doi:10.1038/nature18315

Interdisciplinary research has consistently lower funding success

Lindell Bromham, Russell Dinnage and Xia Hua

Supplementary methods

Supplementary Methods

1. Data

We analysed five years of de-identified research grant application data for research proposals submitted to the Australian Research Council (ARC) Discovery program, including the Field of Research (FOR) codes selected by applicants with their assigned percentages, the number of name chief investigators (CIs), the administering institution (grouped into institutional networks), and whether the proposal was recommended for funding or not. The FOR codes for each proposal, and their funding outcome, are provided in Supplementary Table 1. Discovery Program (DP) applications are submitted in a yearly funding call, for funding in the subsequent calendar year, therefore the application round is known by the funding year (e.g. applications submitted to the Discovery Program in 2010 are given the identifier DP11). Applications to the ARC Discovery Program are made on a confidential basis, so any potentially identifying information cannot be published, such as number of investigators or institution. While the data were provided to us in a de-identified format (with applicant names and institution names removed and with no details on the proposal title or text), several recent studies have highlighted the potential for data re-identification[28,29].

These data allowed us to build a proposal-by-code matrix for each round of applications, in which each proposal is one line, and every possible FOR code is a column. There are 1238 defined 6-digit FOR codes, but since applications only need to select one FOR code, and most proposals have 3 or fewer codes (Extended Data Table 1), most columns are zero for any given proposal. Selected FOR codes were recorded with their percentages. The ARC assigns each proposal to a primary FOR code, identified to four digits, with the percentage assigned to that Group (for example, if all selected codes are within the same 4-digit Group, the percentage on the primary FOR code will be 100%: Supplementary Table 1). This provides a way of assigning each proposal to a primary discipline (e.g. Figure 2).

2. Field of Research (FOR) Codes

Every application to the ARC Discovery program must nominate at least one Field of Research (FOR) code from a defined list of possible codes. FOR codes were jointly developed by the Australian Bureau of Statistics and Statistics New Zealand in 2008 to aid in recording and analyzing the kinds of research and development activities carried out within these two countries. FOR codes use six digits to define a three-level hierarchy, with the first two digits signifying Division, the second two digits signifying Group, and the last two digits signifying Field.

Division	*e.g. 09 Engineering*
Group	*e.g. 0901 Aerospace Engineering*
Field	*e.g. 090101 Aerodynamics (excl. Hypersonic Aerodynamics)*

There are special categories of FOR codes described as “not elsewhere classified”, given the field numbers 99, that pick up fields that have not been adequately captured by other FOR codes – for example “099999 Engineering not elsewhere classified”. Since these may not behave like other 6-digit FOR codes, we ran the analysis both with and without the 99 codes. However, since the results of the analysis did not differ whether the 99 codes were included or excluded, in the text we report results from analyses including all codes (Supplementary Table 2).

On applications to the ARC Discovery program, each FOR code selected must be assigned a percentage weighting, such that the total weighting adds to 100%. By way of illustration, here is an example of the codes selected for an application to the Discovery program:

B2. Field of Research (FOR)

(Select up to three classification codes that relate to your Proposal. Note that the percentages must total 100%.
Click 'Add' after each selection.
Click on the information icon to access the Fields Of Research codes reference.)

Code	Percentage
060309 - Phylogeny and Comparative Analysis	50
060202 - Community Ecology (excl. Invasive Species Ecology)	30
060306 - Evolutionary Impacts of Climate Change	20

The application format for Discovery proposals is constantly being revised, and there was a change in the rules governing the number of FOR codes selected. In the DP11 and DP12 rounds, applicants were able to select more than three codes, but in the DP13, DP14 and DP15 rounds they were restricted to three codes (Supplementary Table 1). To check whether the restriction of number of FOR codes to a maximum of three had any effect on our results, we repeated all analyses on only the last three years of application data (DP13+DP14+DP15). Results were equivalent to those with all years included (Supplementary Table 2).

3. Measuring discipline variety, disparity and evenness.

Any measure of interdisciplinarity should reflect not only the number of different disciplines combined in a project, but also their relative disparity from each other. The hierarchical structure of the Field of Research (FOR) codes allows us to rank proposals with respect to the breadth of fields of research they touch upon, as identified by the applicants.

