A SEAD NETWORK ANALYSIS OF WHITE PAPERS

François-Joseph Lapointe
Department of Biological Sciences, Université de Montréal

Introduction

Much as been said and written about the two-culture paradigm (Snow 1963) separating the world between artists (humanists) and scientists (Katzir-Katchalsky 1972, Meyer 1974, Root-Bernstein 1984, Alfert 1986, Siler 1996, Grillo 2009, Guillemin 2010). On one side of this debate are those who accept and celebrate this cultural art/science divide (e.g., Lévy-Leblond 2010). On the other side are those who reject it altogether to promote a better integration of artscience practices (e.g., Edwards 2008, Wilson 2010). The White Papers submitted to the SEAD network aim precisely at identifying the roadblocks preventing transdisciplinary (or transcultural) research. As such, they present an insider's view of the collaborative process involving artists, designers, scientists and engineers working alongside on common projects. More importantly, these papers offer a representative sample to test the two-culture model by examining in detail the opinions expressed therein. In this survey, I propose a network analysis of the 40 White Papers submitted to the SEAD Network. If it is true that artists and scientists think differently, the papers authored by artists and scientists should fall in different clusters in the network, with papers co-authored by both artists and scientists falling in between. More precisely, I will test the hypothesis that the papers submitted by artists and scientists are significantly disconnected in the corresponding graph, as predicted by the art/science separation. Rejecting this hypothesis will provide support for the alternative artscience integration.

This meta-analysis, performed by a scientist, is organized as a scientific paper with materials and methods, results, and a discussion. It proposes one of many different ways of comparing and analyzing the content of the White Papers. I have opted here for an "objective" statistical analysis of the data, not to be influenced by my own opinions presented in one of the contributions. In other words, I will let the data speak for themselves. The discussion section will then present an interpretation of the results, with personal comments.

Materials and Methods

This analysis was performed on the basis of the 40 White Papers submitted to the SEAD networks and available on the corresponding website (http://seadnetwork.wordpress.com/). The full list is also presented in the Appendix. For comparing the papers with one another, Word Counter 2.10.1 (http://www.supermagnus.com) was applied, and the list of words with at least

five occurrences in each text was recorded. To do so, all words were treated as case-insensitive, converted to their singular forms, and common words were ignored in the analysis. Following this first step, all similar words were then recoded to a single word stem; the corresponding occurrences were tabulated for each paper. The final dataset thus presents 664 words appearing at least five times in one of the 40 White Papers.

Pairwise intertextual (lexicometric) distances (sensu Labbé & Labbé 2001, Merriam 2002) were computed among all papers in R 2.13 (R Development Core Team 2011) based on presence/absence data (function dist). This index measures the distance as a proportion of words used in one text, but not in both. As such it ignores actual word frequencies to prevent shorter papers to be misrepresented in the analysis with respect to the longer papers (Brunet 2003). A null distance (0) between any two papers means that they share exactly the same words, whereas a maximal distance (1) implies that the two papers are characterized by entirely different sets of words.

The 40x40 distance matrix was then submitted to different types of analysis. For one, a hierarchical clustering algorithm was applied to identify relevant groups in the dataset using the R package cluster. Ward's criterion (function hclust) was selected to minimize within-cluster variance and maximize among-cluster variance (Ward 1963), and the resulting dendrogram was used as a template to identify significant clusters. Furthermore, a k-means partitioning algorithm (Hartigan & Wong 1979) was applied to define the optimal number of clusters in the data (function kmeans) and the corrected Rand index (Hubert & Arabie 1985) was used to compare this partition (function randIndex) with an a priori categorization of papers based on the "cultural status" of the author(s).

Three different categories were used to classify the 40 White Papers. Texts with a single author were coded as either "artist" or "scientist" depending of the self-proclaimed status of the author, or as "artscientist" for authors with a dual status (based on unpublished demographics data). Texts with multiple authors were coded as "artist" or "scientist" only when all co-authors were in the same category; they were coded as "artscientist" in every other instance. This classification was further refined by examining the frequencies of the words *art* and *science* in each paper to categorize hybrid contributions. Out of the 40 White Papers, 16 were coded as authored by artists, 7 by scientists, and 17 by artscientists (see Table 1).

