

# Symptoms of quality

## Assessing expert interdisciplinary work at the frontier: an empirical exploration

Veronica Boix Mansilla

*How does one ascertain the quality of interdisciplinary work, when criteria from individual disciplines do not suffice? Assessment is one of the most important and least understood aspects of interdisciplinary research. An empirical study of interdisciplinary work in five established research institutions, reveals that experts prioritize peer review, journal prestige, citation patterns, and successful patent filing, as indicators of quality interdisciplinary work, while also viewing these indicators with skepticism. Three epistemic criteria to assess interdisciplinary work are revealed: (1) the degree to which new insights relate to antecedent disciplinary knowledge in multiple disciplines involved, (2) the sensible balance reached in weaving disciplinary perspectives together, and (3) the effectiveness with which the integration of disciplines advances understanding and inquiry. These criteria may inform the task of reviewers and evaluators of interdisciplinary research outcomes by attending to some of the unique epistemic demands that this type of work presents.*

Veronica Boix Mansilla is Principal Investigator at Harvard Interdisciplinary Studies Project, Project Zero, 124 Mt Auburn St, 5th Floor, Cambridge MA 02138, USA; Tel: 617 496 6949; Fax: 617 496 9709; Email: veronica\_boix-mansilla@pz.harvard.edu.

**T**HE CHALLENGE IS PRESSING. New forms of knowledge production are giving way to dynamic interactions among areas of expertise, rapidly redrawing disciplinary landscapes (EURAB, 2004; Gibbons *et al.*, 1994; Klein, 1990, 1996; Rhoten, 2003; Sung *et al.*, 2003). At the frontier of research today, interdisciplinary work is the norm and pure disciplinary pursuits are the exception. Yet a re-emerging awareness of interdisciplinary research as a vital form of knowledge production is accompanied by an increasing unease about what is often viewed as the dubious quality of interdisciplinary inquiry (Weingart and Stehr, 2000). Central to this controversy is the lingering challenge of assessing interdisciplinary work.

Research evaluation observers recognize the dilemma. On the one hand, adequate measures of quality are essential to adjudicate between more and less acceptable research proposals and outcomes and to evaluate research programs and policies (Feller, 2002, 2005). On the other, assessment parameters developed within tight disciplinary communities are proving increasingly insufficient to evaluate research that typically expands, retools, or challenges the disciplinary canon (Porter and Rossini, 1985; Travis and Collins, 1991). Impact-based measures that prioritize the capacity of research findings to transform their discipline of origin are desirable but complicated by the fact that interdisciplinary “impacts” are often diffused, delayed in time, and dispersed across diverse areas of study and patterns of citation practice (National Academies, 2005; Lewison *et al.*, 2005).

Perhaps most importantly, commonly used markers of quality — that is, the number of peer-reviewed publications associated with a line of

research; the nature and prestige of journals and funding agencies accepting the work — seem to sidestep the problem of defining substantive quality indicators. In these cases, the decision of what counts as quality work in interdisciplinary research is relegated to the black box of ‘peer review’ on the assumption that ‘peers’ will embody the necessary expertise to certify the correctness of inquiry procedures, establish the plausibility of results, and allocate appropriately the scarce resources of funding, journal space, and special recognition. Under the aegis of interdisciplinarity this assumption is problematized by the need to ensure that ‘peers’ and evaluation panels embody forms of expertise that are fit to examine work characterized by the merging of disciplinary traditions. In interdisciplinary research, Helga Nowotny and others have remarked, scientific peers can no longer be reliably identified (Gibbons, Nowotny, and Scott, 2003).

Studies of knowledge validation of the last three decades have often taken as a common point of departure Robert Merton’s classic characterization of the ethos of science — that is, its imperatives of universalism, communism, disinterestedness, and organized skepticism (Merton, 1942; Zuckermann and Merton, 1971). Researchers have examined how values and norms believed to be binding of the person of science are expressed, enacted, and transgressed in the practical context of research evaluation. For example, some have focused on program officers’, editors’, and reviewers’ credentials, tasks, and appointment mechanisms (Amstrong, 1997; Hammermesh, 1994). Others have examined systemic problems of reliability, accuracy, and fairness of the evaluation process (Campanario, 1998; Langfeldt, 2001; Chubin and Hackett, 1990).

In recent years, a small number of empirical studies have begun to reveal the demands that boundary research evaluation imposes on the process and substance of peer review. Michèle Lamont outlined contrasting epistemic stances that reviewers in the social sciences and the humanities bring to their assessment of research proposals. Moving beyond Merton’s universalist imperative these scholars valued rather than bracketed researchers’ reported subjectivity and potential (exogenous) social or political impact to argue for the relevance of the work. Furthermore, work by Guetzkow, Lamont and Mallard revealed cross-disciplinary differences in the way panelists interpret central evaluative criteria such as ‘originality’. Construed as the production of new findings and theories in the natural sciences, the concept is used by social scientists and humanists to refer to work embodying a new topic, a new approach, an understudied area (Mallard *et al*, 2004; Lamont, 2006).

Studies specifically focused on the assessment of *interdisciplinary* work — research proposals and manuscripts — are scarce (Stokols *et al*, 2003; Van Raan, 2000). Recognizing the socially constructed nature of indicators of quality, studies have

addressed the process by which decisions about quality are made. A recent study of the Academy of Finland quality on assessment practices in disciplinary and interdisciplinary research has underscored the need to identify proper panel expertise and the practical difficulties of doing so. It has also illuminated the calibration process that takes place in multi-disciplinary panels, yielding relatively more consensual quality assessments. Such calibration pertains to the standards of scientific merit applied to a research proposal as well as to the standards of relevance emerging in the group as a part of a “collective assertion of interests” (Bruun *et al*, 2005: 135). Also attending to the process by which expertise is harvested and developed in research assessment contexts, Grit Laudel has documented experimental models of peer review in two research network groups (*Deutsche Forschungsgemeinschaft*). In them, a constellation of actors including the research team, delegates of the funding agency, and independent reviewers collaborate over time in the periodic monitoring and final evaluation of research projects. Supportive interactions enable reviewers to gain the substantive grounding needed for an informed evaluation (Laudel, 2003, 2005).