We adapted a biodiversity metric known as Phylogenetic Species Evenness (PSE)[19,20] to measure the degree of interdisciplinarity of a proposal. When used to measure biological diversity, PSE takes into account the evolutionary relatedness of species to each other, as well as their relative abundances. PSE is the abundance-weighted version of Phylogenetic Species Variability (PSV)[29], which only takes into account species relatedness. PSV measures diversity as the expected variance of a hypothetical continuous trait evolving through Brownian motion across the species' phylogeny or evolutionary tree. The expected variance is standardized by dividing by the maximum expected variance for the same number of species that are all equally related (e.g. they have a non-hierarchical relationship with equal disparity from each other). PSE is calculated in the same manner, only each species "tip" in the phylogeny is replaced by a node with m_i new tips with zero branch-lengths (where m_i is the abundance of species i) radiating from it (i.e. a polytomy). The expected variance of this new tree is then standardized by dividing by the largest possible expected variance, which is when the species are connected by a star phylogeny, and all species have equal abundances. This makes the measure independent of species number, and causes it to always fall between 0 and 1, with 1 being the maximum possible diversity for that particular number of species.

The expected variance in a continuous trait of a set of related species under Brownian motion can be calculated using the phylogenetic variance-covariance matrix[30]. If X_i represents the value of a neutrally evolving trait in species i, then the variance of X_i will be proportional to the branch-length between tip i and the root node, assuming this to be proportional to time spent evolving neutrally. Likewise, the covariance between X_i and X_j of species i and j will be proportional to their shared branch-lengths. Thus closely related species will have high covariance in their X_i traits. The variances of each tip form the diagonal of the phylogenetic variance-covariance matrix, and the covariances between species form the off-diagonals. If V is the n by n phylogenetic variance-covariance matrix (where n is the number of species), then one can define a matrix C which summarizes the correlation structure of the hypothetical trait, and is related to V by a scalar σ^2, which gives the total rate of evolutionary divergence across all n species: $V = \sigma^2 C$. Using this C matrix and a vector m_i of length n, giving the abundances of each species i, one can calculate PSE. The formal definition of PSE[19] is:

$$\mathrm{PSE} = \frac{m\,\mathrm{diag[C]'M - M'CM}}{m^2 - \overline{m}_i m}$$

where M is a n x 1 column vector containing the values of m_i, $\overline{m}_i$ is the mean of m_i, m is the sum of m_i ($\sum_{i=1}^{n} m_i$), diag[C] is a n x 1 column vector giving the main diagonal of C, and prime (') denotes the transpose.

In this framework, groups of closely related species will have a lower expected variance in a neutrally evolving trait, and thus have lower diversity as measured by PSE. In this case, this means that groups of closely related Fields of Research (FOR codes) will have a lower disparity than groups of more FOR codes that are further apart on the discipline hierarchy.

In our dataset, there are 2718 proposals that have selected a single code: these single disciplinary proposals have IDD=0. Proposals get the maximum score, IDD=1, when all the selected codes are from maximally disparate fields (i.e. different Divisions under the FOR code hierarchy). For example, a proposal that nominated the following two codes received the maximum IDD score (Supplementary Table 1):

040604 - Natural Hazards (50%)

091307 - Numerical Modelling and Mechanical Characterisation (50%)

Another example of maximum IDD from the database is a proposal that selected four codes from different Divisions and weighted them all equally:

080111 - Virtual Reality and Related Simulation (25%)

091599 - Interdisciplinary Engineering not elsewhere classified (25%)

111705 - Environmental and Occupational Health and Safety (25%)

130199 - Education systems not elsewhere classified (25%)

Note that in this example, the selection of the "99" codes ('not elsewhere classified') does not have any more influence on the IDD score than any other code. The maximal IDD is due to equal weighting of codes from different Divisions.

The lowest possible non-zero value of IDD under 3-level FOR code hierarchy is IDD = 0.0633, which is given to proposals that select codes from the same Group with 95% weight on one code. An example of this score is a proposal which selected the following codes (Supplementary Table 1):

010102 - Algebraic and Differential Geometry (95%)

010110 - Partial Differential Equations (5%)

The observed distribution of IDD scores for all proposals to the ARC discovery program over a five-year period is given in Extended Data Figure 1. This distribution has a median value of 0.56, which is much lower than would be expected if applicants chose FOR codes at random. This demonstrates that applicants tend to choose codes from related fields rather than selecting codes from disparate Divisions.