A network was also built from the distance matrices converted into adjacency matrices with the R package igraph (function graph.adjacency) using different cutoff levels. That is, papers (nodes) with more words in common are connected by a link (edge), whereas more distant papers are

disconnected in the graph. In the final representation, a 10 percent resemblance threshold was selected as an interesting cutoff value for building the network. Lower values produced completely connected graphs and higher values produced disconnected graphs without any structure. The nodes were colored in the final graph with respect to the three different categories of papers.

A statistical evaluation of network indices associated with each category of papers was performed. Namely, the three groups were characterized with degree distribution, clustering coefficient, diameter, and density:

The degree of a node is the number of edges connected to that node. The mean degree for a given category thus quantifies the average number of papers connected with a paper from that category.

The clustering coefficient of a node *z* is defined as the probability that two nodes *x* and *y* which are connected to the node *z* are themselves connected (Milenković, Lai, and Pržulj 2008); the average over all nodes *z* of a given category is the clustering coefficient of that category.

The smallest number of edges that have to be traversed to get from a node *x* to a node *y* in a network is called the distance between nodes *x* and *y,* and a path through the network that achieves this distance is called the shortest path between nodes *x* and *y*. The average of shortest path lengths over all pairs of nodes in a network is called the average network diameter (Milenković, Lai, and Pržulj 2008). For a given category, the average diameter is computed by only counting the shortest paths among nodes of that category.

The density of a graph is simply the ratio of the actual number of edges over the maximum possible number of edges in a fully connected graph. For a given category, density is computed by only considering the subgraph with nodes of that category.

The significance of the corresponding statistics was assessed with pairwise Mann-Whitney tests among the various categories. Moreover, these values were also evaluated with respect to a random graph model (Erdős & Rényi 1959) in which all nodes have the same probability of being connected. To do so, 1000 networks with exactly the same number of nodes and edges as the original network were generated (function erdos.renyi.game) and the categories were also assigned at random by permuting the corresponding colors in the random graphs. For each replicate, the degree distribution, clustering coefficient, diameter and density of each category were computed for comparison with the original values. Under the art/science division

hypothesis, nodes associated with each colors should be significantly more clustered and more densely connected than the same colors would be in random graphs.

Results

The hierarchical classification of intertextual distances is presented alongside the network analysis of the 40 White Papers in Figure 1. That joint representation identifies four clusters in the dendrogram (with distinct colors) corresponding to partly overlapping subgraphs (communities) in the associated network. As such, this classification of papers according to a minimum variance criterion or a graph-theoretical approach reveals a congruent structure in the analysis of the dataset using different analytical methods.

