

**Towards Evaluation as a ‘Public Craft’<sup>1</sup>  
and Evaluators as Stewards of the Public Good**  
**or**  
**On Listening Well**

*Jennifer C. Greene*  
*Associate Professor*  
*University of Illinois, Urbana –Champaign*

*Keynote Address presented at the 2002 Australasian Evaluation Society International Conference  
October/November 2002 – Wollongong Australia. [www.aes.asn.au](http://www.aes.asn.au)*

At this dawning of the new millennium, program and policy evaluation are widely practiced in such important societal domains as economic development, education, environmental management, family welfare, and community well being. But, evaluation rarely fulfills its own potential to be a *powerful* contributor to the health and vitality of these and other critical domains of social life. This is because evaluation is commonly viewed – and I will argue inappropriately so – as an activity practiced at a distance, apart from the contentious fray of political decision making about resources and priorities, in a space somehow shielded from special interests and advocacies, a bit like a little air bubble safe from contamination by bias and disease alike (Mark, Henry, and Julnes, 2000). From this protected space, evaluation smoothly contributes to the many initiatives of the New Public Management movement that has swept westernized democracies over the past decade (and energized or traumatized them, depending on your point of view). You are all familiar with this movement – some of its mantras are:

- Accountability! Just do it!
- Performance measurement for all!
- Succeed in *your* office with results-based management!
- Evidence-based decision making – the latest in what’s hip and cool! (See [www.ebd.com](http://www.ebd.com) now!)

Evaluation within the ideology of the New Public Management is focused on observable, often quantified, hopefully objective indicators of what *difference* programs are making in the lives of participants. And it is focused on grounding decisions in scientific evidence. These are worthy and important aims. But, they are limited ones, precisely because they are constructed from the ostensible safety of the air bubble. They are aims that lack democratic engagement and thus political potency.

In my remarks this morning, I intend to critique this understanding of evaluation and its authentication of selected premises of the New Public Management movement, *and* to offer an alternative vision for evaluation. My remarks will focus on four major points. The first two of these points capture my conceptual argument and the second two my re-envisioned evaluation practice. Here is an overview of these four points.

1. Evaluation *cannot* position itself on the sidelines of political decision making, in a protective air bubble that is neither influenced by nor itself influences the political currents and dynamics of a particular context. This is because evaluation is partly constitutive of these currents and dynamics.
2. So, evaluation cannot be a bystander of democratic discourse; it is inevitably a player therein. And so, strong, worthy, powerful evaluation is conducted – not via the dispassion and detachment of bystander status, but rather via active engagement with democratic politics, especially the politics of difference.
3. A mixed-method way of thinking about evaluation is especially well suited to active engagement with difference towards fulfillment of democratic ideals.
4. A mixed-method way of thinking further serves to position evaluation in society as a “public craft” and to locate evaluators as stewards of the public good, whose *main* task is *not* to generate unbiased truth claims, but rather to advance a stronger community, to build a better society.

Let me move now to the conceptual part of my argument.

---

<sup>1</sup> From Boyte, H.C. (2000)

## **On popping our protective air bubble and becoming politically engaged evaluators**

My good friend and colleague Gary Henry recently asserted:

To stand up in the contentious environs of modern democracies, evaluation must be perceived as producing credible information that is not unduly biased by its sponsor or by bad decisions of the evaluator. . . . Evaluations that tow the line of less ideology and more scientific rigor have had notable influence [in the policy arena]. (Henry, 2001, p. 421, 422)

Gary Henry speaks for many when he urges the community of evaluators to strive for unbiased work, for credibility. Absent credibility, he cautions, our work will not be trusted and therefore will not be influential in policy arenas. Moreover, the surest route to credibility, says Gary, is through dispassionate neutrality, best achieved from a position on the sidelines or in our protected air bubble free from contaminations of all sorts.<sup>2</sup> My responses to Gary's stance comprise the first two points of my remarks. And let me acknowledge at the outset that this is not a new discussion – we have long debated the political dimensions of our profession. But, as we continue to be divided and troubled by these concerns (Mark, 2001), this is a discussion we need to continue to have.

### First . . . Popping our protective air bubble

The essence of my argument is as follows. I believe that “evaluations are not [and cannot be] neutral and innocent technical activities” (Dahler-Larsen and Krogstrup, 2000, p. 1). Rather, they “take place within particular authority structures and cultures” (House, 1993, p. x), and are both shaped by and, in turn, help to shape these structures and cultures, along with their corresponding norms of action and interaction (see also Valovirta, 2002). So, it is not possible for evaluators to assume a position on the sidelines or in a protected air bubble in the hopes that our practice will not perturb the situation or influence it via some form of unwanted bias. Rather, evaluation – like all forms of social inquiry – is at least partly constitutive of the context in which it takes place, particularly of the institutional and interpersonal relationships of power, authority, and voice in that context.