4. Creating hierarchical structure for research fields

The InterDisciplinary Distance (IDD) measure requires that research fields be clustered into related groups. The Australian Research Council uses a set of research field codes that have a hierarchical structure. However, IDD can be applied to any proposal data as long as proposals can be associated with research fields. Any form of research field identifiers, such as key words or nominated subject areas, could be grouped into a hierarchy using clustering analysis based on the co-occurrence of terms or subject areas in proposals, which could then be used to rank proposals by their relative interdisciplinarity. Here we demonstrate how a clustering algorithm can be used to create hierarchical structure in research field identifiers. We demonstrate this by adding an additional level of hierarchical structuring to the FOR code classification.

Within each level of the FOR code hierarchy, there is no grouping of more closely related areas (Supplementary Figure 1). So if we use the 3-level hierarchy as defined we would rank a proposal that combines protein and peptides (030406) and analytical biochemistry (060101) as equivalent in degree of interdisciplinarity with one that combines analytical biochemistry and interactive media (190205).

To better capture disciplinary disparity, we repeated all analyses using FOR codes grouped into an additional fourth level, which we will refer to as research Domains. Domains could be defined subjectively, by grouping disciplines perceived to be similar. However, to avoid subjective decisions, we defined the Domains based on clustering analysis of one year's application data (Extended Data Table 1).

We used Year 1 (DP11) proposal data resolved to Division level (first two digits of FOR code) to cluster FOR codes. Any clustering method would be suitable for this purpose, for example we could use mutual information or Kulczynski dissimilarity. However, here we illustrate the approach using the Bray-Curtis dissimilarity measure[31], a distance measure based on co-occurrence which is often used to compare multispecies ecological communities. The Bray-Curtis method, which reflects between-sample similarity, has a number of features that make it well suited to this application. It scales from zero (when two samples are identical) to a maximum value (no shared codes), so that we can use it to determine the degree to which proposals share codes across discipline boundaries. This scaling is independent of the number the codes that are not shared between any comparisons. Identified problems with the use of Bray-Curtis in community ecology, such as inconsistency of species identification or zero sets[32], do not apply to this dataset where all samples select at least one of a defined set of codes. The Bray-Curtis disimilarity BC_{jk} between proposal *j* and proposal *k* is :

$$BC_{jk} = \frac{\sum_{i=1}^{n} |x_{ij} - x_{ik}|}{\sum_{i=1}^{n} (x_{ij} + x_{ik})}$$

where x_{iz} is the percentage weight (or abundance in ecology) of FOR code *i* in proposal *z*, and *n* is the total number of FOR codes. For the purposes of clustering FOR codes into Domains, we ignored the percentage weights in the proposal data and instead considered only whether the FOR code was present or absent in the proposal (i.e. the x_{iz} were converted to either zeroes or ones).

We used the dendrogram produced by the clustering procedure to define higher-level groups, such that we selected clusters that contained greater than one but less than six Divisions (Extended Data Table 1). Domains were then given a 2-letter code, so that each FOR code was converted to an 8-digit code. For example:

Domain *e.g. 04 Life*
Division *e.g. 0406 Biological Sciences*
Group *e.g. 040603 Evolutionary Biology*
Field *e.g. 04060309 Phylogeny and Comparative Analysis*

The additional level of hierarchical structure gives a finer resolution of levels of interdisciplinarity, upweighting proposals from very disparate fields. Under the 4-level FOR code classification, proposals with IDD = 1 (maximal interdisciplinarity) are a subset of proposals with IDD = 1 under the standard 3-level FOR code. An example of a proposal having IDD = 1 under the 3-level code, but IDD <1 under 4-level code is a proposal that selected and equally weighted two codes from different Divisions belonging to the same high-level cluster (both in the Domain "04 Life"):

060110 - Receptors and Membrane Biology (50%)
110903 - Central Nervous System (50%)

The lowest possible non-zero value of IDD under the 4-level FOR codes is IDD = 0.0475, which is given to proposals that select codes from the same Group with 95% weight on one code (see above for an example). All analyses were repeated both on the existing 3-level FOR code hierarchy (Divisions) and the 4-level hierarchy (Domains: Extended Data Table 1). For analyses involving the fourth-level clustering into Domains, we conducted the analyses only on data from Years 2 to 5, to avoid using Year 1 data in both the definition of Domains and their analysis (Supplementary Table 2). Since the results were similar for both, we report results for the existing 3-level FOR code hierarchy in the main text, but we also include the results for the 4-level clustering in Supplementary Table 2 for comparison.