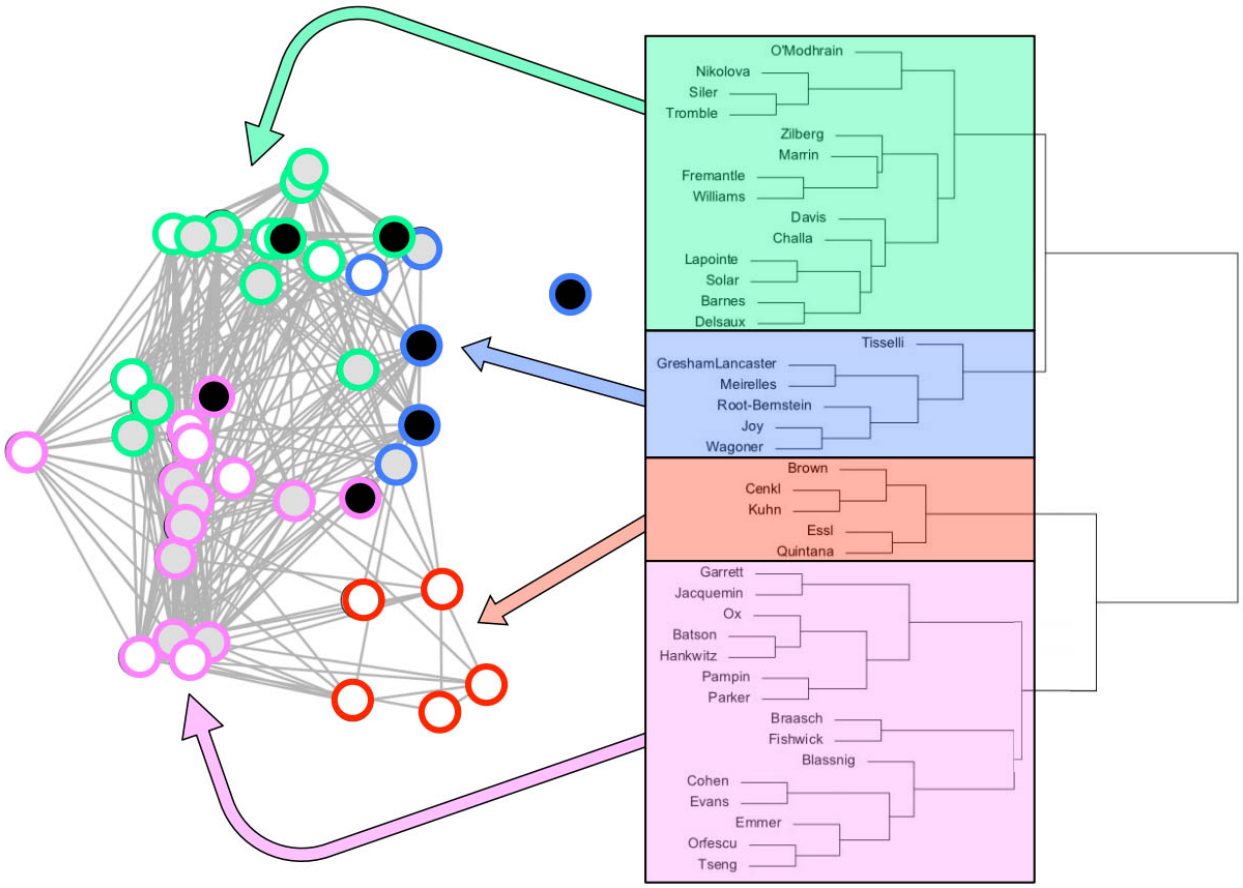

Figure 1. Joint representation of the clustering and network analysis of the 40 White Papers based on intertextual distances. The different colors associated with the four clusters defined on the dendrogram are used to identify the corresponding nodes in the network. The three categories

of papers are also identified on the graph by nodes labeled in white (artist), black (scientist) or grey (artscientist).

More interestingly for the purpose of this meta-analysis is the clustering of nodes representing the three different categories of papers in the network (here depicted in white, black and gray). Under the art/science separation hypothesis, nodes associated with "artists" (white) and "scientists" (black) should be more densely connected within each category than among categories. Moreover, the "artscientists" (gray) should be equally connected to nodes representing either one of the other two categories. This hypothesis was tested by assessing the statistical significance of various network indices for the three different categories.

Results of all pairwise Mann-Whitney tests were not significant, but examination of actual statistics indicates higher degree, clustering coefficient, and density values as well as a smaller diameter for papers assigned to the artscientist category with respect to the artist and scientist categories (Table 1). When these test statistics were compared to those obtained from 1000 random graphs, the artscientist was the only category with observed values more extreme than those expected by chance alone ($p < 0.001$, following Bonferroni correction for multiple tests). In other words, whereas all network indices could not discriminate among papers authored by artists or by scientists, the papers authored by artscientists were clearly more clustered and connected to each other, as well as to other categories.

Table 1. Summary of the network statistics computed from the graph depicted in Figure 1.

Paper Categories	***No. Papers***	***Average degree***	***Clustering coefficient***	***Average diameter***	***Subgraph density***
Artists	16	16.81	0.65	1.64	0.38
Scientists	7	14.15	0.51	2.36	0.29
Artscientists	17	**20.00**	**0.73**	**1.39**	**0.54**

The k-means partition algorithm identified three distinct clusters of papers. The corrected Rand index comparing this objective partition with that defined a priori based on cultural status was statistically significant ($D = 0.067$, $p = 0.048$) with respect to 1000 random assignment of the 40 papers in three groups containing the same number of elements as in the original classification.

Discussion

In this meta-analysis of the 40 White Papers submitted to the SEAD network, I relied on the "scientific method," employing a wide range of statistical and graph-theoretical approaches similar to those I use routinely for the analysis of gene similarity networks (Beauregard-Racine et al. 2011). I intended to look at the "two cultures" from an objective standpoint, testing the corresponding hypothesis that texts authored by artists/scientists would be separated in a network representation of intertextual distances. To my great surprise, this is not what the data said. As a matter of fact, all analyses and statistical evaluation of network indices associated with the different categories of papers revealed an integration of artists and scientists in overlapping groups/partitions. A single cluster in the dendrogram (in red) was formed only with artists. All other clusters included papers representing the three different categories. As such, these results seem to falsify the art/science divide altogether. In terms of word use at least, it is not possible to distinguish papers authored by scientist(s) from those authored by artist(s). Yet, the classification of papers using the k-means algorithm revealed a partition statistically congruent with the cultural status of the author(s) defined a priori. There may well be a (partial) division between artists, scientists and artscientists, but this could to be further evaluated with a fuzzy clustering method allowing for overlapping clusters (Nikhil, Bezdek, and Hathaway 1996).

