Evaluation and usability as a practice area has diversified its approaches, broadened the spectrum of UX issues it addresses, and extended its contribution into deeper levels of product-development decision making. This forum addresses conceptual, methodological, and professional issues that arise in the field's continuing effort to contribute robust information about users to product planning and design.

David Siegel and Susan Dray, Editors

A Professional Empiricist Manifesto

David A. Siegel
Dray & Associates | david.siegel@dray.com

Susan M. Dray
Dray & Associates | susan.dray@dray.com

This inaugural installment of the forum on evaluation and usability is naturally a time for us, as its editors, to let readers know what to expect here and to inspire contributions by defining the forum's scope. We see this as a timely opportunity to reexamine the core value that evaluation contributes as a specialized activity within HCI. The need for this reexamination is driven by the evolution of evaluation as a practice area within HCI, as well as by trends in the evolution of HCI itself, some of which have challenged the centrality or even relevance of evaluation in product planning, development, design, and management. We call this a manifesto because it is an assertion of core principles and a call to a group to take up these principles.

Before discussing why such a manifesto is even necessary, we should clarify what we include under "evaluation and usability." At some time in the past, it may have been appropriate to define usability as the discipline that evaluates things other people have created to see if they are easy to use by testing them in a lab or by reviewing them according to

accepted (and hopefully validated) usability design guidelines or heuristics. This definition became inadequate a long time ago. While many people in usability still specialize in laboratory evaluation, the HCI sub-discipline of usability as a whole includes people using a broad spectrum of methodologies to address issues that are relevant at all stages of product development, from product strategy to interaction design. Ease of use is no longer the only goal of usability, and there is widespread recognition within the community that usability must take into account other issues in user experience, such as utility and engagement.

This broadened scope extends usability in at least two dimensions: beyond the lab and into the field, and beyond the later stages of design into the earlier stages, in which opportunities are assessed and product concepts begin to jell. Blurring a conventional distinction between generative research and evaluation may be controversial, but this distinction is actually pretty fuzzy to begin with. Even lab usability testing is not inherently a pass/fail exercise (although it can be), but potentially a gen-

erative process. Many usability professionals have had the experience that formative evaluation not only yields ideas for elegant new design solutions but also suggests fundamental reframing of the design problems or product concept. Even the distinction between early and late evaluation research breaks down when you consider what people should study before the design team has gone to work. Evaluation research that is late in the development of v.3 is early in the development of v.4. But what about v.1 of an innovative product? If there is a human need, it is guaranteed that people are already doing something to address it, using existing methods and tools, however crude. This is what makes the need potentially discoverable in research. The same spectrum of methods relevant to evaluating what the team will eventually create is also relevant to evaluating these precursor experiences in order to identify needs, opportunities, constraints, and requirements.

Challenges of Growth

Some of the trends that create a need for refocusing on the central

If practitioners are not more sophisticated in their critique of methodologies (including their own) than their audiences, then they are setting themselves up to be marginalized.

contributions that usability and evaluation make are by-products of the growth of usability as a field. One common side effect of the growth of a field can be "deskilling." As with "green" products, usability has become mainstream, at least in that companies now feel obligated to claim their products are usable. Founders of a field tend to differ from the people who enter a more mature field. We are so far beyond the days when usability was the rallying cry of a relatively small group of human-centered visionaries and change agents that usability has sometimes been described as having become a commodity. One manifestation of this is that many—or even possibly most—job descriptions that one finds on the Internet calling for skills in evaluation or research treat it as a small appendage to design and UI development skills, rather than as a specialty in its own right.

It may be that attempts to lower the barriers to adoption of user-centered design practices have inadvertently communicated that getting user input is easy. In fact, the many pitfalls of making sure that data about human behavior is interpretable and that we can draw appropriate, design-relevant implications from it warrant the dedication of a large, specialized discipline. Anyone who does not recognize the existence of these many pitfalls and who is not prepared to judiciously apply a range of strategies for mitigating them is in danger of being misled. There are simply too many opportunities for errors. Certainly, talking informally with a small sample of users may be useful for suggesting design ideas, provided those ideas will later be subjected to more rigorous data-based critique before

too much is invested in them.

Early adherents of a new field have a missionary zeal about promoting the importance of the issue they are trying to address and showing that their approaches are revolutionary. However, as the new practices become established and pervasive, they then become the baseline of practice to be critiqued, and this inevitably leads to more focus on their limitations. To its credit, usability has not been hesitant to critique itself and to evolve new methods. Of course, these will have their own strengths and weaknesses, but hopefully they are at least complementary to those of existing methods. This activity can be seen in academic research and in debates among practitioners. However, we must be clear that this intense self-critical activity does not mean there is no such thing as a valid usability finding. Rather, it is a mark of professionalism and expertise that drives progressive improvement and innovation. If practitioners are not more sophisticated in their critique of methodologies (including their own) than their audiences, then they are setting themselves up to be marginalized.