5. Distribution of Interdisciplinary Distance (IDD) scores

Every proposal was assigned an IDD measure between 0 (single disciplinary) to 1 (maximum disciplinary disparity)[19]. The distribution of scores is uneven (Extended Data Figure 1). In particular, there is a peak of proposals with an IDD of zero, signifying that these proposals selected only one FOR code. We conducted two analyses to make sure that the single-code proposals were not skewing the results. First, we analysed datasets both with and without single-code proposals. Both datasets give the same results (Supplementary Table 2). Second, we generated a binary variable which equals 0 if IDD = 0 and equals 1 if IDD > 0. The continuous IDD metric explains a significantly larger amount of variation in the funding success rate than the binary variable (Supplementary Table 2). From these results we conclude that association between IDD and funding success rate is not simply due to the influence of the single disciplinarity in proposals that select only one FOR code.

The uneven distribution of IDD scores could be due to the structure of the FOR code hierarchy, for example because Divisions contain varying numbers of Groups, which contain varying number of Fields (Supplementary Figure 1). In order to test this, we produced a null distribution of IDD scores by replacing each FOR code of each proposal with a randomly sampled FOR code, so that the number of FOR codes and the relative weighting of codes of each proposal were maintained. We found that the null distribution of randomly sampled FOR codes was significantly different from the observed IDD distribution (two-tailed Wilcoxon rank test, $p < 2.2\times10^{-16}$; Extended Data Figure 1). Because randomly chosen codes have a high chance of being from different Divisions, the peak of the null distribution is at 1 (maximum disparity), whereas codes chosen by researchers tend to cluster closely together, so that the peaks of the observed data are at 0 and 0.4. We conclude that the right skew of the observed IDD values, and the "lumpiness" of observed IDD distribution, is not due to the structure of the FOR code hierarchy, but due to biases in code selection toward codes from related fields rather than sampling evenly across the spectrum of available codes.

6. Data analysis

We used a generalized linear mixed model (GLMM) to model the success of grant proposals, using the lme4 package in R[20]. We chose to use GLMM analysis because it is appropriate to model year effect as random effect. Compared to a fixed effect model, a random effect model has the advantage of reducing the variance of estimates of coefficients, but it has the disadvantage of introducing bias if the covariates are correlated with the random variable[33]. Since covariates in our model are not likely to be correlated with year, and we are not primarily interested in estimating the year effect, GLMM is the most appropriate form of analysis. The response is a vector of ones (Recommended for Funding) and zeroes (Not Recommended for Funding). Fixed and random variables are summarized in Supplementary Table 2. We centered the IDD value before each GLMM analysis in order to interpret the coefficient of IDD as the expected change in success rate from IDD = 0 to IDD = 1.

We used a parametric bootstrap to generate confidence intervals for Figure 1 and to test our initial results. Briefly, we used the model parameterized from the observed data to generate many simulated datasets, then refitted the same statistical model to the simulated data in order to generate a distribution of model parameters. We used the bootstrap confidence interval in Figure 1, but in the text we report p-values of the fixed effects based on a Wald's Z test, an analytical approximation that yielded highly similar results to the parametric bootstrap, but which was computationally much more efficient.

6.1 Year

Throughout the analyses, year is treated as a random variable to account for year-to-year variation in success because it is reasonable to treat each year's success rate as a random draw from a population of possible yearly success rates, and we were not interested in measuring the year-to-year variation itself, only controlling for it as a source of non-independence among proposals. The variance estimate of the year by IDD random effect is close to zero, and much less than the variance explained by year itself, which suggests that there is little variation across years in the effect of interdisciplinarity on success (likelihood ratio test, $\chi_1^2 = 0.29$, $p = 0.87$; Supplementary Table 2). When analyzing 4-level clustering of FOR codes, we did not include Year 1 data in the analysis, as it was used to estimate the clusters. We also repeated the 3-level IDD analysis on the last three years of application data because the application format changed to restrict the number of FOR codes selected to a maximum of three (see above).