Complementing these lines of research, empirical studies that focus on epistemic qualities that render a piece of interdisciplinary work acceptable in reviewers’ eyes are necessary, especially given the pivotal role played by peer evaluation in the rapidly growing industry of interdisciplinary inquiries. In this spirit, I introduce an study of experts’ perceptions of ‘quality’ in interdisciplinary research. My findings reveal that researchers systematically (and somewhat reluctantly) rely on *indirect* quality indicators (eg number of patents and publications or type of journals and funding agencies associated with the work). Measures that directly address epistemic dimensions of interdisciplinary work (eg explanatory power, aesthetic appeal, comprehensiveness) were less well articulated and converged on three core qualities: *consistency* with multiple disciplinary antecedents, *balance* of disciplinary perspectives in relation to research goals, and *effectiveness* in advancing knowledge through disciplinary integrations. These criteria may inform the task of reviewers and evaluators of interdisciplinary research outcomes by attending to some of the unique epistemic demands that this type of work presents. The study was part of a larger investigation of interdisciplinary research and education in which our research group examined social, intellectual, and epistemic dimensions of interdisciplinary work.<sup>1</sup>

## Interdisciplinary research defined

In this study, we defined “interdisciplinary research” as a form of inquiry that integrates knowledge and modes of thinking from two or more disciplines (eg history, physics) or established fields of study

(eg ethics, law, the visual arts) to produce a cognitive or practical advancement (eg explain a phenomenon, create a product, develop a method, find a solution, raise a question) that would have been unlikely through single disciplinary means. Three qualities characterize the definition proposed. First, interdisciplinary research is *purposeful* in that it embraces the goal of advancing basic or applied understanding in ways that would have not been possible through single disciplinary means. That is, interdisciplinary research is not an end in itself but a means to explain phenomena, advance categorizations, create methods and instruments, craft products, find solutions, pose new questions.

Second, interdisciplinary research is *disciplined* in the sense that it embodies concepts, tools, methods, theories, findings, and forms of communication developed by disciplinarians and experts in established fields of study, such as physics, biology, law, and anatomy. Such disciplinary inputs and practices have survived the scrutiny of expert communities who grant them provisional acceptability at a particular historical time. In my characterization of this aspect of interdisciplinary work I am less concerned with boundary disputes over disciplinary jurisdictions or distinctions as to whether 'biology' or 'molecular biology' are preferable units of analysis. Instead, I focus my attention on the distinction between validated expert knowledge and less scrutinized everyday common sense claims — since interdisciplinary work as here defined builds on insights stemming from the former, not the latter.

Finally, interdisciplinary research as here defined is *integrative* in that it moves beyond the juxtaposition of disciplinary insights toward their productive articulation. Forms of integration vary greatly in interdisciplinary research. In some cases, analogous phenomena (eg cell mutation and evolution) are examined with a single cognitive tool (eg complexity theory), yielding a more acute understanding of both phenomena and tool. In other cases, disciplines are positioned in complementary roles, with one set of tools (eg a bill of law) leveraging the capacity to engage a phenomenon typically examined by a different discipline (eg genetic testing technologies). Whether interdisciplinary researchers seek to produce multi-causal explanations of socio-biological

phenomena like incest taboo; contextualize a scientific discovery like Einstein's relativity theory in its broader historical époque; or embed macro-social, economic, and environmental phenomena in the micro-cosmos of a case study for in-depth exploration, interdisciplinary research involves a form of synthesis in which the interdisciplinary whole is more than the sum of its disciplinary parts.

The above definition is informed by epistemological characterization of knowledge construction within and across disciplines (Becher and Trower, 2001; Elgin, 1997, 1999; Goodman, 1978; Habermas, 1987; Putnam, 1981), conceptual analyses of interdisciplinarity (Gibbons *et al* 1994, 2003; Gould 2003; Klein, 1996; Kockelmans, 1979; Newell, 1998; Weingart and Stehr, 2000), and empirical analyses of expert interdisciplinary work (Boix Mansilla *et al*, 2002; Boix Mansilla, 2005; Feller, 2002; Guetzcow, Lamont and Mallard, 2004; Klein *et al*, 2001; Latuca, 2001; Nowotny, 2003; Rhoten, 2003).

## Methods

### *Informants*

Fifty individuals working in five interdisciplinary research institutes were interviewed between 2001 and 2002. The Santa Fe Institute (SFI), New Mexico, is a basic research center founded in 1984 to study common themes that arise in natural, artificial, and social systems through lenses such as chaos and complexity theory. The MIT Media Lab (ML), in Cambridge, Massachusetts, was founded in 1980 to study the future of human computer interaction. The Research in Experimental Design group at XEROX-PARC (RED) in Palo Alto, California, was the research division of Xerox Corporation and worked with individuals whose skills ranged from architecture and cultural theory to programming and video production in the design and exhibition of future technologies.<sup>2</sup> The Center for the Integration of Medicine and Innovative Technologies (CIMIT) Cambridge, MA, is a multi-institutional organization that facilitates collaborations among physicians, scientists and engineers to develop innovative medical technologies emphasizing minimally invasive diagnosis and therapy. The Center for Bioethics at the University of Pennsylvania (CB-UP) brings together experts in philosophy, social sciences, law, and life sciences to conduct empirical research in bioethics and inform practice in the life sciences and medicine.

Research centers were selected on four grounds:

1. They reflected a long-standing commitment and accumulated experience (five years or more) in quality interdisciplinary research.
2. Leadership and researchers showed willingness to reflect about the nature of interdisciplinary research and its challenges.

---

**Interdisciplinary research is not an end in itself but a means to explain phenomena, advance categorizations, create methods and instruments, craft products, find solutions, pose new questions**

---

3. Collectively, the centers represented a broad range of disciplinary emphases and combinations (eg history and mathematics, physics and biology, music and computer science).
4. Researchers were dedicated to exploring novel disciplinary combinations as opposed to more traveled and institutionalized paths such as art history, biochemistry, and sociology of science.

I expected that the difficulties associated with developing novel integrations would have engendered a certain epistemological awareness among these researchers — a trait we were interested in capturing. The four criteria (years of experience, reflective stance, diversity and novelty of disciplinary combinations) were also used by senior administrators to propose particular researchers as informants for our study (see Table 1).

*Data collection*

The data corpus for this paper consisted of 50 in-depth, semi-structured interviews. Selected samples of researchers’ work (publications, exhibits, reviews), and institutional documents (Internet homepages, publications describing the centers and brochures) were used as background. Interviews of an average length of 1.5 hours were conducted at the research centers by two interviewers in each case. To prepare for the interviews we familiarized ourselves with each center’s institutional mission and procedures as well as with our informants’ biographies and published work. Interviews were fully transcribed and transcripts reviewed by 90% of the participants. The interview protocol covered organizational, social, and intellectual dimensions of interdisciplinary work. A considerable portion was dedicated to quality assessment, a central area of concern in our study (see sample interview questions in Appendix 1). For example, researchers were asked to describe their current interdisciplinary work in detail and the qualities they appreciated or found problematic about their own work. Researchers were asked to name indicators of quality interdisciplinary work (produced by themselves and others) and describe challenges of quality assessment.