From a different perspective, however, the network analysis tells an even more interesting story than the nonseparation of artists and scientists. The statistical analysis of graph-based indices exhibited the special status of artscientists, a category different from artists and scientists. That is that, papers authored by hybrid individuals (artscientists) as well as those submitted by co-authors with multiple status (artists + scientists) are significantly more clustered in the network representation that papers authored by artists and/or scientists alone. These contributions to the SEAD Network are characterized by larger clustering coefficients and higher density, among others. This implies that artscientists are probably better at collaborating with each other, but more importantly, that they could also collaborate with artists and scientists.

This paper revealed the power of network analysis for the study of intertextual comparisons. Using presence/absence information, I was able to detect significant patterns in lexicometric data. Yet, this simple analysis is far from being complete. The literature abounds with other types of similarity/distance measures among texts. I opted here for a binary measure, while others have relied on distances that account for word frequencies (Merriam 2002, 2003). Just as for the analysis of ecological data, taking species abundance as opposed to presence/absence might reveal different patterns in the community structure (Hubalek 1982). Likewise, various metrics

may provide different results, particularly when very long texts are compared with shorter ones. In such situations, the longer texts usually end up being clustered together, just because they are more likely to share words. A wide range of corrected indices is available to circumvent this problem (Brunet 2003), but it was not the scope of the present paper to assess the relative performance of different intertextual distances. Alternatively, I could have use a measure of semantic relatedness to compare the papers based on their meanings, not just word contents. A plethora of algorithms are currently being published for doing so in different fields (Pevzner 1992, Budanitsky & Hirst 2006, Ferreira & Couto 2010). In the present case, I am willing to admit that comparing papers using a semantic metric would produce a more precise characterization of intertextual distances, and possibly different outcomes.

Concluding Remarks

As a final note, a caveat is mandatory. The results and conclusion of this meta-analysis are based entirely upon the classification of papers into one of three possible categories – that of the "cultural status" of the author(s): artist, scientist, and artscientist. I suspect that if one were to survey the authors of the White Papers to build such a classification, most of them would probably check "all of the above" if asked about their status. For that matter, I decided for the present analysis to categorize the papers not only based on their author's status, but also their contents. I fully understand and accept that this categorization is somewhat inaccurate and can be improved. Namely, I have voluntarily ignored other categories such as engineer, designer, or humanities, for the sake of simplicity. It is particularly telling to have obtained significant results based on such a crude classification. The White Papers have spoken: artists and scientists are not distinguishable, but artscientists are a different breed – individuals who thrive in transdisciplinary (and paradisciplinary) contexts. This is, of course, what the SEAD network is all about; this meta-analysis provides statistical support for promoting such collaborative endeavors.

Acknowledgments

I would like to thank the White Papers Steering Committee for asking me to perform this meta-analysis. I am especially thankful to Roger Malina for introducing me to the SEAD network, Carol LaFayette for assembling and providing the demographics data upon which the author's status was determined, and Margaux Le Cam for helping out with the network analysis. This work was supported by NSERC grant OGP0155251.

APPENDIX

List of 40 White Papers with corresponding categories assigned for the network analysis