The expanded purview of the field of evaluation has also brought about challenges stemming from diversification. The challenge is that this broadening and diversification create increased overlap with other disciplines, while at the same time raising questions about what the unique contribution of evaluation is and what the core factors are that unify it. The risk is that the definition will default back to the most distinctive and visible thing that usability "owns" in organizations, which is testing.

External Challenges

Other challenges that inspire our manifesto come from the larger context of interactive product development. One notion is that the evolution of digital technology has made usability less relevant because it has been superseded by other considerations in the user experience. The proliferation of digital interactions aimed at supporting ongoing experiences that do not have discrete success/failure endpoints has challenged evaluation to expand the range of data it looks at and the types of phenomena it tries to understand. Usability, narrowly defined as focused on ease of use, is only one factor that contributes to desirability, attractiveness, engagingness, and all the other virtues at which interactive product design aims, although it contributes to these things in varying degrees, depending on the type of product.

Social networking provides a special example of this issue. No one would deny that ease of use is important for accomplishing tasks such as creating an account, managing privacy, understanding the relationship between what you do and what others will see, and using the range of various communicational/sharing features on a social networking site. However, when the most powerful driver of engagement and stickiness of the experience is thought to be the self-organizing behavior of the networks of users and the content they contribute through using the site, the usability of these tools can be seen as a peripheral issue that has little impact on the overall success of the site.

Evolution of technology has led some to suggest that the traditional concerns of usability have become less troublesome. There

is probably some truth to the idea that as people have learned to use a wider range of basic interaction modalities, they are becoming increasingly adaptable and efficient when learning new ones, and that some design patterns have stabilized and been internalized by both users and designers. On the other hand, each new product presents a particular combination of constraints and trade-offs, and there will always be new uncertainties as patterns abstracted from one set of contexts are applied to another specific context. While transferable knowledge may provide a head start, it is not necessarily sufficient. It may simply push forward the leading edge of design questions on which evaluation should be brought to bear.

Empiricism as a Unifying Principle

To define the core contribution of evaluation and usability in a way that clarifies its deep relevance in the face of these and similar challenges, we must first consider the broadened scope of the field, beyond a narrow focus on ease of use and laboratory evaluation. But then, what is the range of possible activities relevant to this broader definition? For example, what unifies unstructured contextual research, whether from a contextual inquiry or ethnographic tradition, with laboratory testing? Furthermore, if usability is defined as including, for example, research to discover basic user needs that could be addressed by new products, what distinguishes it from other practices used to uncover user needs?

Our proposed answer assumes a distinction between what it means to be a professional as opposed to a technician. A profession is not

reducible to proficiency in using a set of techniques or methods. Instead, it implies fluency in applying a certain body of knowledge and certain types of thinking to problems in a particular domain. For example, you would not define the core skills contributed by design as a profession as "skill in using Photoshop and similar tools." Fundamental skills support the collection, interpretation, and integration of complex and sometimes inconsistent information using judgment. Techniques are certainly important, but professionalism is reflected in the ability to intelligently recognize their limitations and to innovate to adapt them to the particularities of each given real-life context.

With this in mind, we propose this definition:

Usability and evaluation as an area of professional practice represents the effort to introduce disciplined empiricism about human behavior into the product-development process.

Some hallmarks of this professional focus include:

• the ability to identify and articulate beliefs, assumptions, and predictions about human behavior that are often implicit in product planning and design discussions, and to use them to formulate hypotheses or to frame questions at a useful level of abstraction, for which data (evidence) is at least potentially relevant;

• appreciation of the interconnectedness and multilayered nature of factors influencing human behavior (cognitive, emotional, social and interpersonal, and cultural, among others), which enables broad consideration of the range of data topics that are potentially relevant;

• appreciation of the strengths

We urge practicing evaluation and usability professionals to live up to the calling of professional empiricist, and to use it as a basis for collaborating usefully with others who contribute to product development by being actively involved in integrating information across disciplines.

and limitations of different data types—for example, behavioral, attitudinal, and structured versus unstructured and qualitative versus quantitative;

• appreciation of the factors that influence the quality of data, including potential biases and differing sensitivities of specific data-gathering and measurement techniques, sampling approaches, and so on;

• critical evaluation of the overall quality of evidence relevant to a proposition;

• skill in assessing the incremental cost and benefits of achieving different levels of rigor according to the needs of a project, taking into account the risks and benefits of varying degrees of certainty;

• sensitivity to the common pitfalls of interpreting data about human beings, such as our tendency to default to habitual, simplistic, often self-serving ways of accounting for human behavior ("They won't adopt it because they are resistant to change"). This also includes managing the risks of post hoc reasoning, use of anecdotal evidence, stereotyping, insensitivity to confounds, attribution of intentionality to impersonal or statistical processes, stereotyping, and so on; and

• a habit of thinking critically about the challenges of generalizing findings, which helps avoid the problems caused by extrapolating from a sample of people to a population of people, from evidence obtained in a limited sample of usage contexts to other contexts, and from one technical environment to others.