*Analytic strategy*

Claims about quality assessment were preliminarily grouped in three categories: general assessment challenges (What is difficult about evaluating interdisciplinary research?); agents and processes of evaluation (Who evaluates interdisciplinary research and how?); and assessment criteria (What indicators of quality are put forth to assess interdisciplinary research?). A content analysis of the data under each category enabled further grouping of claims into 17 initially tenable new categories representing indicators of quality interdisciplinary work (eg venue prestige, peer acceptance, market use, field impact,

**Table 1. Expert sample by institution and main disciplinary affiliation**

Institution and number of faculty/experts interviewed	Informant	Main disciplinary affiliation of informant
Bioethics, University of Pennsylvania [BioE] N = 6	XUP01	anthropology/communications
	XUP02	history/philosophy
	XUP03	sociology
	XUP04	philosophy
	XUP05	sociology
	XUP06	philosophy
CIMIT N = 7	XC01	engineering
	XC02	medicine (cardiology)
	XC03	physics (medical instruments)
	XC04	medicine (cardiology)
	XC05	medicine
	XC06	medicine (pediatric transplant surgeon)
	XC07	engineering
MIT Media Lab N = 13	XML01	computer science
	XML02	computer science
	XML03	computer science/art
	XML04	linguistics/comparative literature/psychology
	XML05	history/technology in education
	XML06	computer science
	XML07	computer science/science
	XML08	journalism
	XML09	computer science/artificial intelligence/poet
	XML10	engineering (electrical)
	XML11	history/computer science
	XML12	computer science musician (composer + performer)
	XML13	computer science
Santa Fe Institute N = 15	XSF01	physics
	XSF02	biology/genetics
	XSF03	physics
	XSF04	physics
	XSF05	liberal arts/marketing
	XSF06	finance/economics
	XSF07	physics
	XSF08	liberal arts
	XSF09	biology/physics
	XSF10	chemistry
	XSF11	English
	XSF12	history/sociology/public policy
	XAS01	music/physics
XAS02	film making/media (video artist)	
XAS03	music	
Xerox Parc N = 9	XRX01	audio engineering/design
	XRX02	computer science/theater/fine arts
	XRX04	communications research
	XRX05	architecture/computer science
	XRX06	engineering/film/education
	XRX07	engineering
	XRX08	art/poetry
	XRX09	design/technology
	XRX10	writer/artist (“media art”)
	XRX11	music (composition)/art/ engineering

Source: XAS Artist collaborators at Art and Science Lab, Santa Fe, NM

explanatory power, aesthetic appeal, viability, generalizability, coherence, replicability). A clear distinction between primary epistemic measures (eg explanatory power) and secondary level indicators

of quality (eg accepted by peers) emerged. We re-grouped epistemic indicators under categories such as *innovation*, *foundation/standards*, *self-critique*. We grouped secondary level indicators of quality under *productivity*, *prestige*, *acceptance*.

## Results

### Overview

A vast majority of our informants referred to the validation of interdisciplinary research as an obscure and challenging topic. Expressions such as “puzzle”, “struggle”, “opinion-driven”, “not obvious”, “problematic” and “unclear” were used. Researchers referred extensively to the difficulties associated with the assessment of manuscripts for publication (eg determining the relevance, potential impact and acceptability of interdisciplinary research outcomes). They also highlighted difficulties in the evaluation of interdisciplinary grant proposals, and the associated challenge of examining the institutional environment for the work proposed, as well as researchers’ academic trajectories. Describing this last difficulty one informant said, succinctly: “Interdisciplinarity is a luxury of seniority.”

Overall, researchers tended to highlight three sources of difficulty for evaluation directly associated with the interdisciplinary nature of the work. First, they noted that disciplines themselves bring a variety of often-conflicting standards of validation to the interdisciplinary meeting ground. Second, our subjects pointed to a lack of clarity about the nature of interdisciplinary work and its assessment, recognizing the need for a more systematic reflection about it on the part of reviewers. Third, they emphasized that in highly innovative work where novel territories are charted and few precedents are available, developing validation criteria is part of the inquiry process itself. One informant’s description of the decision to try a new medical technology on patients for the first time illustrates the demands. He highlights the challenge (and excitement) of working in uncharted territories, the need to uphold standards that stand beyond formal approval from the Food and Drug Administration, and the tension and

complementarity between engineering and medical risk assessments.

As product developers [of non-invasive medical technologies] we tend to work with early adapters in the clinical world — who enjoy having the first publication describing a new idea as a first in humans. As an engineer without an MD I’ve always been a team member working with the physicians. I typically find myself being the one saying, “I don’t think it is quite ready to try yet.”

How do you decide when it is time to try something in a human patient? You can do 200 dogs or goats but at some point you say okay, we are ready ... We get all the regulatory approval and even if you have the approval it is still a personal decision by an individual physician to try this [technology] on one of his or her patients for the first time. That is a momentous decision and sometimes more than others. Usually the more innovative ... the larger the gap, the greater the risk.

So every time we get to that point I look at the risk benefit curve. I am looking at the prototype. I know all the things that could go wrong with it and might still go wrong, even though I have done everything I can to prevent that from happening. Mine is a different risk benefit analysis than the one of the physician, who is looking at the patient saying: “What are my choices? What would I do if I didn’t use this new thing? How would it affect the patient? What are the patient’s chances? They make the decision, I make the decision and we decide together ... the frames of reference are different but complementary. [XC01]

### Indirect measures of acceptability

Faced with the task of making their assessment criteria explicit, all interviewees (with the exception of two) referred to *indirect measures* of quality, placing greater emphasis on indicators of reception of the work than on its substantive quality. They pointed to indicators such as the *number* of accepted patents, publications, devices, and citations stemming from the work; the *prestige* of the universities, funding agencies, and journals in which it is placed; and the *approval* of peers and a broader community. A member of CIMIT put it clearly:

We have external reviewers (technical or clinical persons) who are usually based in the disciplines that we are using. If anything that we do makes it into a publication, or into patents that make it through the system, that’s an outside sign that the research is functioning well. [XC02]

A researcher at the MIT ML added:

---

**Our subjects pointed to a lack of clarity about the nature of interdisciplinary work and its assessment, recognizing the need for a more systematic reflection about it on the part of reviewers**

---

Getting an award, competing against all the professional designers is one piece of evidence and ... citations are another source of evidence that the work is strong. [XML04]

In multiple cases, quality descriptors of this kind were followed by critical appraisals. As the Director of the Office of Technology Implementation at CIMIT claimed:

Simply counting things are easy answers, as far as I'm concerned — How many patents have you filed? How many patents have been licensed? How many new companies have been started? How many *Science* papers? How many *Nature* papers? We need better measures of quality.