Now I would like to briefly elaborate on these ideas – ideally to persuade you of their merit but, failing that, to provoke the best of your counter-arguments. The idea that the practice of social science can influence the phenomena being studied has been around a long time. For example, I recall from my own graduate school days of yesteryear studying the infamous Hawthorne effect. This effect arose during a study of changes in working conditions in the Hawthorne plant of Western Electric in Cicero, Illinois. It refers to documented observations that some factory workers knew about the intervention, management's interest in its intended effects, and the study thereof. With this knowledge, the workers changed their behavior to conform to the desired effects, thus entangling the effects of the intervention with the effects of the study of the intervention. Students of social science are also familiar with challenges that arise from the reactivity of measurement. Reactivity refers to the idea that the very practice of measuring a given phenomenon, say knowledge of how HIV/AIDS is transmitted, can itself change that phenomenon, so that those measured have different kinds or levels of HIV/AIDS knowledge resulting from the measurement process alone.<sup>3</sup> We studied this as well back when I was in graduate school. More currently, all evaluators and applied social researchers who have used open-ended interviews in their work are aware that good interviews invoke reflection on the part of our respondents, so that their thoughts about the meaningfulness of their program experiences actually change and develop through the process of asking about them, especially if the questioning takes place over a period of time. So, it is not controversial to state that the social process of inquiring about a phenomenon can alter or change that phenomenon in particular ways.

Nor is it controversial to claim that the context of inquiry, in turn, can affect what is being studied, how, and with what results. There is a large body of research on matching interviewer and interviewee characteristics in order to maximize validity. And the more contemporary form of this research asks the question, who can legitimately and meaningfully study whom (Fine et al., 2000; Hood, 2001; Wong,

---

<sup>2</sup> Gary and his colleagues Mel Mark and George Julnes have elaborated upon this position in their new “integrative” evaluation theory, which I find admirable and highly defensible in most respects save this and one or two others. See Mark, Henry, and Julnes, 2000, and see Greene and Walker, 2001 for a review of this new book.

<sup>3</sup> Measurement reactivity can ostensibly be addressed by a Solomon 4-group design.

1998)? Moreover, the evaluation field itself is keenly attuned to the importance of context, to organizational norms for behavior, to the influence of administrative hierarchies on what is revealed and what remains undisclosed, and to the importance of tailoring evaluation designs and methods to the footprints of particular contexts.<sup>4</sup>

So, why does it remain controversial to claim that evaluation cannot assume a neutral, non-politicized bystander position, cannot be protected from idiosyncratic predispositions of the evaluator *or* the context? Why do the specters of bias and advocacy haunt us so (see Greene, 1997)? Why do so many in our evaluation community share Blaine Worthen's recent lament?

I continue to believe [that] to be useful, [evaluation must be objective, meaning] the evaluator must be sufficiently neutral, nonpartisan, and dispassionate about that which is evaluated to avoid unrecognized biases from coloring his/her portrayals and evaluative judgments in ways that alter facts or dilute reality. (Worthen, 2001, p. 414)

Clearly, the challenge here is to evaluation's cherished position of political neutrality. Our traditional job indeed has been to help settle political disputes via empirical study that yields unbiased truth claims. And our traditional concerns about bias – the Hawthorne effect, reactivity, interviewer characteristics, and so forth – are indeed concerns about the epistemic quality of the knowledge generated in evaluation, concerns about the truth value of our claims to know.

Yet, as a long line of argument has demonstrated, values, ethics, and politics are intertwined with, rather than separable from epistemology and knowledge generation (Schwandt, 2002). Michael Scriven and Ernie House have urged us for decades to reject the fact-value dichotomy and the fiction that social science can be value free. And Tom Schwandt has challenged us for just as long to understand that knowledge generation is not a neutral activity, but rather is both constituted by and constitutive of the values, norms, ethics, and relationships of power and privilege in the context in which it takes place. This is to say that evaluative results *cannot* be generated from nor represent a neutral position within the politics of the contexts in which we work. So, like it or not, the practice of evaluation itself either sanctions and reinforces, or alternatively challenges and disrupts, key dimensions of these contexts, notably:

- who has the right to be heard about what
- what counts as legitimate knowledge
- how decisions are made – who participates, what happens publicly and what happens behind the scenes
- what factors or criteria are valued in making decisions, and who gets to determine these
- and also, the ways that people relate in a given context – with trust or suspicion, respect or disregard, reciprocity or selfishness, caring or neglect

Let me offer a snapshot as an illustration of these ideas. This comes from my own work.

