RE-EXAMINING THE IMPLICATIONS OF SYSTEMS THINKING FOR EVALUATION

BY

EMILY F. GATES

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Educational Psychology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2016

Urbana, Illinois

Doctoral Committee:

    Professor Jennifer C. Greene, Chair
    Professor Thomas A. Schwandt, Director of Research
    Professor Lizanne DeStefano
    Professor Martin Reynolds, Open University

ABSTRACT

Over the last twenty years, many prominent evaluators have been borrowing and using ideas, theories, and methods from the systems and complexity fields with little research on the implications of this trend for the evaluation field. This thesis examines this borrowing to identify over-arching implications for evaluation theory and practice. The first paper reviews inter-disciplinary literature on systems thinking and complexity science with regards to evaluating social policies and programs and identifies major implications for how evaluators theorize evaluation practice. The second paper reports on an analysis of eight cases of evaluation practice that use systems and complexity ideas and techniques and presents findings regarding how evaluators conceive of and practice evaluation. The third paper advances an argument for how evaluators can use critical systems heuristics to surface, reflect on, and make explicit the values that influence and should influence an evaluation. Collectively, these papers support the potential of borrowing from the systems and complexity fields to expand on and re-define evaluation theory and practice.

ACKNOWLEDGMENTS

This dissertation would not have been possible without the mentorship of Dr. Thomas Schwandt and Dr. Jennifer Greene. I am thankful to Tom for overseeing this dissertation, sharing his wisdom, providing generous editorial comments, and continuously challenging and expanding my thinking. I am thankful to Jennifer for her early and ongoing encouragement of this research, for being a supportive advisor throughout the years, and for continuously inspiring me to be a more thoughtful and compassionate scholar and evaluator. I would also like to thank Dr. Lizanne DeStefano for fostering my development as an evaluator, Dr. Martin Reynolds for paving a way in critical systems and evaluation, and Bob Williams for sharing his brilliant enthusiasm for evaluation and systems thinking – and to all five for investing significant time and energy in this project as committee members. I want to express my gratitude to the eight evaluators I interviewed for their time, insights, and leadership in the field. Additionally, I want to acknowledge the University of Illinois Graduate College for providing funding to complete this dissertation.

I could not have accomplished this project without the support of my colleagues, friends, and family. My thanks to my QUERIES, I-STEM, WIDER, and ISSS Ph.D. colleagues for providing motivation, feedback, and a sense of community, to Carolina (for friendship from day one), Ayesha (for nudging me to come to UIUC and stick with it), Gabi (for sharing the ups and downs), Sarai (for reminding me to smile), Megan-Brette (for reminding me of the reasons we do this), Lorna (for providing a space to unwind), Priya (for your refreshing visits), Molly (for your intellectual companionship), Lisa (for giving me candles to light the way), and to Mbikayi (for listening to my ramblings and providing constant encouragement). Last but not least, I want to thank my parents, Gerry and Cindy, for their unconditional emotional and financial support and Tyler, Austin, and especially Carlye for being there in more ways than I can name.

TABLE OF CONTENTS

INTRODUCTION

Throughout the development of the evaluation field, evaluators have borrowed knowledge from other fields and disciplines. Consider, for example, the intersection of anthropology and evaluation to better understand the contexts of evaluands (Fetterman, 1984; Butler & Copeland-Carson, 2005); contributions of organizational theories of learning to strengthen evaluation use (Russ-eft & Preskill, 2009); political philosophy for envisioning the roles of evaluation in democracy (House & Howe, 1999); and econometric methods for assessing resource use and value (Yates, 2012). In the last fifteen years, many prominent evaluators have been borrowing ways of thinking (i.e. ideas, concepts, theories) and techniques (i.e. methods, tools, approaches) from the systems and complexity fields--an interdisciplinary area of inquiry including operational research, systems science, complexity science, and cybernetics that broadly focuses on the study of systems and the design of interventions to address problematic situations (Eoyang, 1999; Sanderson, 2000; Barnes et al, 2003; Williams and Imam, 2007; Callaghan, 2008; Rogers, 2008; Morell, 2010; Forss, Marra, & Schwartz, 2011; Patton, 2011; Westhorp, 2012; Byrne, 2013).