Indirect measures of this kind seemed to side-step the question of what constitutes warranted interdisciplinary knowledge by uncritically relying on institutionalized procedures of peer review, inter-subjective agreement, and ultimately consensus as *generators*, rather than *monitors*, of acceptability.

A new media art expert studying human computer interaction at the MIT Media Lab highlighted the need for substantive reflection about quality eloquently:

The critical audience does not know what to look for. 'Oh, it's got computers, it must be good' ... is the level of critical discourse about this [new media art] work. Unfortunately, museums [as institutions involved in identifying and construing excellence] have been pushed to include this in their collections so we put computers everywhere. But what's the rationale for it? Is it any good? Is it relevant? Is it worthwhile? — has not been answered yet. You have contests that look to crown the latest great integrative art. The entries are so poor so you have to award less than optimal art because that's all there is to choose from. The reality is that [clarifying the rationale for quality standards] has to be done somehow. [XML03]

Often, our subjects criticized these 'proxy' criteria because they saw them as ultimately reducing the assessment of their interdisciplinary work to the criteria of particular disciplines.

Being interdisciplinary means that there aren't any established disciplines or communities to judge. Each faculty or reviewer brings his or her own sense of value to assessing the work. If the work is more clearly situated in a discipline then it is easier to publish the paper. [XML11]

Many interpreted the problem as a difficulty in finding adequate peers to assess their work. As one ML informant proposed:

It is very hard. Sometimes it is hard to get a work peer-reviewed because there are no peers. Yes, people are not doing likeminded or similar work so it is very difficult to get reviewed in non-disciplinary constraining terms. [XML13]

Here the disciplines are so different that there are really few measures of success that are the same across them. Everybody agrees you should have impact, success, be the best in the world, but what constitutes those things is quite different [across disciplinary communities]. [XML09]

Highlighting the need for more communication between researchers and reviewers one member of ML said:

To use the word cautiously, "paradigms" are ill-structured in the interdisciplinary research. So it is still very important to have feedback and assessment and criticism but the way you have to do it is different than you would in do in the hard sciences or disciplinary sciences. So it really has more to do with the interaction and commentary. [XML08].

All interviewees recognized criteria such as productivity, venue prestige, and peer acceptance as the standard way — however flawed — in which the quality of interdisciplinary work is determined at the forefront of knowledge production today. When probed, however, the majority of informants also referred to more *primary* or *epistemic measures* of acceptability — that is, epistemological indicators directly addressing the substance and constitution of the work. Researchers referred to a broad range of epistemological criteria, that is, qualities they look for in good interdisciplinary work (eg experimental rigor, aesthetic quality, fit between models and data, power to address previously unsolved questions in a discipline). We distilled three epistemic principles underlying these criteria and fit to assess interdisciplinary work. They are introduced and illustrated below.

#### *Toward an epistemic framework*

Most interviewees highlighted the complexity of knowledge validation at disciplinary borders. In their view, interdisciplinary findings, theories, or exhibitions were not assessed as a sum of independent claims to be tested against equally independent disciplinary bars. Rather, when addressing the issue epistemologically, researchers tended to provide a dynamic picture of knowledge validation, suggesting three fundamental grounds to examine the quality of interdisciplinary research outcomes:

1. The way in which the work stands in relation to what researchers know and find tenable in the

disciplines involved (*consistency with multiple disciplinary antecedents*).

2. The way in which the work stands together as a generative and coherent whole (*balance in weaving together perspectives*).
3. The way in which the integration advances the goals that researchers set for their pursuits and the methods they use (*effectiveness in advancing understanding through the integration of disciplinary views*).

1. *Consistency with multiple disciplinary antecedents* While the impetus of their interdisciplinary work was to move beyond established disciplinary boundaries, the vast majority of researchers evaluated the degree to which their work was reasonably consistent with antecedent disciplinary knowledge (ie accepted methods, preferred conceptualizations, and epistemic values). They referred extensively to the act of satisfying multiple -sometimes conflicting- disciplinary standards at once. Researchers typically presented the disciplinary arrangements of universities and funding agencies as constraints, but they invariably viewed disciplinary knowledge as providing important tools to advance knowledge.

In researchers' view, borrowed disciplinary theories, methods, and communicative genres embodied epistemic values, which collectively informed the acceptability of interdisciplinary outcomes. For example, seeking to satisfy "two masters", one of our informants at SFI expected that his computer models of political life in Renaissance Florence would meet standards of scientific elegance *and* historical significance. He valued scientific elegance:

What I mean by elegance is the ability to explain coherently, highly heterogeneous phenomena. So the more heterogeneous the phenomenon, the more elegant an argument would be. A classic physicist could have uttered that sentence that I just uttered. Most physicists would agree ... explaining heterogeneity with simple principles. That's really what it's all about.

At the same time he viewed his work as needing to meet standards of historical relevance:

[I seek to] reveal important interactions among people at the time. In a hundred years, will anyone read it? Historians care a lot about that. [XSF012]

The disciplinary canon was often a basic parameter against which researchers assessed their work. If a new finding was consistent with the "the laws of physics" or "current predictions in biology" it gained credibility.

"There is a tremendous sense of freedom associated to breaking [disciplinary] rules" commented a researcher from RED, as he described his

group's work developing experimental reading systems — technologies that seek to understand and inform the phenomenon of reading. At the same time, he added,

You can do a lot of wild things, [but you need] somebody down the hall ... who adheres to the scientific method [and is] squarely involved in the disciplines ... to say, well, this is against the laws of physics. [XR06]

Researchers often depicted the disciplines and fields of knowledge from which they borrowed as a dynamic complex, and not without internal tensions. They did not seek consistency with the whole of the disciplines they embraced but with particular theories, traditions, methods or schools of thought. An collaborator at SFI who conducts micro-recording and non-linear modeling of animal underwater communication, which he eventually also employs in musical compositions. He explained:

I call myself a composer, a sound artist. My personal aesthetic is more a hard core modernist one. I am really hard to convince about traditional concepts of beauty. I am not judging something with classic, not even 20th-century, concepts of beauty. I value the act of exploration itself. I am constantly trying to expand the frame of reference in a way that challenges my own assumption about reality and the world. [XAS01]

When interdisciplinary findings violated fundamental disciplinary tenets or revealed their limitations, additional justification was often seen as required. A ML expert in affective computing explained:

There have been people from the statistical psychology community questioning some of our methods. Having difficulty understanding our methods — too many variables not enough data, why didn't she run the standard ANOVA analysis, and what is the p, and all that. And I say, I don't think these distributions are Gaussian, and the method we are using accounts for these different forms of representation of the data, and we are looking at much harder criteria of classification and accuracy. So the burden is on me to sit down and get a deeper understanding of their methods and show them how their methods do and don't relate to our interdisciplinary methods. In a sense I feel that I have to do my work and their work and show them. [XML09]

Ensuring appropriate fit between interdisciplinary products and findings and their antecedent disciplinary counterparts was not without challenges. Disciplinary communities often conflicted *vis-à-vis* what they considered worth studying and what they viewed as warranted understanding. "What is this

physicist doing writing a sociology proposal?” asked a SFI researcher [XSF07] as he imagined how colleagues in physics would critique his work on social networks. Illustrating differences in validation standards, another researcher at SFI commented:

Computer models are looked down on much more in economics than they are in physics. Mathematical proofs are regarded as much more important in economics than they are in physics. Physicists are more comfortable with approximations [XSF04].