***Vignette #1. Evaluating the Grandview High School Science reform***

\* In 1995, the Grandview High School science department initiated a major reform of their curriculum and instructional practices, a reform fully in sync with national recommendations for more problem-based and active learning, and for cooperative, heterogeneous learning groups. The curriculum reform was not controversial. However, the shift to more heterogeneous grouping in science classes – that is, a shift away from academic streaming – was met by considerable opposition from well heeled, well educated, politically powerful high school parents. These vocal parents were concerned that their own children's science competitiveness would decrease and they desired an evaluation that would assess whether or not students' science learning, as best captured in standardized test scores, was suffering as a result of the reform. Not heard at the public forums held for this science reform were the opinions or points of view of parents of other kinds of students – students of color, students from low-income families, students from rural areas.

\* A bystander evaluation in this context would seek to remain apart from the political fray and just provide the facts. But what facts? Standardized science achievement test scores, as desired by the opposition? Points

---

<sup>4</sup>In the spring of 2000 I conducted a small study on the major influences on evaluators' practice decisions with a group of graduate students. Our respondents clearly reported that their practice decisions were most powerfully shaped by their desire to be responsive to contextual demands.

of view of the silent parents? Experiences of students in their new science classes? Faculty concerns and experiences?

\* Just deciding what questions to pursue in this contentious evaluation context would serve to shape the consequent discourse and debate, and thus the relationships of power and privilege therein. Focusing the evaluation on standardized student achievement as the political elite desired would not change this context much in terms of power, voice, and diversity of perspective. Focusing the evaluation on other, multiple dimensions of this science reform could diversify and broaden both the educational conversation and the political discourse.<sup>5</sup>

Now onto point number two.

Second ... Powerful evaluation comes not with dispassion but with engagement

Given the understanding that evaluation and its context are mutually constitutive – notably with respect to values, norms, knowledge legitimization, and relationships of power and influence – the question becomes, what kinds of values, norms, knowledge claims, and power relationships do we as evaluators wish to influence? What ends do we wish our evaluative practice to advance? Whose interests do we wish to serve?

These questions point to the intersection of evaluation with democratic theory, for the ideals of democracy provide the most inspiring and inarguably the most defensible normative agenda for evaluation. Ideas about democracy have surfaced in multiple evaluation discourses in recent years, anchored in the historic work on democratic evaluation by Barry MacDonald and colleagues at CARE, including the esteemed Saville Kushner – work that has also been advanced by a number of activist-scholars here in Australia, including Robin McTaggart, Stephen Kemmis, Yoland Wadsworth, and Marie Brennan.

Many participants in these conversations have aspired to a vision of a full participatory democracy, a strong democracy (Barber, 1984), in which all citizens are able to participate meaningfully, fully, and equally in all matters of social policy. Because this “ideal of a participatory democracy is a demanding one, and it can never fully be realized in practice” (at least not right now) (House and Howe, 1999, p. 11), ways to approach or to get close to this ideal have also been advanced in our evaluation literature.

The work of Ernie House (1980, 1993; House and Howe, 1999, 2000) has emphasized the rationally deliberative dimensions of democratic decision making. “In our view it is not enough that participants [in democratic dialogues and deliberations] simply register their opinions and preferences. Their views and preferences should be subject to critique and rational analysis” (House and Howe, 2000, p. 12). Evaluation then provides one forum for such dialogues and deliberations among all legitimate stakeholders. Anders Hanberger (from Umea University in northern Sweden, 2001) offers a related perspective on the value of deliberation in democracy. Hanberger distinguishes what he calls discourse democracy (democracy with the people) from both representative democracy (for the people) and limited participatory democracy (by the people). He also very thoughtfully analyzes the linkages between these visions of democracy and evaluation practice, as summarized in the following table.<sup>6</sup>

<b>Democracy</b>	<b>Evaluation approach</b>	<b>Role of the evaluator</b>
Elitist ( <i>for</i> the people)	Technocratic	Expert
Constrained participatory ( <i>by</i> the people)	Advocacy	Advocate Facilitator
Discourse ( <i>with</i> the people)	Mediating	Counselor Mediator

These views of democracy clearly advocate for inclusion of all legitimate voices and interests, within a framework of rational deliberation. Deliberation is advanced because it can help to forge some kind of agreement amidst multiple and diverse, even conflicting stakeholder interests, argue these theorists. Rational deliberation can thus help avoid the risks of raw, undisciplined pluralism – risks like cross-

<sup>5</sup> I have written fairly extensively on this evaluation experience – a kind of self-debriefing – and would be happy to provide additional references for those interested.

group misunderstandings, intolerance, rejection, which are all fodder for societal disintegration and fragmentation.