6.2 Number of disciplines

We tested whether the number of codes selected correlated with IDD or with funding success. We found no evidence on non-linearity between funding success and the number of FOR codes selected, so we assume a linear relationship between them in the GLMM analyses. When we treat IDD and the number of codes selected as fixed variables in separate analyses, we find that both IDD and number of codes are significantly negatively correlated with funding success. However, when we include both IDD and the number of codes as two fixed variables in a single analysis, number of codes is no longer a significant predictor of funding success, only IDD is significantly correlated with funding success (Supplementary Table 2). There is minimal collinearity between number of codes and IDD, as indicated by small value of variance inflation factor (VIF) around 1 (Supplementary Table 2). These results suggest that the number of codes selected does not explain variation in funding success other than the amount explained by IDD.

6.3 Primary research field

Research fields differ in their overall success rates (Extended Data Table 1, Extended Data Figure 1). To control for variation in success rates between major fields of research, we included the

Division (or the Domain) of the primary FOR code as fixed variables. The negative correlation between IDD score and funding success remained significant when either Division (likelihood ratio test, $\chi^2_{22} = 57.94$, $p = 4.5\times10^{-5}$; Extended Data Table 3) or Domain was included (Supplementary Table 2). 'Mathematical Sciences' and 'History And Archaeology' have the highest success rates, but they do not have the most proposals or the lowest IDD values (Extended Data Table 1). Similarly, 'Agricultural And Veterinary Sciences' and 'Built Environment And Design' have the lowest success rate, but they also have the fewest proposals (Extended Data Table 1). The negative correlation between IDD score and funding success rate is not driven by more successful Divisions being less interdisciplinarity or less successful Divisions having a greater degree of interdisciplinarity

6.4 Number of investigators

It may be that interdisciplinary proposals are more likely to succeed if they have more investigators, as it may allow them to draw on expertise in a number of different disciplines. To test if number of investigators is a potential confounding factor to the relationship between interdisciplinarity and funding success, we added number of named chief investigators (CIs) as a fixed variable in the GLMM analysis. Since there is no evidence of non-linearity between funding success and the number of CIs, we assume a linear relationship between them in the analysis. While the number of investigators tends to have positive impact on success, the negative correlation between IDD score and funding success remained significant (Supplementary Table 2). There is minimal collinearity between number of investigators and IDD, as indicated by small value of variance inflation factor (VIF) around 1 (Supplementary Table 2). These two results suggest that number of investigators and IDD have a weak association which does not confound the relationship between interdisciplinarity and funding success rate.

6.5 Institution

Institutions vary in the number of research proposals submitted to the ARC Discovery Program and in their overall funding success rates (Extended Data Table 2). Therefore, we tested whether the primary administering institution for each proposal influences the relationship between interdisciplinarity and funding success (Extended Data Figure 2). Considering all institutions separately would have too little power, so instead we use independently defined groups of universities into higher education networks, which represent alliances of similar institutions (Extended Data Table 2). All other institutions were classified as "unaligned". The overall funding success rate varied between institution networks (Kruskal-Wallis $\chi^2 = 17.08$, df=4, $p = 0.002$). On average, the research-intensive "Group of Eight" (Go8) group of universities had 1.8 times higher funding success than non-G08 universities, according to Fisher's exact test ($p < 2\times10^{-16}$). Although institutional networks differed in their overall success rates and their median IDD, there is still a significant negative relationship between IDD and success when institutional network is accounted for as an additional fixed effect in GLMM (Supplementary Table 2).

6.6 Balance vs disparity

Our IDD scores take into account both the disparity (degree of relatedness between disciplines) and balance (evenness, or the degree to which disciplines are equally weighted) of the FOR codes selected by each proposal. To examine whether it is evenness or relatedness or both that drives the negative correlation between interdisciplinarity and funding success, we conducted two additional tests.

