key affiliations, and all key subject headings for the last 10 y of each scientist's publication output. An outside researcher confirmed these results by comparing our self-reported demographic information with those provided by the third party. In addition, we randomly sampled the scientific profile data and were able to confirm their accuracy with alternative and proven sources of publication data (PubMed, NIH, and Thomson Reuters).

Regarding the new scale of exploratory work behavior and its semantic differential format, exploration (our focal variable) and exploitation (the converse of exploration: applying, extending, and refining current knowledge) originally were conceptualized as incompatible choices that are inversely related (2). Particularly for individuals (compared with organizations), the behaviors are mutually exclusive in that a given resource (time, energy, cognitive effort) that is applied to one is not available for the other (3). Thus, higher levels of exploration connote lower exploitation, and vice versa.

To assess depth and breadth, we used Medical Subject Headings (MeSH) for each publication authored or coauthored by each questionnaire respondent. MeSH terms have been produced since 1960 by the National Library of Medicine (NLM), a member of the US government's National Institute of Health. One of the primary missions of the NLM is the maintenance and oversight of the PubMed/Medline databases. Together, these publication databases contain over 22 million citations for biomedical literature. The MeSH database represents the NLM's controlled vocabulary for subject indexing and searching journal articles in the PubMed and Medline databases. The MeSH thesaurus of terms is published in multiple languages, and used by a number of institutions worldwide. It is widely considered to be one of the most sophisticated structured thesauri in existence (4, 5).

As an example of the hierarchy, one of the broad top-level terms is "organisms"; under "organisms" there are five level-two terms, one of which is "bacteria"; and under "bacteria", there are 19 level-three terms, and so on. The NLM staff assigns each publication between one and twelve MeSH terms. In our sample the mean was 8.1, with a SD of 1.2, indicating a tight distribution around the mean.

Table S1 shows descriptive statistics and intercorrelations. Tables S2 and S3 show the regressions for depth and breadth when controlling for total publications rather than network centrality. The latter are reported in the article, with significance levels and directions fully replicated.

1. Sauermann H, Roach M (2010) Increasing web survey response rates in innovation research: An experimental study of static and dynamic contact design features. *Res Policy* 42(1):273–286.
2. March JG (1991) Exploration and exploitation in organnizational learning. *Organ Sci* 2(1):71–87.
3. Gupta AK, Smith KG, Shalley CE (2006) The interplay between exploration and exploitation. *Academy of Management Journal ARCHIVE* 49(4):693–706.
4. Nelson SJ, Johnston D, Humphreys BL (2001) Relationships in medical subject headings. *Relationships in the Organization of Knowledge*, eds Bean CA, Green R (Kluwer Academic, New York).
5. Omta SW, de Leeuw AC (1997) Management control, uncertainty, and performance in biomedical research in universities, institutes and companies. *J Eng Technol Manage* 14(3):223–257.

Table S1. Descriptive statistics and bivariate correlations

No.	Variables	Alpha	Mean	SD	1	2	3	4	5	6	7	8	9	10	11	12	13
1	Breadth	—	425.21	152.20													
2	Depth	—	254.54	98.25	0.59*												
3	Centrality	—	4.50	7.44	0.43*	0.52*											
4	University researcher	—	0.74	0.44	0.04	0.05	0.02										
5	Sex	—	1.43	0.50	−0.11*	−0.05	−0.04	0.11*									
6	US researcher	—	0.42	0.49	−0.08	0.00	−0.06	0.13*	0.15*								
7	Date of first publication	—	1993	11.20	−0.12*	−0.00	−0.16*	0.01	0.16*	−0.18*							
8	Conscientiousness	0.84	5.42	0.92	−0.11*	0.10	−0.05	0.25*	0.06	0.13*	0.02						
9	Openness to experience	0.78	5.76	0.85	0.38*	0.11*	0.10	0.22*	−0.07	0.04	−0.11*	0.15*					
10	Performance goal orientation	0.83	4.06	1.05	−0.05	0.25*	−0.06	−0.08	−0.06	0.04	0.20*	0.03	−0.12*				
11	Learning goal orientation	0.83	6.14	0.81	0.12*	0.02	−0.03	0.31*	0.13*	−0.01	0.01	0.23*	0.23*	−0.01			
12	Exploration	0.76	5.20	1.20	0.25*	−0.15*	0.01	−0.02	−0.21*	−0.22*	−0.08	−0.16*	0.30*	−0.23*	0.17*		
13	Competitiveness	0.89	4.21	0.91	0.12*	0.24*	−0.00	−0.05	−0.16*	−0.15*	0.13*	0.04	0.01	0.39*	0.12*	−0.04	
14	Total publications		22.5	31.7	0.47*	0.65*	0.52*	0.07	−0.15*	0.01	0.30*	−0.03	0.18*	−0.09	−0.01	0.14*	0.12*

*$P < 0.05$; $n = 466$.

Table S2. Depth regression analysis controlling for total publications instead of network centrality

	Standardized coefficients			
	Model 5 controls		Model 6 full model	
Dependent variable: Depth	Coefficient	SE	Coefficient	SE
Constant	−8.766	(8.430)	3.333	(8.007)
Total publications	0.282***	(0.001)	0.301***	(0.001)
University researcher	−0.021	(0.104)	0.000	(0.097)
Sex	−0.029	(0.096)	−0.006	(0.093)
US researcher	0.002	(0.095)	−0.066	(0.091)
Date of first publication	0.048*	(0.004)	−0.030	(0.004)
Conscientiousness			0.058	(0.049)
Openness to experience			0.153***	(0.054)
Performance goal orientation			0.252***	(0.046)
Learning goal orientation			0.010	(0.056)
Exploration			−0.186***	(0.041)
Competitiveness			0.097*	(0.033)
R^2	0.081		0.229	
Improvement over Base (ΔR^2)			0.148**	

*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; SEs in parentheses.

Table S3. Breadth regression analysis controlling for total publications instead of network centrality

	Standardized coefficients			
	Model 7 controls		Model 8 full model	
Dependent variable: Breadth	Coefficient	SE	Coefficient	SE
Constant	17.900	(8.555)	7.000	(7.791)
Total publications	0.380***	(0.001)	0.352***	(0.001)
University researcher	0.031	(0.106)	−0.005	(0.094)
Sex	−0.011	(0.097)	0.002	(0.090)
US researcher	−0.089*	(0.097)	−0.053	(0.089)
Date of first publication	−0.094*	(0.004)	−0.058*	(0.004)
Conscientiousness			−0.096*	(0.048)
Openness to experience			0.274***	(0.052)
Performance goal orientation			0.038	(0.045)
Learning goal orientation			0.221***	(0.055)
Exploration			0.134**	(0.040)
Competitiveness			0.001	(0.032)
R^2	0.171		0.361	
Improvement over base (ΔR^2)			0.190***	

*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; SEs in parentheses.