My own thoughts about the intersections of democratic theory and evaluation are highly similar to those of House and Howe and of Hanberger, although I am a bit less fearful of relativism because I am more committed to engaging with diversity and difference. Like these theorists, I wish to forge an evaluation practice that serves broad democratic ideals of participation, justice, and equity by meaningfully engaging with people in their lived daily experiences. I agree with Anne McKee and Bob Stake that “democracy in evaluation means honoring the issues, experiences, and values of people, especially the poor and minorities” (2002). In my evaluative work, I am especially concerned with insuring that the interests of all legitimate stakeholders are included, particularly those who are traditionally left out of the conversation. I am a supporter of rational deliberation, but a skeptic about the possibility or even desirability of consensus. Rather, I strive to find ways for diverse stakeholders to talk and dialogue with one another toward greater mutual understanding, respect, tolerance, and acceptance. I am not sure that evaluation, even with full fledged, formalized deliberative forums (see McNeil, 2002), can ever settle important public debates. I am, however, committed to a vision of evaluation as one important site for engaging with the differences that are important in these debates – differences of perspective, experience, values, and political ideology, and differences of privilege, power, prestige, and possibility. I believe there is no more important political challenge in today’s western democracies than finding ways to genuinely understand and respect the differences that currently besiege our everyday lives.

Evaluation cannot meaningfully engage with difference from the cocooned safety of the air bubble. Evaluation cannot be politically meaningful or powerful from a position of political detachment and dispassion. Rather, “influence comes from engagement, not detachment” (Cronbach et al. 1980, p. 53), engagement with democratic ideals and our ongoing struggles to realize them. In other words, evaluation is an intrinsic part of the institutional fabric of public discourse and decision making about public issues. We are among the weavers of this fabric, not just observers or admirers or critics of the weaving. And so let us help ensure that all the many different kinds of weavers are present and join with them in weaving a cloth resplendent with diverse yarns, multi-hued colors, and varied patterns of plaiting.

My remarks will now turn to questions of just how to practice this democratically engaged form of evaluation.

### **Democratically engaging with difference via a mixed-method way of thinking**

One approach to democratically engaged evaluation is to adopt a mixed-method way of thinking. This is, of course, not the only approach, but it remains a potentially powerful and heretofore under-utilized way of democratically engaging with diversity. Mixing methods in evaluation connects to democratic evaluation through their shared valuing of diverse perspectives and multiple ways of knowing. Good mixed-method evaluation actively invites diverse ways of thinking and valuing to work in concert toward better understanding. In good mixed-method evaluation, difference is thus constitutive and generative (from Greene, Benjamin, and Goodyear, 2001).

Let me sketch three inter-related dimensions of evaluation practice and discuss how a mixed-method enactment of each invites democratic engagement with diversity: (1) how we practice our craft or how we use our methodologies; (2) how we think about or theorize the phenomena we are studying; and (3) how we position ourselves in our work. Then, my final point will address the issue of how we position our work in society

---

<sup>6</sup> Hanberger’s work also makes linkages to various visions of civil society.

<sup>7</sup> Difference, says Hannah Arendt, is not a human condition to be overcome. Plurality is not a threat to social life. Instead, the uniqueness of human agency (our capacity to initiate action) is the fundamental condition of human interaction and thus the fundamental aim of politics. The main task of political action is to make plurality possible (Biesta, 2001). See also later discussion on Arendt’s concept of “visiting.”

### Mixing methods in practice

Today, evaluators of all stripes routinely use a variety of methods in their work. We use mail surveys and personal interviews, case studies within quasi-experimental designs, both structured and participant observations, document as well as discourse analysis, and so forth. We use this variety of methods in the broad service of better understanding. And available evidence suggests we make our mixed-method decisions mostly pragmatically, in response to the character of our evaluation charge and the demands of our contexts, rather than for philosophical reasons or on explicitly ideological grounds (see Greene and Caracelli, 2002). That is, mixing methods is not routinely done in the service of democracy.

Even so, even if many mixed-method evaluation studies are crafted for pragmatic rather than political reasons, I contend that mixed-method evaluation practice inherently engages the challenges of human diversity and does so with profound respect for human difference – and thereby invites democratization into the evaluation enterprise. Let me support this contention with a brief (hypothetical) mixed-method example.

#### ***Vignette #2. Community development evaluation***

\* A community development organization is partnering with a local city government to sponsor a series of sports leagues for young teens in the area. The idea is to provide an attractive and healthy alternative for teens' time, and to lessen the allure of gang membership and other anti-social activities. Coaches and officials for these teen sports teams are being specially recruited from groups often victimized by teen mischief and crime – like ethnic shop owners, elders/retirees, and residents of poor neighborhoods.

\* A local foundation is supporting this initiative by funding an evaluation of it. (Obviously, there are many who would *not* view evaluation as support! But, there it is.) The foundation is particularly interested in the meaningfulness of this program experience for its various participants – youth and adults alike. Perfect opportunity for a qualitative case study evaluation!

\* But, the savvy and creative evaluation team, led by the esteemed Saville Kushner – realizing the limits of any single methodology – opts to try to locate their case studies in a broader context of city demography. They seek not only to assess program meaningfulness on a case by case level, but also in terms of neighborhood or community characteristics. They plan to do so by using census data, survey data (if available), or perhaps even spatially via GIS (Geographic Information Systems). So the evaluation will include selected qualitative case stories, as contextualized within more standardized or spatialized descriptions of the character of city neighborhoods. And, we could reasonably assume that the evaluators fashioned this mixed-method design primarily for practical reasons – to enable a better, more comprehensive understanding of the phenomenon of program participation.