Firstly, we calculated an alternative metric (PSV) that accounts only for relatedness not evenness[19]. We did a GLMM analysis using PSV as the fixed effect (Supplementary Table 2). Results show that interdisciplinarity measured by PSV also has a significantly negative association with funding success (slope=-0.35, $p = 1.7 \times 10^{-10}$), suggesting that discipline disparity is an important factor in the funding success of proposals, independently of the weighting between disciplines.

Secondly, we calculated Pielou's Evenness index (or Pielou's J'), a measure of evenness that does not take disparity into account. To calculate Pielou's J', we calculated Shannon's diversity index then divided it by the natural log of the number of FOR codes selected for each proposal, so that this index only accounts for balance (evenness of weighting between disciplines) not variety (number of disciplines selected). An analysis using Pielou's J' as a fixed effect suggests that the evenness of interdisciplinarity spread is also significantly associated with funding success (with single code proposals: slope=-0.27, $p = 2.1 \times 10^{-7}$; without single code proposals: slope=-0.71, $p = 2.2 \times 10^{-4}$). Moreover, PSV is not correlated with Pielou's J' for proposals with multiple FOR codes (Spearman's $\rho = 0.004$, $p = 0.66$), suggesting that these two metrics successfully separate out the qualities of balance and disparity. These results suggest that both balance and disparity among disciplines contribute to the negative impact of interdisciplinarity on funding success.

7. Effect size

In addition to the regression coefficient (Figure 2), we calculated another metric value to represent the effect size of IDD on success rate. We first predicted the success rate of each proposal by a GLMM analysis with IDD, Division, and their interactions as fixed variables and year as random variable. We then calculated, for each year and each division, the mean and variance of the Cohen's D between the predicted success rate of the funded proposals and the predicted success rate of the unfunded proposals. Last, we calculated, for each division, the weighted average of Cohen's Ds over the 5 years. The weight of each year is the inverse of the variance of Cohen's D of the year. The 5-year average effect size of IDD on success rate in each Division is reported in Extended Data Table 3.

8. Keyword analysis

We performed a keyword analysis as an independent test to verify if IDD measures interdisciplinarity and if the negative correlation between IDD and success reflects the negative impact of interdisciplinarity on success. We identified a proposal as interdisciplinary if its title, abstract, or impact statement contained certain keywords. These keywords include cross discipline; crossdisciplin; cross-disciplin; integrated applied research; integrated assessment; inter discipline; interdisciplin; inter-disciplin; multi discipline; multidisciplin; multi-disciplin; post-disciplin; trans discipline; transdisciplin; trans-disciplin.

There are 597 proposals that had at least one of these keywords in the title or summaries. These proposals have significantly higher IDD than proposals without keywords (one-tailed Wilcoxon rank test, $p = 5.8\times10^{-9}$), indicating that IDD score is positively correlated with text-based identifiers of interdisciplinarity. The percentage of funded proposals with relevant keywords (17.9%) is lower than the percentage of funded proposals without these keywords (20.7%), and this difference in funding success is marginally significant ($\chi_1^2 = 2.61$, $p = 0.053$). This result is consistent with our finding that interdisciplinarity has negative impact on funding success.

Additional References

28 de Montjoye, Y.-A., Radaelli, L. & Singh, V. K. Unique in the shopping mall: On the reidentifiability of credit card metadata. *Science* **347**, 536-539 (2015).

29 Gymrek, M., McGuire, A. L., Golan, D., Halperin, E. & Erlich, Y. Identifying personal genomes by surname inference. *Science* **339**, 321-324 (2013).

30 Felsenstein, J. Phylogenies and the comparative method. *American Naturalist*, 1-15 (1985).

31 Bray, J. R. & Curtis, J. T. An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs* **27**, 325-349 (1957).

32 Clarke, K. R., Somerfield, P. J. & Chapman, M. G. On resemblance measures for ecological studies, including taxonomic dissimilarities and a zero-adjusted Bray–Curtis coefficient for denuded assemblages. *J. Experimental Marine Biology and Ecology* **330**, 55-80 (2006).

33 Clark, T. S. & Linzer, D. A. Should I use fixed or random effects. *Political Science Research and Methods* **3**, 399-408. (2015).