Notice that this definition does not equate empiricism with reductionism, quantification, or isolation of variables, although

these can be consistent with it. It addresses something more basic and therefore more inclusive: the devotion to applying critical thinking to the gathering and interpretation of evidence. We mean to include research that strives to be holistic and qualitative, as well as quantitative. An ethnographer who is sensitive to the inherent strengths and weaknesses of the ethnographic approach, who critically evaluates the strength of her own data and is careful to not make claims beyond what it can support, who considers alternative ways of accounting for findings and plans ways of gathering evidence to help select among them, and who considers approaches other than hers when they might provide more relevant and stronger data is probably operating as a professional empiricist contributing to the evaluation enterprise (even if she is understandably reluctant to use the term "evaluation" because of other narrow associations it has with usability testing). This definition does not automatically place people who have a specialized focus on a particular narrow technique in the "technician" category. In fact, one cannot be a professional without having mastery of technique.

Finally, evaluation is neutral about the role of deductive, inductive, and abductive thinking. The professional empiricist simply asks what evidence is relevant to assessing a proposition, regardless of which of these forms of thinking led to it, but he or she should also be capable of using all of those forms of thinking to generate hypotheses.

Design, Evaluation, and Professional Empiricism

With the growth of a professional

design community focused on interactive products, a sometimes polarized relationship has evolved between design and evaluation. Post hoc analyses of successful products in which the hand of a creative designer seems outwardly very evident can create the impression that design skill alone ensures usability, that design ensures engagement so that usability is irrelevant, or that design renders explicit evaluation based on user data irrelevant. Designers may talk about the usefulness of contextual research and ethnographic data as a source of inspiration, but sometimes reserve for themselves the authority to decide what to do in response to this inspiration. Some question the value of evaluation's role as an input to critiquing design, or even accuse evaluation of being potentially harmful to the design endeavor by squelching innovation. Hopefully this is not simply because many designers have been sent back to the drawing board on the basis of usability findings. After all, design is a profession that prides itself on its culture of criticism. We therefore need to consider that there may be a legitimate basis for their complaints.

Part of the criticism of usability arises from its real limitations when it is narrowly defined as laboratory evaluation of ease of use, especially when practitioners do not recognize these limitations and draw questionable conclusions as a result. For example, laboratory usability evaluation is not optimized to study the process of learning and adaptation that users may undergo in the process of adopting initially unfamiliar technologies and interaction models. If this is not taken into account,

usability findings from the lab may too easily be used to shoot down innovative ideas. Similarly, even when ease of use is a highly relevant dimension of user experience to evaluate, it is very difficult in lab testing to study and take into account the interaction between people's motivation to use something and their experience of ease of use over time, and the interaction among ease of use and other user experience dimensions. A professional empiricist should be prepared to address these issues.

There can be a tone of mutual negative stereotyping in the polemics between design and evaluation. However, we need to acknowledge that the accumulated experience that each side has with members of the other discipline may all too often confirm these mutual stereotypes. The field of evaluation contains many shining examples of professional empiricism, both in academic research settings and in applied product development settings, but we cannot assume that everyone whose job title includes the word "usability" or "evaluation" exemplifies the best practices. On the other hand, using the existence of methodological limitations in practice to discount the *inherent* contribution of evaluation underestimates the potential of professional empiricists to factor in their awareness of methodological limitations as they interpret their data and communicate their findings.

Every phase of the process of product development generates propositions and uncertainties about humans worthy of disciplined evaluation based on evidence. We urge practicing evaluation and usability professionals to live up to the calling of

professional empiricist, and to use it as a basis for collaborating usefully with others who contribute to product development by being actively involved in integrating information across disciplines. This means they need to understand the mind-set or complementary thinking skills of their fellow disciplines and the dilemmas with which they wrestle. Sometimes usability professionals take the stance that their job is to provide findings, leaving it up to others to figure out what they mean for the product. Failure to take responsibility for how your data is interpreted, weighed, and integrated with other information is the stance of a technician, not a professional. Indeed, if too many usability and evaluation people actually behave like technicians rather than professionals, they will provide justification for being stereotyped and marginalized. We hope that this forum will play at least a small role in helping to prevent that, and we invite readers to join us in this effort with by contributing thought-provoking articles on any of the many important professional issues facing evaluation and usability.

ABOUT THE AUTHORS
David Siegel, vice president of Dray & Associates, is a user-centered design consultant who has contributed his research skills to a wide range of technologies. He uses methods ranging from contextual field research to laboratory evaluation.

Susan Dray, president of Dray & Associates, is a practitioner and consultant carrying out both generative and evaluative field research, and has taught many practitioners how to design, conduct, and interpret field research, among other things.