\* Even so, this mix of methods offers opportunities to engage more meaningfully with issues of diversity than a single-method study. In addition to the developmental and identity characteristics likely invoked in the individual case studies, different dimensions of diversity may well be engaged with this mixed-method design – neighborhood characteristics such as ethnic/racial homogeneity, transience, sense of community, historical continuities, belongingness, even intensity of social capital. Moreover, *all* dimensions of diversity generated in this mixed-method evaluation are viewed as potentially important and are valued as relevant to the people who named them, for this is the mixed-method way of thinking. And a wider variety of methods is likely to engage a more diverse set of differences. And difference in mixed-method evaluation is respected as generative. And democratic values of equity and justice are potentially served.

### Mixing methods and thereby mixing theories, concepts, ways of thinking and knowing

Much of my own mixed-method work has advanced the value of mixing methods intentionally and explicitly at the level of knowledge paradigm, or how we think about the phenomena we are endeavoring to understand. That is, I believe we can do better mixed-method work if we mix our methods with thoughtful attention to different ways of knowing, not just for practical reasons. My arguments here are (1) epistemological – that we can know something better if we bring multiple ways of knowing to bear on it, and (2) political – that all ways of knowing are partial and thus multiple, diverse ways of knowing are to be valued and respected.

Let me illustrate this admittedly abstract argument with a fanciful example that uses competing contemporary ideas about social reality and what we can know of it.

### ***Vignette #3. The visiting anthropologists from Mars and Venus***

\* Two anthropologists, one from Mars and one from Venus, independently came to earth to evaluate various cultural practices.

\* The one from Mars had prepared for his trip by studying the recent evaluative work of scientific realists – notably, Ray Pawson and Nick Tilly from England (1997) and my American colleagues, Mel Mark, Gary Henry, and George Julnes (2001). These theorists maintain that social reality does indeed exist independent of our thinking about it, but it is much more layered and contextual than previously understood. Causes of human and social phenomena are neither simple nor easily discovered. Rather, such phenomena are assumed to be multiply determined by complex interactions of causal factors that vary across context, space and time and that may well be deeply, almost impenetrably layered. Even so, our job as scientists remains one of figuring out, as best we can, these causal, contextual complexities that *are* social reality – as in Pawson and Tilly’s notions of CMO, context–(causal) mechanism–outcome (effect) configurations.

\* The anthropologist from Venus had prepared for her field work by studying the evaluative theories of social constructionists, including Bob Stake’s responsive evaluation and the ideas of Egon Guba and Yvonna Lincoln (1989). In contrast to realists, these theorists maintain that much of what is important in social reality – namely, meanings and motivations for action – are socially constructed. The demographer markers of race, gender, and class, for example, do not exist out in the world; rather, they are cultural constructions – albeit with real-world power and consequences (because we act *as if* race exists even though it exists only with the particular form in which we construct it – clearly implying that we can deconstruct and then reconstruct it in a different form). For social constructionists, what is important to know, therefore, are the various meanings that people construct about their experiences and the processes by which these meanings are developed.

\* Now, it turns out that both of our alien anthropologists ended up studying at the same time a particularly unusual cultural activity, practiced by some natives of the state of New South Wales on the continent of Australia. The natives called it “cricket” which appeared to both aliens to be some kind of recreational game, even though there is a singing insect on earth with the same name. Here are excerpts from each anthropologist’s fieldnotes made while attending an extended cricket game (it went on for days!!) between the New South Wales Blues and the Victorian Bushrangers on the baggy green fields of Wollongong. And recall that, like all anthropologists in the known universe, these two worked hard to use and understand indigenous languages in their work.

#### Mars’ anthropologist:

##### *Descriptive observational notes:*

The striker stood with his bat held down in front of his wicket. He tapped the bat on the pitch three times.

The non-striker stood behind the other return crease. The bowler took a run-up from behind the non-striker's wicket. He passed to one side of the wicket, and when he reached the non-striker's popping crease he bowled the ball towards the striker, bouncing the ball once on the pitch. The striker hit the ball with his bat. It went 40 meters to the off side. The striker and the non-striker ran to each other’s popping creases two times.

Meanwhile a fielder ran and picked up the hit ball and threw it back to the pitch. The batsmen were run out.

This sequence was repeated as the bowler delivered the ball again to the striker. The striker hit the ball, about 15 meters to the leg side. The batsmen did not run.

##### *Interpretive thoughts:*

This is indeed a complex reality. It remains unclear just what causes what in cricket. Clearly, the ball has to be bowled before it is hit, and it has to be hit before it is fielded, but what causes the batsmen to run? Perhaps it’s the angle at which the ball is hit or the distance or the speed. More logically, it is some complex interaction of these and other factors, like the weather or the hats that are worn by some of the players.