One significant motivation for this borrowing arises from criticisms of longstanding tendencies to define the objects of evaluation by pre-determined objectives and rather linear and mechanistic processes and to study effects primarily as intended changes in individuals. Many evaluators as well as evaluation commissioners are arguing, to the contrary, that initiatives that address persistent, interconnected issues (often called wicked problems) operate within continually changing circumstances influenced by many factors at multiple scales and amidst differing perspectives, values, and interests. For the evaluation field to be relevant and useful to both stakeholders and to the wider public affected by such initiatives, a turn to the systems and complexity fields is needed (Stern, Saunders, & Stame, 2015).

However, at present, the significance and implications of this turn are not yet clear (Mowles, 2014; Stern et. al, 2015). The literature includes examinations of various theories and ideas for evaluation (Sanderson, 2000; Barnes et. al, 2003; Davies, 2004; Rogers, 2008; Callaghan, 2008; Hummelbrunner, 2011; Byrne, 2013; Marra, 2015); new ways to conduct evaluations using systems and complexity concepts (Morell, 2010; Patton, 2011; Forss, Marra, and Stern, 2011; Levin-Rozalis, 2014); cases of evaluations that used systems and complexity thinking and techniques (Williams & Imam, 2007; Patton, McKegg, and Weipeihana, 2015; Fredericks, Deegan, & Carman, 2008; Dyehouse, Bennett, Harbor, & Childress, 2009; Morell, Hilscher, Magura, & Ford, 2010; Durland & Fredericks, 2005; Reynolds & Williams, 2011); and literature reviews (Mowles, 2014; Walton, 2014). Additionally, there has been considerable research in neighboring fields, including public health, social services, community psychology, and international aid and development, that addresses the implications of systems and complexity thinking and techniques for designing, implementing, and evaluating policies and programs (e.g., Milstein, 2008; Ramalingham, 2013; Foster-Fishman, Yang, & Nowell, 2007). Much of this research has focused on 'hard' (i.e., quantitative, realist) traditions and largely overlooked 'soft' (i.e. qualitative, dialogic) and 'critical' (i.e., mixed methods, pragmatic) traditions within the systems and complexity fields. There is a need for broad reviews of this research to identify both current implications for evaluation theory and practice and future directions and to explore potential contributions of lesser-examined traditions.

Overview of the Three Papers

The three papers in this thesis conceptually and empirically examine the implications of borrowing from the systems and complexity fields for evaluation theory and practice. The first paper reviews recent literature on systems thinking and complexity science in evaluation and related fields and identifies major implications for evaluating social interventions. Findings suggest re-thinking foundational dimensions of evaluation theory. The second paper reports on an analysis of eight cases of evaluation practice that use systems and complexity ideas and techniques and presents findings regarding how evaluators conceive of and practice evaluation. The study provides insights into the variety of ways and circumstances in which evaluators are applying systems and complexity ideas and techniques. The third paper informs the conversation about valuing by explaining what critical systems heuristics (CSH) can contribute to how evaluators think about and practice valuing. The central argument advanced is that CSH can help evaluators identify, question, and justify the values and assumptions that influence and should influence an evaluation.

Implications of Systems Thinking and Complexity Science (STCS) for Evaluation

This paper reviews recent literature on STCS in evaluation and related fields (i.e., public health, international aid and development, community psychology, and social services) and identifies major implications of STCS for evaluating social interventions. The argument advanced is that STCS poses six challenges to the ways evaluators conventionally think about and practice evaluation in relation to the following: 1) supporting social problem solving, 2) framing an intervention and its context, 3) selecting and using methods, 4) engaging in valuing, 5) producing and justifying knowledge, and 6) facilitating use. As a critical literature review, this article addresses a trending topic and a gap in the research and provides insights about the

practices of evaluation and program planning. This article is most relevant to evaluators and program planners. It was submitted to *Evaluation and Program Planning*, and I am in the process of revising for resubmission on the basis of generally positive reviewer feedback.