Occasionally, standards stemming from different disciplines appeared as openly incompatible. For instance, the above mentioned Experiments in the Future of Reading was a collaborative exploration among designers, computer scientists, linguists, cultural critics, and artists. Their goal was to develop and test new technologies designed to augment the experience of reading (eg interactive books, image- and sound-enhanced texts). One RED researcher spoke about what he and his colleagues perceived as the conflict between aesthetic and technological dimensions of their experimental technologies. He claimed,

A discipline gives you the economy of value that you can use to make your decisions, when you're faced with multiple choices. And how do you judge when the question is not just which is truer, which is more beautiful, which is more elegant? Actually, one discipline may say, the more beautiful [design option] is always the right choice, Other may say the more elegant is always the best choice ...

We're constantly negotiating what the underlying value is to assign to the different elements of the choices. This [design option] gets us speed, this gets us beauty. For some, speed is always better and if it's beautiful that's great, but it's gravy. Or, if we were artists, speed would be inconsequential. It doesn't matter if it takes five hours to download on a 14.4 modem, it's a beautiful piece of work. These tensions are very common and often you cannot have it both ways. [XRX04]

In sum, while a reasonable fit with antecedent knowledge in multiple disciplines strengthened the credibility of interdisciplinary outcomes, it clearly did not suffice as the sole source of rigor in deeming outcomes acceptable. Quality interdisciplinary understanding did not rest on a sum of established disciplinary rules, but rather on a unique coordination of disciplinary insights where disciplines played particular roles in the overall composition of the work. It is not surprising then, that our interviewees viewed reflective “balance” as a second symptom of quality interdisciplinary work.

---

**While a reasonable fit with antecedent knowledge in multiple disciplines strengthened the credibility of interdisciplinary outcomes, it clearly did not suffice as the sole source of rigor in deeming outcomes acceptable**

---

2. *Balance in weaving together perspectives* Assessing interdisciplinary work involves an appreciation of how disciplinary insights are intertwined and the relatively different roles that they play in yielding an overall composition. When disciplinary values are in conflict, compromises and negotiations are in order. Several interviewees told us that they valued work that exhibited a thoughtful balance of perspectives where “the whole makes sense”, “is coherent”, “is well put together”, “you can see the forest and the trees”. Yet only a few informants were able to articulate this sense of coherence in detail. Reflective balance did not imply an equal representation of disciplines in a piece of work, but a sensible one. Illustrating this point, one of our informants, Arthur Caplan, Director of CBUP, described the relative contribution of law and philosophy in his work. He illustrated how inquiry goals largely determined what counted as a workable balance.

There is some tension between law and philosophy, as to what is the best way to talk, literally [about matters such as organ donation or human cloning]. Should we talk like lawyers and use case precedents and analogical reasoning? Do we use principles? That battle goes on. I think each [view] has a case and I think healthy tension is OK.

Caplan crafted a pragmatic balance.

For certain issues you do want to know the legal framework in which you are operating. And for some other issues like, ‘Should we ban cloning?’ — starting with the law is really not a good idea. For those, you really need to think philosophically about what cloning is and why it would be bad. You can make a law later.

Caplan critiqued work that made legal recommendations “prematurely, before there is consensus about the values” as well as other cases where “there’s a lot of consensus about the values and you don’t need to dig in the same old ethical holes again”.

Relatedly, our subjects referred to finding an appropriate balance *vis-à-vis* the levels of depth at



which various disciplines were engaged. Again in this case, specific inquiry purposes seemed to inform the weighing of options against each other toward a sensible balance overall. In assessing the Experiments in the Future of Reading designs, one RED informant noticed how the inclusion of an animal character in his piece made the technology more effective (ie accessible and interesting to his users).

[the piece] succeeded at being a dog that could read aloud, rather than a computer exhibit. [Such success] allows you to relax some of the more stringent requirements of technology. For example, in terms of engineering, the performance of the reading was not 100% accurate, but it was, after all, only a dog. [XR06]

Several researchers found isolated disciplinary assessments of interdisciplinary work dissatisfactory because they failed to capture the knowledge composition as a whole — a critique often applied to peer-review panels composed of specialists working in isolation. As one informant put it, “Sometimes very good interdisciplinary papers may be viewed in a very negative light simply because narrow disciplinary criteria are used to assess them.”

As our interviewees described it, the interdisciplinary “balancing act” [XUP02] seemed to involve maintaining generative tensions and reaching legitimate compromises in the selection and combination of disciplinary insights and standards. Such a delicately balanced whole gained credibility if it did not violate central tenets of the disciplines involved. It gained relevance and acceptability if it afforded new understandings, solutions, products, and questions — including proposed transformations in established disciplinary practices. Determining the *effectiveness* of the leverage afforded by an interdisciplinary integration was a most informative criterion to ascertain the success of interdisciplinary enterprises — the third principle in our categorization.

3. *Effectiveness in advancing understanding* Not surprisingly, researchers overwhelmingly tended to assess the success of their work in light of the aims of their inquiry. Interdisciplinary inquiries varied broadly in their specific aims, and their favored validation criteria varied accordingly. When SFI informants assessed their mathematical theories of innovation and network behavior respectively, they favored qualities such as their theories’ ability to “*predict* unstudied social and biological phenomena” [XSF03] and their “*tangible success in explaining* something that wasn’t explained by somebody else before” [XSF07]. At CIMIT, the combination of physiology, molecular biology, nano-physics, and material sciences brought scientists closer to the creation of an unprecedented entity — a vascularized artificial human liver that “*works*” and whose creation could have a “*transforming effect*” on organ transplantation surgical practice [XC06].