Underlying mediating causal mechanisms may well include acquired skills and confidence in running. I shall have to observe more closely.

#### The anthropologist from Venus:

##### *Excerpts from an interview with a cricket batsman:*

That was an exceptional over, don’t you agree? 18 runs -- my personal best! And that ball on the full over the boundary! Wasn’t that just grand! Especially because it was off a marvelous googly from the bowler. And then there was that glance right through the legs of the square leg! I almost fell down laughing with that one! No fallen wickets for me this time round! Maybe I can do as well again before it’s stumps.

##### *Interpretive thoughts:*

Huh?

More seriously, clearly this interviewee is enthusiastic about something related to cricket. It is not exactly clear what he is enthusiastic about. I believe it has something to do with his constructed meanings of *boundaries* and *glances*. These seemed especially meaningful to him. But, his laughter – notably about the legs of the square leg – may also signal other dimensions of significant idiosyncratic constructions and understandings. I shall have to listen more carefully.

Beyond the silliness of this language game, the point being made is that these two anthropologists can readily complement and enhance each other's work, and thereby the resultant depth of understanding reached. If Drs. Venus and Mars could accept and appreciate each other's way of conceptualizing and making sense of human activities, they would jointly do much better than each alone. Mars may advance his causal understanding of running in cricket by asking the runners themselves – who are likely to embed important contextualities in their causal explanations. And Venus may more successfully interpret her interviewees' enthusiasms with a more standardized and representative portrayal of the game. And, in the process of sharing and dialoguing, one with the other, democratizing respect for diverse ways of knowing could be enacted, along with the concomitant values of equity and reciprocity – not to mention significant inter-planetary harmony.

### Being a mixed-method evaluator

Mixed-method evaluations can take many forms, and many different kinds of evaluators can ply a mixed-method trade. But, there are likely to be certain practices, certain ways of being an evaluator that underlie good mixed-method practice. Figuring out what just these practices are is an important task for future thought and study. To inaugurate this line of work, I have one fine candidate to bring forward this morning – and that is the practice of listening, of listening well.

I was struck once again by the connections between the idea of listening well and conducting good mixed-method evaluative work when I read a journal article recently featuring some of the ideas of Hannah Arendt (Coulter and Wiens, 2002). Extracting heavily from this article ... Arendt was an important 20th century Jewish philosopher, who fled Germany in 1933 and went first to France and then to the US. Arendt's post-war work can be viewed as an effort to understand how and why the Holocaust could have happened. In particular, Arendt pursued links between thinking, judging, and acting – trying to understand how really smart people, really good thinkers (such as her colleague and lover Heidegger), could be such bad judges and do such awful things.

In this work, Arendt attended specifically to inter-connections between good judgment and wise action, toward developing a portrait of the “good judging actor.” I think evaluators could and should aspire to be “good judging actors.” Arendt first articulated her views of what constitutes good judgment.

One judges [said Arendt] as a member of a community... Good judgment ... is not a matter of objective knowledge or of subjective opinion, but a result of intersubjectivity; becoming a good judge depends largely on one's capacity to consider other viewpoints of the same experience, “to look upon the same world from another's standpoint, to see the same in very different and frequently opposing aspects” (Arendt, 1968, p. 51). (Coulter and Wiens, 2002, p. 17)

For Arendt, a respect for plurality is thus a necessary attribute of a good judging actor. Arendt rejects both pluralism-without-judgment, or unfettered diversity, and judgment-without-pluralism, or judgments made outside the web of human diversity (Biesta, 2001). Equally important for Arendt is a respect for the uniqueness and agency of diverse others. So, to develop an appreciation for and understanding of the viewpoints of others, to develop good judgment,

Arendt advocates “visiting,” which involves carefully listening to the perspectives of others because “the more people's standpoints I have present in my mind while I am pondering a given issue, ... the better I can [judge]” (Arendt, 1968, p. 241). [So] actors must go into a plural public world and engage with others [by visiting and listening carefully].... Arendt's public [world] is not an abstract public sphere, but a world of diverse and unique individuals, all capable of agency.... Respecting diverse standpoints requires dialogue with other people, listening to their stories, and relating to their uniqueness without collapsing these divergent views into a generalized amalgam (Coulter and Wiens, 2002, p. 18) ... without essentializing them, and without losing your own unique standpoint. “Visiting is therefore not to see through the eyes of someone else, but to see with your own eyes from a position that is not your own ... in a story very different from [your] own” (Biesta, 2001, p.398).

What a wonderful vision for the practice of mixed-method inquiry – going out into the public world, visiting with diverse others, listening well to each of them – and thereby being enabled to make good judgments in the service of doing wise actions.



And finally ...