Borrowing Knowledge from the Systems and Complexity Fields: Implications for Evaluation Practice

This study examines eight cases of evaluation practice employing systems and complexity thinking and techniques as defined by particular evaluators. Case materials included semi-structured interviews with each evaluator and documents describing evaluations these evaluators conducted as written in journal articles, book chapters, and evaluation plans and reports Evaluation practice is defined as how evaluators think about what being an evaluator means and what they are doing when they are practicing evaluation and, more specifically, the activities evaluators carry out when designing and conducting evaluations in particular circumstances. Three questions guided this study: 1) What ideas and techniques are these evaluators borrowing in their evaluation practices? 2) How do these ideas and techniques influence the way these evaluators design and conduct evaluations? 3) In these evaluators' views, what key challenges and opportunities does borrowing pose for evaluation practice? Findings highlight the variety of ways and circumstances in which evaluators are taking up and using systems and complexity ideas and techniques. Rather than a wholesale change in evaluation practice, this study found that these evaluators are re-thinking specific roles, relationships, and methods and re-configuring these in some new ways.

This article is targeted for *Evaluation: The International Journal of Theory, Research, and Practice.* Addressing a topic of interest in the international evaluation community, this paper examines the practices of evaluators from the United States and Europe in the context of a broader examination of the phenomenon of interdisciplinary borrowing to improve evaluation

practice. This article fits the aims of the journal as well as recent articles published on systems and complexity ideas and responds to a call made by the journal's editor (Stern et. al, 2015).

Valuing in Evaluation with Critical Systems Heuristics

This paper explains what critical systems heuristics (CSH) can contribute to how evaluators think about and practice valuing – the process and product of judging the merit, worth, or significance of an evalaund. The central argument advanced is that CSH can help evaluators be more critically reflective about valuing. Being critically reflective means identifying, questioning, and justifying the values and assumptions that influence and should influence an evaluation and, particularly, a judgment of the merit, worth, or significance of an evaluand. The paper introduces CSH and three dimensions that comprise the foundation of its contributions to valuing in evaluation and proposes four ways in which CSH can help evaluators be more critically reflective about valuing: 1) in considering the social value of evaluation; 2) in framing the evaluand and the evaluation; 3) in selecting and justifying criteria with consideration for exclusion and marginalization; and 4) in developing a witness role. The paper suggests that addressing valuing more systematically and explicitly, as proposed, can enhance the credibility, defensibility, and legitimacy of an evaluation.

As a critical essay, this article, to be submitted to the *American Journal of Evaluation,* examines a central yet unsettled issue in evaluation and identifies contributions from a tradition of systems thinking that has received little attention in the United States evaluation field. This article supports a current focus of the journal, on evaluation and systems thinking, and the journal's aim of exploring challenges related to conceptualizing, designing and conducting evaluation.

Significance

Evaluation practice is at a critical juncture. Current ways of thinking about social problems, interventions designed to address those problems, and methods of assessing the value and impact of those interventions are ill suited to the kinds of adaptive management required of governments, foundations, and development organizations to address complex social and environmental issues in constantly changing environments. Strong interest in examining the implications of systems and complexity thinking and techniques for evaluating social interventions is evident, yet responsible research on this matter remains to be done. This dissertation adds value to the field by drawing on analysis of inter-disciplinary literature, empirical cases, and overlooked systems theories to advance a much-needed, alternative way of framing evaluation practice and new approaches to conducting evaluations that are better suited to managing complex social change processes.

REFERENCES

Barnes, M., Matka, E., & Sullivan, H. (2003). Evidence, understanding and complexity: evaluation in non-linear systems. *Evaluation, 9*(3), 265-284.