Artscientists	**Artscientist**
Barnes, Steven J., and Carlos Castellanos. 2012. "(Re)Structuring Innovation: Community-Based Wet Labs for Art-Science Collaborations." Evans, Kathryn, and Roger Malina. 2012. "Bridging the Silos: Curriculum Development in the Arts, Sciences and Humanities" Jacquemin, Christian. 2012. "Emergence of New Institutions for Art-Science Collaboration in France and Comparison of Their Features with Those of a Longer Established One." Marrin, D.L. 2012. "Interactions among Scientists/Engineers and Artists/Designers in Developing a Common Language and Unique Perspectives on Today's Challenges." Meirelles, Isabel. 2012. "The Cross-Disciplinary Challenges of Visualizing Data." Nikolov(a), Jennifer. 2012. "Towards a Taxonomy of the Challenges Within Typologies of Collaborations Between Art – Design – Engineering – Science – Humanities – A Practical Guide." Ox, Jack, and Richard Lowenberg. 2012. "SARC (Scientists/Artists Research Collaborations)." Parker, Jennifer. 2012. "The Openlab Network Facilitates Innovative, Creative and Collaborative Research with Art, Community, Design, Technology, and Science at the University of California, Santa Cruz." Root-Bernstein, Robert, and Michele Root-Bernstein. 2012. "The Importance of Early and Persistent Arts and Crafts Education for Future Scientists and Engineers." Zilberg, Jonathan, Barry Kitto, Helen-Nicole Kostis, Linda Long, and Kathryn Trenshaw. 2012. "Can Art Advance Science? A Hypothetical SEAD Experiment."	Batson, Glenna. 2012. "Ex-Scribing the Choreographic Mind—Dance & Neuroscience in Collaboration." Challa, Krishna Kumari. 2012. "Science-Art Interactions in Asia With Particular Reference to India." Emmer, Michele. 2012. "Interdisciplinary Courses, Positions, PhD, in Italy." Hankwitz, Molly. 2012. "Environmental Equity: Enabling Excellence in Media Art and Science in Under-Served Communities." Lapointe, François-Joseph. 2012. "How I Became an Art[Scient]ist: A Tale of Paradisciplinarity." O'Modhrain, Sile. 2012. "Building an Interdisciplinary Research Team." Solar, Myriam. 2012. "Complexity Art: A Pattern of Transdisciplinary Emergent Properties."

Artists	**Artist**
Braasch, Jonas. 2012. "Creative Artificially-Intelligent Agents for the Arts: An Interdisciplinary Science-and-Arts Approach." Orfescu, Cris. 2012. "The Nanoart 21 Project." Pampin, Juan, and James Coupe. 2012. "The Coming of Age of a PhD Program in Digital and Experimental Arts Practice: Lessons Learned and Challenges for the Future." Quintana, Joan, and Jose Quintana. 2012. "How SEAD Network Can Advance Experimental Economics: A Case Study of Innovation and Entrepreneurship in Support of Rural Community and Economic Development." Tseng, Yu-Chuan, and Antoanetta Ivanova. 2012. "Chaos, Computers, and Cyborgs. Developing the Art and Technology Practices in Taiwan." Williams, Roy, Jenny Mackness, and Simone Gumtau. 2012. "Learning Across Cultures."	Brown, Ron. 2012. "Using 'Processing' as a Stimulus for Producing STEAM." Cenkl, Pavel. 2012. "A New Ecology of Learning: Ecological Systems as Pedagogical Models." Cohen, Nathan. 2012. "Bridging the Divide: Collaboration, Communication and Education in art And Science." Davis, Josie E. 2012. "A Case Study in IP Arising in Art/Science Performance Research and Transdisciplinary Collaboration." Delsaux, Jean. 2012. "From Workshop to Academic Laboratory, an Artistic Experience of Transdisciplinarity." Essl, Georg. 2012. "Between Barriers and Prospects: Merging Art Performance and Engineering in Mobile Music Education and Research." Fremantle, Chris, John Mullins, and Donald Urquhart. 2012. "CoRE Challenges: the Artist in Residence Programme at the British Heart Foundation Centre for Research Excellence, Queens Medical Research Institute, University of Edinburgh." Garrett, Marc. 2012. "DIWO (Do-It-With-Others): Artistic Co-Creation as a Decentralized Method of Peer Empowerment in Today's Multitude." Gresham-Lancaster, Scot. 2012. "Data Sonification; An Emerging Opportunity for Graduate Music/Sound Design Departments to Expand Research in an Art and Science Collaboration." Kuhn, Sarah. 2012. "Thinking with Things: Feeling Your Way into STEM."

Scientists	Scientist
Blassnigg, Martha, and Michael Punt. 2012. "Transdisciplinarity: Challenges, Approaches and Opportunities at the Cusp of History." Siler, Todd, and Geoffrey Ozin. 2012. "Cultivating Artscience Collaborations that Generate Innovations for Improving the State of the World." Wagoner, Cynthia L., and Robin Wilkins. 2012. "Process Driven Potentials for Interdisciplinary Learning: Ubeats, a Model for Science and Music Learning."	Fishwick, Paul. 2012. "Learning Computing through Game Experiences." Joy, Anu. 2012. "Mapping Space: Introducing Geographical Information Systems in Indian School Classrooms." Tisselli, Eugenio, Juanita Schlaepfer-Miller, and Angelika Hilbeck. 2012. "Sauti Ya Wakulima: Using Mobile Phones to Make the Voices of Rural Farmers in Tanzania Heard." Tromble, Meredith. 2012. " Vitamin A: A Modest Proposal to Introduce Trace Amounts of Contemporary Art into Research by Preparing Students in Art, Design, Engineering, and Science for Collaborative Creative Work, With the Intention of Saving Earth."