No single set of assessment criteria can do justice to the enormous variation in inquiry aims.<sup>4</sup> Still it is worth noticing that, among our interviewees, contributions oriented toward pragmatic problem solving and product development seemed to place a premium on standards of *viability*, *workability* and *impact*. “All the original work that I have done professionally is completely driven by my patients’ needs. It has never been because of a piece of science or a piece of technology,” described one CIMIT researcher [XC06], who seeks to address the lack of available human organs for transplantation by advancing the development of an artificial human. Similarly, one CB-UP informant called attention to the effectiveness of solutions (and the need for empirical evidence of impact) as a criterion to assess the work on living will policies in bioethics:

There are philosophers fighting for individual rights by God, the autonomous individual to be fully self-governing even at the time of greatest vulnerability ... They created a whole department in the FDA ... to ensure that every State would have a living will policy, and all of this without ever conducting a study to see if the advanced directives work. Not one! [XUP02]

Contributions that sought formal algorithmic models of complex phenomena seemed associated to measures of *simplicity*, *predictive power*, and *parsimony*. One SFI informant illustrated the point:

There’s a great puzzlement as to how to compare [computer] models with the data. If they fit the data, if they were to fit the data perfectly it would be embarrassing. Why would such a crude model possibly fit the data? But fitting data is usually the best way to judge a theory. [So quality modeling demands that] you look for patterns, regularities, middle level theory, phenomenological principles of some sort that are known to hold or that you discover in the data, and look for those in the model. Try to find a model such that as you continuously proceed from the real situation with greater and greater and greater simplification, these regularities persist. Then you can explain them in the simple model. That explanation might still be valid in the much more complicated reality. [XSF01]

Contributions aiming at a more grounded understanding of multidimensional phenomena (eg lactose intolerance or organ donation viewed in their intertwined biological, cultural, and psychological dimensions) tended to favor work that reached new levels of *comprehensiveness*, *careful description*, and *empirical grounding*. In her analysis of bioethics as a sociological and historical phenomenon, one informant at CB-UP critiqued an account of medical ethics in China for its lack of critical comprehensiveness.

They [researchers focusing on medicine and ethics] made no reference to Confucianism, or to Daoism, no intimation that the two thousand years or more of Chinese history and culture could possibly be in any way involved. They got it wrong [XUP03].

In addition to assessing the substantive leverage afforded by interdisciplinary work, researchers highlighted their methodological contributions. For example, an artificial intelligence expert at ML claimed that his computer models of animal behavior provided novel method for cognitive scientists to ‘test out’ their hypotheses.

Increasingly, computation is going to be a very valuable way to test out models [in psychology and cognitive science]. Because it is one thing to write a book and say this is how it [animal intelligence] must be organized. The proof is, could you take those ideas and actually implement them? [XML08]

Enhanced methodological options, in turn, raised the standards for the interdisciplinary inquiry that followed. “What we used to do in the past are now things that 16-year-olds can do on the Internet,” explained ML expert on affective computing. “We were the only ones doing that 10 years ago.”

Because of their non-canonical approaches to knowledge production, our interviewees often confronted the challenge of a lack of precedents or viable contenders against which to compare their achievements. Working in uncharted terrains implied that “there is no higher authority to appeal to adjudicate what’s relevant knowledge and what’s not or what is useful and not,” claimed a scientist at RED. “We don’t know if we are doing better than people working by themselves [in mono-disciplinary contexts] ... we don’t have these kinds of measures,” noted one CIMIT director. Another CIMIT researcher described the problem of innovation compellingly:

[When] you are at the cutting edge of anything, by definition you’re taking risks that most do not take. So having somebody who can, in an expert way, help you is problematic because there’s this vested interest problem, where the status quo and building on the status quo is most of what goes on. [XC06]

For these researchers, the effective advancement of interdisciplinary understanding involved not only developing new insights and methods, but also fashioning criteria with which to gauge their progress.

## Conclusion

Arguably, all research, interdisciplinary or not, strives for standards of disciplinary robustness,

coherence, and innovation. However, under closer scrutiny, interdisciplinary work poses a qualitatively new challenge to the organization and evaluation of knowledge production. Our findings suggest that while researchers in the institutions that we visited deemed a piece of work acceptable if it survived the test of peer review, journal prestige, citation patterns, or successful patent filing, they also viewed these procedures with a certain degree of skepticism. Recognizing the unique demands associated with the evaluation of interdisciplinary work, science policy makers and journal editors are experimenting with new procedural approaches. For example, in the USA, the National Institute of Health’s *Roadmap* initiative, has begun to include ‘interpreters’ in their multidisciplinary review panels — individuals able to understand and negotiate cross-disciplinary differences emerging in the group. Others are increasing the use of small ad hoc expert committees enhanced with videoconferencing capabilities; inviting grant and manuscript authors to propose reviewers and interact with their evaluators to clarify aspects of the work (National Academies, 2005; Feller, 2002, 2005; Laudel, 2001a, 2001b, 2003). These efforts are geared to harvesting and developing adequate reviewer expertise. The findings here presented complement such efforts by drawing attention not to the ‘procedural how’ but to the ‘epistemic what’ of interdisciplinary research evaluation — thus sharpening the focus of discussions.

As the informants in our study portrayed it, quality interdisciplinary understanding does not rest on an accumulated set of established disciplinary or interdisciplinary rules and standards. Instead, each piece of interdisciplinary work reveals an idiosyncratic coordination of disciplinary insights geared to accomplish researchers’ cognitive and practical goals. Distilling workable criteria to assess the epistemic dimensions of interdisciplinary work requires that we tackle the problem at a productive level of analysis. Criteria too local (eg innovative experimental methods, rich original sources) will fail to account for the formidable diversity of aims and approaches legitimately characterized as ‘interdisciplinary’. Categories too generic (eg validity, coherence, accuracy, parsimony) will be ill-fit to capture the particular challenges associated with interdisciplinary integration.

---

**Recognizing the unique demands associated with the evaluation of interdisciplinary work, science policy makers and journal editors are experimenting with new procedural approaches**

---

The three assessment categories introduced build on interdisciplinary researchers' own views on quality assessment — a perspective that can inform evaluation practice in productive ways by shaping review guidelines (see examples in Appendix 2). Further research is needed, however, to examine how reviewers interpret indicators of quality in interdisciplinary work across a range of domains of study and disciplinary combinations, and the relative weight that they attribute to them in the process of deliberation. Productive research in this area will stand at the crossroads of cognitive psychology, epistemology, sociology, and science policy to shed light on the tacit expertise embodied in experienced evaluators of interdisciplinary work, how such expertise plays out in review panel negotiations, and what kinds of selection outcomes it yields.

Greater clarity about central dimensions of interdisciplinary work to be assessed may enhance the validity of evaluative claims and lessen the chances that proposals and manuscripts are assessed solely in fragmented disciplinary ways. Yet even if the focus of quality assessment were to be agreed upon, disagreements are to be expected and valued as importantly contributing to the advancement of knowledge. Such disagreements are the result of legitimate differences in views among experts about what counts or should count as quality work, or

which lines of research are considered relevant. Deliberations during panel reviews serve not only to certify acceptable work and make the work better (eg by filtering errors, suggesting new methods), but also to calibrate standards in the research community. In this sense each review embodies an opportunity to advance a collective sense of the core principles by which we judge interdisciplinary work, even if we disagree on particular details.