### **Evaluation as a public craft**

Making good judgments in the service of doing wise actions, in the service of contributing to wise practice – that is, to social, educational, environmental, and development programs that are strong and successful; that enable participants to have healthier, more productive, more meaningful lives; that help forge networks of compassion and trust; that help build caring and vital communities – doesn't this sound like a worthy vision for evaluation? Obviously, I think it does and I believe that the points I have made so far are consonant with this vision.

In lofty terms, I am talking about positioning evaluation as a public craft and evaluators as stewards of the public good. A public craft, from Harry Boyte of the Center for Democracy and Citizenship (2000), is “work that is undertaken for public purposes and in public ways ... [it is work that] adds public judgment or wisdom to knowledge ... [work that invokes the technical canons of science and is] attentive to the local setting ... and to the civic implications of practices” (p. 1). And Harry Boyte makes the same argument I have regarding the importance of conducting professions, like evaluation, in democratizing ways and regarding the value of a mixed-method way of thinking for this purpose. Says Harry,

[Traditional scientific] models of knowledge ... emphasize the detached, rational, analytic observer as the highest judge of truth and the most effective problem solver. This approach is in conflict with communal common sense, folk traditions and appreciation for craft knowledge mediated through everyday life experience.... This is not necessarily to say that science is “wrong,” and folkways are “right” (for instance, in medicine, it is good to have a highly trained specialist to fix a broken limb). But it is to propose that there are different sorts of valuable knowledge in public life. A singular celebration of the scientifically educated expert as the actor or initiator in public affairs marginalizes the amateur, while it produces mainly information and knowledge – neither wisdom nor public judgment. We need a very different and far more civic craft model of professional practice if we are to see any widespread democratic renewal in our time. (p. 3)

So, the point is not to get rid of science, but rather to forge a partnership between science and citizenship, to think of evaluators as public citizens, as good judging actors, and evaluation as the advancement of wise action.

In more concrete and practical terms, let me share one last vignette, this one based on recent experience, an experience integrally wrapped up in the tangles of the New Public Management.

#### ***Vignette #4. Evaluation in and for the community***

\* Seated around the comfortable table were representatives from four local funders and myself. The funders included city and county governments, as well as a local chapter of the United Way (an umbrella non-profit organization that raises local funds and then redistributes them to community social service agencies). The purpose of the meeting was to plan a workshop on outcomes and performance measurement for staff in local agencies that receive grants from these funders. The funders were hoping that the workshop would enable local agencies to begin move beyond their customary focus on *outputs*, or what activities they conducted and how many people came, to the identification of key *outcomes*, or what difference their services have made in the lives of participants. When I asked the funders, “why are you so interested in outcomes and performance measures,” they replied, “because *our* funders are demanding this kind of information. We need to now report on outcomes to *our* funders.” Sigh. I believe this scenario has been repeated thousands of times in meeting rooms all over the western world during this past decade. Many of you have probably attended such meetings.

\* My challenge at this meeting and the subsequent workshop I conducted was not to belittle or ignore these accountability pressures. While I am far from enamored with most performance measurement systems, I agree with many that good information on the outcomes of public services is long overdue. AND, in embracing diversity and difference, one cannot be selective. No, instead my challenge was to *reconnect* evaluation – in this context, in its incarnation as performance measurement – with wise practice, to reposition evaluation as not only a management tool but also as a meaningful engagement with diversity and difference, a time for visiting with others and listening well, a democratizing advancement of the public good, a public craft conducted by good judging actors.

\* The workshop that I co-led aspired to the following vision. In this vision of a performance measurement workshop, participants would discuss together the following simple but profound questions:

1. *What is the purpose of each of our various community, social, and educational programs? Why are we doing them? What are we hoping to accomplish?*

Here, we would create a vision of the kind of community we wished to live in, and we would acknowledge that program work is extremely challenging and complex.

2. *How would each of us know (or how do we know) if we are reaching these ambitions, if we are creating the kind of community we aspire to, if each of our programs is a good one?*

Here, we would talk about the importance of reflective critique on practice and the valuable contributions that evaluative data can make to such reflections and critique. We would also talk about how difficult this is to do well.

3. *Let us each pick one important program outcome and talk about ways that we could come closer to understanding how well it is being fulfilled by our program and what program changes might enable better fulfillment and thus a stronger, healthier community.*

Here, we would talk about holding group interviews or local forums in which community people could share *their* ideas about the meanings of positive youth development, substance abuse rehabilitation, social capital networks, environmental sustainability, and so forth – and we would talk about how such diverse views could be incorporated into revisions to wise practice. We would also talk about existing measures, social indicators, or even brief surveys that could chart progress on important program and community outcomes. And we would talk about how these various kinds of information could be reported to funders, yes, but also and importantly, used to get closer to our collective community vision of democratic health and vitality.