Better Evaluation. (2014). Retrieved from: http://betterevaluation.org

Butler & Copeland-Carson, (2005). Creating Evaluation Anthropology: Introducing an Emerging Subfield. *Annals of Anthropological Practice*. 24(1), 1-181.

Byrne, D. (2013). Evaluating complex social interventions in a complex world. *Evaluation, 19*(3), 217-228.

Callaghan, G. (2008). Evaluation and negotiated order: developing the application of complexity theory. *Evaluation, 14*(4), 399-411.

Davies, R. (2004). Scale, complexity and the representation of theories of change: Part II. *Evaluation, 11*(2), 133-149

Durland, M. M., & Fredericks, K. (2005). An introduction to social network analysis. *New Directions for Evaluation*(107), 5-13.

Dyehouse, M., Bennett, D., Harbor, J., Childress, A., & Dark, M. (2009). A comparison of linear and systems thinking approaches for program evaluation illustrated using the Indiana Interdisciplinary GK-12. *Evaluation and Program Planning, 32*(3), 187-196.

Eoyang, G. H., & Berkas, T. H. (1999). Evaluating Performance in a Complex Adaptive System. In M. Lissack & H. Gunz (Eds.), (pp. 1-21). Westport, Connecticut: Quorum Books.

Fetterman, D. (1988). Qualitative approaches to evaluating education. *Education and Educational Research.* 17(8), 17-23.

Forss, K., Marra, M., & Schwartz, R. (Eds.) (2011). *Evaluating the Complex: Attribution, Contribution, and Beyond.* Vol. 18. London, United Kingdom: Transaction Publishers.

Foster-Fishman, P. G., Nowell, B., & Yang, H. (2007). Putting the system back into systems change: a framework for understanding and changing organizational and community systems. *American Journal of Community Psychology, 39*(3-4), 197-215.

Fredericks, K., Deegan, M., & Carman, J. (2008). Using system dynamics as an evaluation tool: experience from a demonstration program. *American Journal of Evaluation, 29*(3), 251-267.

House, E., & Howe, K. (1999). Values in Evaluation and Social Research. Thousand Oaks, CA: Sage Publications, Inc.

Hummelbrunner, R. (2010). Beyond logframe: critique, variations, and alternatives. In N. Fujita (Ed.), *Beyond logframe; using systems concepts in evaluation* (pp. 1-33). Japan: Foundation for Advanced Studies on International Development.

Levin-Rozalis, M. (2014). *Let's Talk Program Evaluation in Theory and Practice* Tel Aviv, Israel Dekel Academic Press.

Marra, M. (2015). Cooperating for a more egalitarian society: Complexity theory to evaluate gender equity. *Evaluation, 21*(1), 32-46.

Milstein, B. (2008). Hygeia's Constellation: Navigating Health Futures in a Dynamic and Democratic World (pp. 1-124).

Morell, J. A. (2010). *Evaluation in the Face of Uncertainty: Anticipating Surprise and Responding to the Inevitable* New York, New York: The Guilford Press.

Morell, J. A., Hilscher, R., Magura, S., & Ford, J. (2010). Integrating Evaluation and Agent-based Modeling: Rationale and an Example for Adopting Evidence-based Practices. *Journal of Multidisciplinary Evaluation, 6*(14), 32-57.

Mowles, C. (2014). Complex, but not quite complex enough: The turn to the complexity sciences in evaluation scholarship. *Evaluation, 20*(2), 160-175.

Patton, M. Q. (2011). *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*. New York, New York: The Guilford Press.

Patton, M.Q., McKegg, K., and Weipeihana, N. (Eds.) (2016). *Developmental Evaluation Exemplars: Principles in Practice*. New York, NY: The Guilford Press.

Ramalingam, B. (2013). *Aid on the Edge of Chaos: Re-thinking International Cooperation in a Complex World*. United Kingdom: Oxford University Press.