In the end, while the assessment categories emerging from our study might contribute to a more reasoned and reasonable consideration of interdisciplinary work, they will not render interdisciplinary work immune from “the unfortunate propensity for error” that characterizes human knowledge construction (Elgin, 1999:12). Indeed, interdisciplinary work gains its strength from its keen awareness of the provisional epistemic status of its findings. A serious assessment of interdisciplinary work should not seek to establish “warranted truths” nor, on the contrary, to let “all interdisciplinary flowers bloom”. Such assessment should instead yield illuminating evidence to grant provisional credibility to the work in question. Thus the acceptance of an interdisciplinary insight (much like that of the framework here proposed) rests on the assumption of the inherent provisional nature of understanding and the endless human capacity to “retrench, retool, and try again” (Elgin, 1999:12).

### Appendix 1. Sample questions from semi-structured interview protocol

*Note:* The questions below were selected from sections of the interview protocol that are relevant to quality assessment. Excluded are questions about the researchers' background, organizational culture, collaborations and future directions of the work.

- What are you trying to accomplish in your work right now? What problem are you currently studying? Could you describe how you do your work?
- Do you view your approach to this problem as interdisciplinary? If so, what disciplines inform your work? Explain
- How do these disciplines contribute to advancing your work? Explain
  - How easy or difficult was it for you to distinguish disciplinary contributions in your previous description?
  - Why?
  - Do you normally distinguish between the disciplinary and interdisciplinary dimensions of your own work?
- What can you do with your interdisciplinary approach that you could not do if you used only a D1 or D2 approach?
- Have you ever had difficulty in blending elements of different disciplines in your work?
  - If so, would you describe the situation?
  - How did you resolve the conflict?
- How do you evaluate the work that you do? What standards do you (or others) use to judge your work or that of others?
  - Do the disciplines of D1 and D2 contribute to these standards? If so, how?
  - Do you think there are new forms of interdisciplinary standards? If so, what are they?
  - People say, “You can recognize quality work when you see it.” How can you tell when you see good interdisciplinary work?
- What is difficult about assessing interdisciplinary work?
- Where do you go for informed peer critique and commentary (inside and outside the institution)? Has this been a problem?
- Do you ever have to consult with disciplinary experts in order to get your work done? Please offer an example.
- What journals do you read?
- What criteria do they use to judge papers?
  - Do these criteria work?
- Have you seen interdisciplinary ventures or research fail? What make you think they failed and what do you think led to failure?

## Appendix 2. Sample guideline questions

The questions below briefly illustrate how the epistemic merit of interdisciplinary research might be assessed based on the categories introduced. Additional relevant assessment categories such as *significance* of the problem under study, *research environment*, and *researchers' background* are not included in this selection.

### Purpose

To clarify the aims and interdisciplinary scope of the work and calibrate assessment, reviews may begin by asking: What is the purpose of the work proposed? Does it call for an interdisciplinary approach? How?

#### I. Consistency with multiple disciplinary antecedents

Is the selection of frameworks, methods, designs, tools, analyses, concepts from multiple disciplines adequate to address the purpose of the work? Are such frameworks tools and insights employed or developed in accordance with disciplinary standards; or explicated if used in 'non-traditional' ways to transform such standards?

#### II. Balance in weaving together perspectives

Does the work hold together as a coherent and productive whole — where theories, methods, approaches stemming from different disciplines are well integrated, and their relative weight adequately considered to advance the purpose of the work? Or does the inclusion of theories methods insights and perspectives seem artificial, or excessive — beyond the point of diminishing returns

#### III. Effectiveness in advancing understanding

Does the work develop or propose a novel integrative framework, solution, explanation, approach or description that would have been unlikely through single or isolated disciplinary means? Do such resulting framework, explanations and so on challenge existing paradigms and practices, address an innovative topic, put forth a promising approach, tools, method in this field in ways that would have been unlikely through single disciplinary means? Does the work introduce the criteria by which its contribution should be gauged and its limitations considered?

## Acknowledgements

An earlier version of this paper entitled "Assessing expert interdisciplinary work: symptoms of quality" by Veronica Boix Mansilla and Howard Gardner was presented on 1 December 2003 at *Rethinking Interdisciplinarity*, a virtual seminar supported by the CNRS project Society of Information, and moderated by Gloria Origi and Christophe Heintz. I would like to thank our informants in the study for their time and illuminating insights, participants in the seminar for their helpful analyses of our findings, and the Atlantic Philanthropies for its generous funding of the Harvard Interdisciplinary Studies Project.

## Notes

1. For further information see <[www.pz.harvard.edu/interdisciplinary/](http://www.pz.harvard.edu/interdisciplinary/)>, last accessed 10 May 2006.
2. RED closed operations in 2003.
3. See <<http://www.parc.xerox.com/about/pressroom/xfr/>>, last accessed 10 May 2006.
4. See also Crutchfield and Schuster, 2003; Caplan *et al.*, 2002; Durham, 1991; and Machover, 2004 for examples of variation in quality standards at work.

## References

- Armstrong, J S 1997. Peer review for journals: evidence on quality control, fairness, and innovation. *Science and Engineering Ethics*, 3, 63–84.
- Becher, T and P Trower 2001. *Academic Tribes and Territories: Intellectual Enquiry and the Culture of Disciplines*. 2nd edn. Philadelphia: Society for Research into Higher Education and Open University Press.
- Boix Mansilla, V 2005. Interdisciplinary work at the frontier: an empirical examination of expert interdisciplinary epistemologies. Article in preparation.
- Boix Mansilla, V, D Dillon and K Middlebrooks 2002. *Building Bridges across Disciplines*. Harvard Graduate School of