Was I able to conduct this workshop according to this vision? With the assistance of the esteemed Saville Kushner, oh yes indeed. No, just kidding. The actual workshop was much less dialogic and much more focused on outcomes and their measurement than imagined in this vision. But, the vision remained as an inspiration, and the actual workshop *was* successful in assertively connecting outcomes and performance measurements to the substantial challenges and vital importance of wise practice, of engaging with diversity and thinking multiplistically, of embracing difference, of being a good judging actor.

And that I believe is what each of us as evaluators can do to connect our public craft with stewardship of the public good. May we all become good listeners.

## References

- Barber, B. (1984). *Strong democracy: Participatory politics for a new age*. Berkeley, CA: University of California Press.
- Biesta, G.J.J. (2001). How difficult should education be? *Educational Theory*, 51 (4), 385-4000.
- Boyte, H.C. (2000). "Professions as public crafts." A background paper prepared for the working conference on New Information Commons, Racine, WI.
- Coulter, D., and Wiens, J.R. (2002). Educational judgment: Linking the actor and the spectator. *Educational Researcher*, 31 (4), 15-25.
- Dahler-Larsen, P., and Krogstrup, H. (2000). "Constitutive effects of evaluations." Paper presented at the annual conference of the American Evaluation Association, Honolulu.
- Fine, M. et al. (2000). For whom? Qualitative research, representations, and social responsibilities. In N.K. Denzin and Y.S. Lincoln (eds.), *Handbook of qualitative research, second edition*. Thousand Oaks, CA: Sage.
- Greene, J.C. (1997). Evaluation as advocacy. *Evaluation Practice*, 18, 25-35.
- Greene, J.C., Benjamin, L., and Goodyear, L. (2001). The merits of mixing methods in evaluation. *Evaluation*, 7 (1), 25-44.
- Greene, J.C., and Caracelli, V.J. (2002). Making paradigmatic sense of mixed method practice. In A. Tashakkori and C. Teddlie (eds.), *Handbook of mixed methods in social and behavior research*. Thousand Oaks, CA: Sage.
- Greene, J.C., and Walker, K.C. (2001). Review of – *Evaluation: An integrated framework for understanding, guiding, and improving policies and programs*, by M.M. Mark, G.T. Henry, and G. Julnes. *Evaluation and Program Planning*, 24, 368-371.
- Guba, E.G., and Lincoln, Y.S. (1989). *Fourth generation evaluation*. Thousand Oaks, CA: Sage.
- Hanberger, A. (2001). Policy and program evaluation, civil society, and democracy. *American Journal of Evaluation*, 22, 211-228.
- Henry, G.T. (2001). How modern democracies are shaping evaluation and the emerging challenges for evaluation. *American Journal of Evaluation*, 22, 419-429.
- Hood, S. (2001). Nobody knows my name: In praise of African American evaluators who were responsive. In J.C. Greene and T.A. Abma (eds.), *Responsive evaluation. New Directions for Evaluation no. 92*. San Francisco: Jossey-Bass.
- House, E.R. (1980). *Evaluating with validity*. Thousand Oaks, CA: Sage.
- House, E.R. (1993). *Professional evaluation: Social impact and political consequences*. Thousand Oaks, CA: Sage.
- House, E.R., and Howe, K.R. (1999). *Values in evaluation and social research*. Thousand Oaks, CA: Sage.
- House, E.R., and Howe, K.R. (2000). Deliberative, democratic evaluation. In K.E. Ryan and L. DeStefano (eds.), *Evaluation as a democratic process: Promoting inclusion, dialogue, and deliberation. New Directions for Evaluation no. 85*. San Francisco: Jossey-Bass.
- Mark, M.M. (2001). Evaluation's future: Furor, futile, or fertile? *American Journal of Evaluation*, 22, 457-479.
- Mark, M.M., Henry, G.T., and Julnes, G. (2000). *Evaluation: An Integrated Framework for Understanding, Guiding, and Improving Policies and Programs*. San Francisco: Jossey-Bass.
- McKee, A., and Stake, R.E. (2002). Making evaluation democratic in a climate of control. In L. Bresler and A. Ardichvili (eds.), *Research in international education: Experience, theory, and practice*. New York: Peter Lang.
- McNeil, C. (2002). Evaluator as steward of citizen deliberation. *American Journal of Evaluation*, 23, 45-54.
- Pawson, R., and Tilly, N. (1997). *Realistic evaluation*. London: Sage.

- Schwandt, T.A. (2002). *Evaluation practice reconsidered*. New York: Peter Lang.
- Valovirta, V. (2002). Evaluation utilization as argumentation. *Evaluation*, 8, 60-80.
- Wong, L.M. (1998). The ethic of rapport: Institutional safeguards, resistance, and betrayal. *Qualitative Inquiry*, 4, 178-199.
- Worthen, B. (2001). Whither evaluation? That all depends. *American Journal of Evaluation*, 22, 409-418.