Reynolds, M., & Williams, B. (2011). Systems Thinking and Equity-focused Evaluation (pp. 115-141): UNICEF.

Russ-eft, D., & Preskill, H. (2009). *Evaluation in Organizations: A Systematic Approach to Enhancing Learning, Performance, and Change* (Second ed.). New York, New York: Basic Books.

Rogers, P. J. (2008). Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions. *Evaluation, 14*(1), 29-48.

Sanderson, I. (2000). Evaluation in complex policy systems. *Evaluation, 6*(4), 433-454.

Stern, E., Saunders, M., & Stame, N. (2015). Standing back and looking forward: Editors' reflections on the 20th Anniversary of Evaluation. *Evaluation, 21*(4), 380-390.

Walton, M. (2014). Applying complexity theory: a review to inform evaluation design. *Evaluation and Program Planning, 45*, 119-126.

Westhorp, G. (2012). Using complexity-consistent theory for evaluating complex systems. *Evaluation, 18*(4), 405-420.

Williams, B. & Imam, I. (Eds.) (2007). *Systems Concepts in Evaluation: An Expert Anthology*. Point Reyes, CA: EdgePress of Inverness.

Yates, B. T. (2012). Step Arounds for Common Pitfalls When Valuing Resources Used Versus Resources Produced *New Directions for Evaluation, 133*, 43-52.

PAPER 1

IMPLICATIONS OF SYSTEMS THINKING AND COMPLEXITY SCIENCE
FOR EVALUATION

Throughout the development of the evaluation field, new trends gain traction with bold promises of transforming how we evaluate social interventions (i.e., policies, programs, practices). For example, over the last three decades, stakeholder approaches have influenced norms about stakeholder involvement in evaluations (Rodriguez-Campos, 2012) and, more recently, results-based management has increased demand for results-based monitoring and evaluation systems (e.g., Kusek & Rist, 2004). Systems thinking and complexity science (STCS) are the latest new trends in evaluation (Reynolds, Forss, Hummelbrunner, Marra, & Perrin, 2012). Growing interest in these ideas is evident across books (Eoyang & Berkas, 1999; Williams & Imam, 2007; Williams & Hummelbrunner, 2011; Morell, 2010; Patton, 2011; Forss, Marra, & Stern, 2011; Wolf-Branigin, 2013; Levin-Rozalis, 2014), journals (Cabrera, Colosi, & Lobdell, 2008; Mowles, 2014), conference themes of professional associations (Parsons, Keene, & Dhillon, 2014), and reports from agencies commissioning evaluations (e.g., Fujita, 2010; GIZ, 2011). Driving this interest are myriad ways in which evaluators and evaluation commissioners regard the potential of systems thinking and complexity science to transform how social interventions are evaluated.

There are a variety of efforts to translate insights from STCS for evaluation theory and practice. Efforts focus on methods and methodologies as well as conceptual and theoretical issues. At least since the 1980's, scholars have been bringing systems and complexity methods into the evaluation field (Ulrich, 1988; Gregory and Jackson, 1992ab; Midgley, 1996) with the recent book *Systems Concepts in Action: A Practitioner's Toolkit* (Williams & Hummelbrunner, 2011) as, perhaps, the latest attempt. Some of the methods being explored include causal loop

diagrams and system dynamics (Fredericks, Deegan, & Carman, 2008; Dyehouse, Bennett, Harbor, & Childress, 2009); agent-based modeling (Morell, Hilscher, Magura, & Ford, 2010); soft systems methodology (Attenborough, 2007); social network analysis (Durland & Fredericks, 2005); and critical systems heuristics (Reynolds & Williams, 2011). Evaluators have developed new evaluation conceptual frameworks and guides for practice (e.g., Cabrera & Trochim, 2006; Parsons, 2007; Hargreaves, 2010; Gopalkrishnan, Preskill, & Lu, 2013; Preskill & Gopalkrishanan, 2014; Marra, 2011a; Cabrera, Colosi, & Lobdell, 2008; Wasserman, 2010) based on STCS for evaluating complex (i.e., emergent processes and outcomes) and systems change (i.e., intended to modify social systems such as communities, schools, healthcare) interventions. New theoretical approaches to evaluation practice have been developed, for example *Developmental Evaluation* (Patton, 2011), *Systemic Evaluation* (Boyd, Geerling, Gregory, Kagan, Midgley, Murray, & Walsh, 2007), *Systematization* (Tapella & Rodriguez-Bilella, 2014), and several conventional approaches have been modified to incorporate STCS including *Responsive Evaluation* (Gregory, 1997) and *Theory-based Evaluation* (Davies, 2004; Stame, 2004; Rogers, 2008; Callaghan, 2008; Hummelbrunner, 2010).