References

Alfert, M. 1986. “Creativity and merit in art and science.” *Leonardo* 19: 323-28.

Beauregard-Racine, J., C. Bicep, K. Schliep, P. Lopez, F.-J. Lapointe, and E. Bapteste. 2011. “Of Woods and Webs: Possible Alternatives to the Tree of Life for Studying Genomic Fluidity in E. Coli.” B*iology Direct* 6: 39.

Brunet, E. 2003. “Peut-on mesurer la distance entre deux textes?” *Corpus* 2. http://corpus.revues.org/30

Budanitsky, A., & Hirst, G. (2006). “Evaluating WordNet-based Measures of Semantic Distance.”
Computational Linguistics, 32:13-47.

Edwards, D. 2008. *Artscience: Creativity in the Post-Google Generation*. Cambridge, MA: Harvard University Press.

Erdős, P., & Rényi, A. 1959. “On random graphs.” *Publicationes Mathematicae* 6: 290-97.

Ferreira, J., & Couto, F. 2010. “Semantic Similarity for Automatic Classification of Chemical Compounds.” *PLoS Computational Biology 6*: e1000937.

Grillo, Y. A. 2009. “When Does Art Become Science and Science Art?” *Leonardo* 42: 103-4.

Guillemin, R. 2010. “Similarities and Contrasts in the Creative Processes of the Sciences and the Arts.” *Leonardo* 43: 59-62.

Hartigan, J. A., & Wong, M. A. 1979. “A K-means Clustering Algorithm.” *Applied Statistics* 28: 100-8.

Hubalek, Z. 1982. “Coefficients of Association and Similarity, Based on Binary (Presence Absence) Data: An Evaluation.” *Biological Reviews* 57: 669-89.

Hubert, L., and Arabie, P. 1985. “Comparing Partitions.” *Journal of Classification* 2: 193-218.

Katzir-Katchalsky, A. 1972. “Reflections on Art and Science.” *Leonardo* 5: 249-53.

Labbé, C., and D. Labbé. 2001. “Inter-textual Distance and Authorship Attribution: Corneille and Molière.” *Journal of Quantitative Linguistics* 8: 213-31.

Lévy-Leblond, J.-M. 2010. *La science (n')e(s)t (pas) l'art: Brèves rencontres.* Hermann, Paris.

Merriam, T. 2002. “Intertextual Distances between Shakespeare plays, with Special Reference to Henry V (Verse).” *Journal of Quantitative Linguistics* 9: 261-73.

———. 2003. “Intertextual Distances, Three Authors.” *Literary and Linguistic Computing* 18: 379-88.

Meyer, L. B. 1974. “Concerning the Sciences, the Arts—and the Humanities.” *Critical Inquiry* 1:163-217.

Milenković, T., J. Lai, and N. Pržulj. 2008. “GraphCrunch: A Tool for Large Network Analyses.” *BMC Bioinformatics* 9: 70.

Nikhil, R. P., J. C. Bezdek, and R. J. Hathaway. 1996. “Sequential Competitive Learning and the Fuzzy c-Means Clustering Algorithms.” *Neural Networks* 9: 787-96.

Pevzner, P. A. 1992. “Statistical Distance between Texts and Filtration Methods in Sequence Comparison.” *Computer Applications in Bioscience* 8:121-27.

R Development Core Team. 2011. “R: A Language and Environment for Statistical Computing.” Vienna, Austria: R Foundation for Statistical Computing. http://www.R=project.org/

Root-Bernstein, R. S. 1984. “Creative Process as a Unifying Theme of Human Cultures.” *Daedalus* 113: 197-219.

Siler, T. 1996. “Creating Artscience.” *Art Journal* 55: 42.

Snow, C. P. 1963. *The Two Cultures: And a Second Look.* Cambridge: Cambridge University Press.

Ward, J. H., Jr. 1963. “Hierarchical Grouping to Optimize an Objective Function.” *Journal of the American Statistical Association* 58: 236-44.

Wilson, S. J. 2010. A*rt + Science Now: How Scientific Research and Technological Innovation Are Becoming Key to 21st-Century Aesthetics.* London: Thames and Hudson.