- Education, Project Zero <<http://www.pz.harvard.edu/interdisciplinary/pdf/BuildingBridges.pdf>>, last accessed 30 December 2005.
- Bruun, H, J Hukkinen, K Huutoniemi and J T Klein 2005. *Promoting Interdisciplinary Research: The Cast of the Academy of Finland*. Publications of the Academy of Finland. Helsinki: Academy of Finland.
- Campanario, J M 1998. Peer review for journals as it stands today: Part 1 and Part 2. *Science Communication*, 19(3), 181–211 and 19(4), 277–306.
- Caplan, A L, D Magnus and G McGee eds. 2002. *Who Owns Life?* Amherst, NY: Prometheus Books.
- Chubin, D E and E J Hackett 1990. *Peerless Science: Peer Review and US Science Policy*. Albany, NY: State University of New York Press.
- Crutchfield, J P and P Schuster eds. 2003. *Evolutionary Dynamics: Exploring the Interplay of Selection, Accident, Neutrality, and Function*. New York: Oxford University Press.
- Durham, W 1991. *Co-evolution: Genes, Culture and Human Diversity*. Stanford, CA: Stanford University Press.
- Elgin, C 1997. *Between the Absolute and the Arbitrary*. Ithaca: Cornell University Press.
- Elgin, C 1999. *Considered Judgment*. Princeton: Princeton University Press.
- EURAB, European Union Research Advisory Board 2004. Interdisciplinarity in research. EURAB, April <[http://europa.eu.int/comm/research/eurab/pdf/eurab\\_04\\_009\\_interdisciplinarity\\_research\\_final.pdf](http://europa.eu.int/comm/research/eurab/pdf/eurab_04_009_interdisciplinarity_research_final.pdf)>, last accessed 10 May 2006.
- Feller, I 2002. Performance measurement redux. *American Journal of Evaluation*, 23(4), 435–452.
- Feller, I 2006. Multiple actors, multiple settings: multiple interpretations of quality as a criterion for assessing interdisciplinary research. This volume, pp. 5–15.
- Gibbons, M, C Limoges, H Nowotny, P Scott, S Schwartzman and M Trow 1994. *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies*. London: Sage.
- Gibbons, M, H Nowotny and P Scott 2003. "Mode 2" revisited: the new production of knowledge. *Minerva*, 41, 179–194.
- Goodman, N 1978. *Ways of World Making*. Hassocks, Sussex: Harvester Press.
- Gould, S J 2003. *The Hedgehog, the Fox, and the Magister's Pox: Mending the Gap between Science and the Humanities*. New York: Harmony Books.
- Guetzkow, J, M Lamont and G Mallard 2004. What is originality in

- the humanities and the social sciences? *American Sociological Review*, **69**, April, 190–212.
- Habermas, J 1987. *Knowledge and Human Interest*. Cambridge: Polity.
- Hammermesh, D S 1994. Facts and myths about refereeing. *Journal of Economic Perspectives*, **8**(1), 153–163.
- Klein, J T 1990. *Interdisciplinarity: History, Theory, and Practice*. Detroit, MI: Wayne State University.
- Klein, J T 1996. *Crossing Boundaries: Knowledge, Disciplinarity, and Interdisciplinarity*. Charlottesville, VA: University Press of Virginia.
- Klein, J T, W Grossenbacher-Mansuy, R Häberli, A Bill, R W Scholz and M Welti eds. 2001. *Transdisciplinarity: Joint Problem Solving among Science, Technology, and Society: An Effective Way for Managing Complexity*. Basel: Birkhäuser Verlag.
- Kockelmans, J J ed. 1979. *Interdisciplinarity and Higher Education*. University Park: Pennsylvania State University Press.
- Lamont, M 2006. *Cream Rising: What Defines Excellence in the Social Sciences and the Humanities*. Manuscript in preparation.
- Langfeldt, L 2001. The decision making constraints and processes of grant peer review, and their effect on the review outcome. *Social Studies of Science*, **31**(6), 820–841.
- Lattuca, L R 2001. *Creating Interdisciplinarity: Interdisciplinary Research and Teaching among College and University Faculty*. Nashville, TN: Vanderbilt University Press.
- Laudel, G 2001a. Collaboration, creativity and rewards: why and how scientists collaborate. *International Journal of Technology Management*, **22**(7–8), 762–781.
- Laudel, G 2001b. What do we measure by co-authorships? *Proceedings of the 8th International Conference on Scientometrics and Informetrics, Sydney, Australia, Vol. 1*, pages 369–384.
- Laudel, G 2003. Are there special criteria for assessing interdisciplinary work? Response posted at the *Rethinking Interdisciplinarity* virtual seminar. December. <<http://www.interdisciplines.org/interdisciplinarity/papers/6/5#>>, last accessed 16 December 2005.
- Laudel, G 2006. Conclave in the tower of Babel: How peers review interdisciplinary research proposals. This volume, pp. 57–68.
- Lewis, G, I Rippon and S Wooding 2005. Tracking knowledge diffusion through citations. *Research Evaluation*, **14**(1), 5–14.
- Machover, T 2004. *Toy Symphony*. <<http://www.toysymphony.net>>, last accessed 10 May 2006.
- Merton, R 1942. *The Normative Structure of Sciences in the Sociology of Science*. Chicago: University of Chicago Press.
- Mallard, G, M Lamont and J Guetzow 2004. Epistemological pluralism and fairness in peer reviews: evidence from the social sciences and the humanities. Under review at *AJS*, Harvard Department of Sociology.
- National Academies 2005. *Facilitating Interdisciplinary Research*. Washington DC: National Academies Press.
- Newell, W H ed. 1998. *Interdisciplinarity: Essays from the Literature*. New York: College Board.
- Nowotny, H 2003. The potential of transdisciplinarity. Paper presented at the *Rethinking Interdisciplinarity* virtual seminar. May. <<http://www.interdisciplines.org/interdisciplinarity/papers/5>>, last accessed 2 January 2006.
- Porter, A L and F Rossini 1985. Peer-review of interdisciplinary research proposals. *Science, Technology and Human Value*, **10**, 33–38.
- Putnam, H 1981. *Reason Truth and History*. New York: Cambridge University Press.
- Rhoten, D 2003. A multi-method analysis of the social and technical conditions for interdisciplinary collaboration. *National Science Foundation, Final Report BCS-0129573*. September. <<http://www.hybridvigor.net/publications.pl?s=interdis>>, last accessed 16 December 2005.
- Stokols, D, J Fuqua, J Gress, R Harvey, K Phillips, L Baezconde-Garbanati et al 2003. Evaluating transdisciplinary science. *Nicotine and Tobacco Research*, **5**(1), 21–39.
- Sung, N, J I Gordon, G D Rose, E D Getzoff, S J Kron, D Mumford et al 2003. Educating future scientists. *Science*, **301**, September, 1485.
- Travis, G D L and H M Collins 1991. New light on old boys: cognitive and institutional particularism in the peer review system. *Science, Technology, and Human Values*, **16**, 322–341.
- Van Raan, A F J 2000. The interdisciplinary nature of science, theoretical framework and bibliometric empirical approach. In *Practising Interdisciplinarity*, eds. P Weingart and N Stehr. Toronto, Ontario, Canada: University of Toronto Press.
- Weingart, P and N Stehr eds. 2000. *Practising Interdisciplinarity*. Toronto, Ontario, Canada: University of Toronto Press.
- Zuckermann, H and R K Merton 1971. Patterns of evaluation in science: institutionalisation, structure, and functions of the referee system. *Minerva*, **9**(1), 66–100.

Copyright of Research Evaluation is the property of Beech Tree Publishing and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.