Beyond these implications discussed within the evaluation community, scholars in related fields are examining the transformative implications of STCS for designing, implementing, and evaluating social interventions. In public health, international aid and development, community psychology, and social services, scholars argue that STCS challenge and transform the ways these fields are conceptualized and practiced (in public health see Leischow & Milstein (2006), Sterman (2006), Trochim et. al, (2006), Leishow et. al, (2008), and Milstein (2008); in international aid and development see Ramalingham et. al (2008), Jones (2011), Ramalingham

11

(2013); in community psychology see Foster-Fishman (2007) and Foster-Fishman & Watson (2012); in social services see Wolf-Branigin (2012)).

Many evaluators still feel uncertain and, in some cases, utterly confused about these ideas and what, if anything, they mean for practicing evaluation differently. There have been few efforts to broadly examine this systems and complexity trend in the evaluation field and identify major implications for evaluators. Walton (2014) has identified implications of complexity theory for evaluation design, and Mowles (2014) critically reviewed the turn to complexity science in evaluation. The only broad examinations of systems thinking in evaluation have focused on conceptualizing systems thinking (Cabrera, Colosi, & Lobdell, 2008) and exploring the use of systems concepts (e.g., interrelationships, perspectives, boundaries) and methodological approaches in specific evaluation cases (Williams & Imam, 2007). Evaluators curious about what systems thinking and complexity science are and what they mean for evaluation practice will find a flourishing and lively discussion, but little clarity or agreement on the most significant implications of these ideas for evaluation practice.

This paper reviews recent literature on systems thinking and complexity science in evaluation and identifies over-arching implications for conceptualizing and practicing evaluation. The breadth and plurality of the systems and complexity fields, the evaluation field, and the recent conversation at their intersection make it impossible to comprehensively cover this literature. Instead, the review is purposefully organized around the central ideas of systems thinking and complexity science and a basic conception of evaluation practice. The terms *systems thinking* and *complexity science* are defined in multiple ways. Systems thinking can refer to: 1) a way of thinking characterized by core systems concepts (such as interrelationships, perspectives, and boundaries or distinctions, systems, relationships, perspectives); 2) a way of

thinking about real-world phenomena as systems (such as a school system or health care system; and 3) an inter-disciplinary field of theories, approaches, and methods. The term *complexity science* can mean: 1) an alternative paradigm of science; and 2) a set of ideas, theories, approaches, and methods for studying complex systems. The abbreviation STCS, is used to refer to the combination of these two ideas. For the purposes of this paper, evaluation practice is conceptualized as a set of six interrelated activities: 1) supporting social problem solving; 2) framing interventions and contexts; 3) selecting and using methods; 4) engaging in valuing; 5) producing and justifying knowledge; and 6) facilitating use. These activities were selected based on two well-known frameworks for analyzing theories of evaluation practice (Shadish, Cook, & Leviton, 1991; Alkin & Christie, 2003; Christie & Alkin, 2013). The review includes academic and grey literature that discusses the implications of STCS for evaluating social interventions. The review does not include systems and complexity theoretical research or case applications of evaluations using STCS.

The argument advanced here is that STCS pose six challenges to the ways evaluators conventionally think about and practice evaluation: 1) shift from assuming a linear, predict-act-evaluate approach to social problem solving to building capacity for a more iterative, adaptive approach; 2) consider multiple ways to frame a social intervention and its context; 3) recognize the partiality and inevitable influence of methods on social interventions; 4) judge the value of social interventions by comparing descriptive and normative models of these interventions; 5) expand evaluative knowledge to be more problem-oriented, collaborative, and inclusive of multiple kinds of evidence and causal relationships; and 6) embed evaluation use into an intervention and/or agency by designing evaluations as feedback loops, social learning processes, or normative reflection and direction setting. Taking up these implications would, in theory,

transform evaluation practice. However, STCS also suggest shifts in the role of evaluation in governing and managing ongoing change, the institutional contexts for conducting evaluations, and evaluators' training, capacities, and responsibilities in order to fully realize these transformations in evaluation practice.

Reviewing the Literature

Literature addressing implications of systems thinking and complexity science for evaluating social interventions was drawn from a variety of academic fields and included peer-reviewed journal articles as well as grey literature (e.g., conference proceedings, evaluation guides, policy briefs). Journal articles were drawn primarily from three areas: evaluation, systems and complexity science and practice, and intervention-driven fields. The latter area included research in public health, international aid and development, organizational management, community psychology, education, and social services.

Search engines (i.e., EBSCOHost, Google Scholar) were used to identify journal articles for inclusion in the review. Searches were restricted to 1988 to 2015 and used combinations of the following terms: systems thinking; systemic thinking; complexity thinking; complexity science; systems perspective; systems approach; system; evaluation; systems evaluation; systemic evaluation. Additionally, select journals in evaluation and systems and complexity science and practice were searched. Evaluation journals, specifically *Evaluation, American Journal of Evaluation, Evaluation and Program Planning,* and *Journal of Multi-disciplinary Evaluation* were searched using the terms complexity, complex, and systems thinking. Systems and complexity science and practice journals, specifically *Systemic Practice and Action Research, Journal of the Operational Research Society,* and *Emergence: Complexity & Organization* were searched using the terms evaluation, evaluate, and evaluating. A separate

search was conducted to identify grey literature. The terms systems thinking, complexity science, and evaluation were used on Google and IssueLab to identify potential guides and statement papers published by foundations, research think tanks, and development agencies. Some additional articles were identified indirectly through a snowball technique of searching the references of articles found to identify additional articles relevant to the review and through word of mouth including articles sent by colleagues and through listservs.

Articles included in the review were limited to those that addressed systems or complexity thinking, science, theories, and/or methodologies in relation to evaluating social interventions. Articles included were mostly conceptual or theoretical. Articles were excluded for five main reasons. First, articles discussing systems or complexity thinking in relation to non-evaluation activities such as knowledge management, complex decision-making, capacity development, operational research, management science, and project management were excluded. Second, empirical case studies of systems or complexity approaches or methodologies used in evaluations were not included if they lacked discussion of evaluating social interventions more generally. Third, in evaluation journals, articles were excluded if they focused on a specific evaluation methodology (e.g., QCA) or approach (e.g., realist evaluation, developmental evaluation) in relation to an issue, intervention, or set of circumstances, which was described as complex or a system. Following this criterion, articles discussing evaluation systems, systems of care, and complex interventions were excluded if this was the only discussion of systems and complexity. Fourth, in systems journals, articles were excluded if they were only about evaluating a systems theory (e.g., Maturana's constructivist family therapy) or systems methodology (e.g. total systems intervention, interactive planning) and did not discuss the broader issues involved in evaluating social interventions. Additionally, articles applying

evaluation frameworks based on specific systems methodologies (e.g., soft systems

methodologies) to specific circumstances and/or interventions were not included if they did not

make theoretical connections to evaluating social interventions. Table 1 provides the count of

articles included organized by field.

Table 1.1. Number of Articles and Books Included in Review by Field

Field	Articles	Books
Evaluation	61	9
Systems and/or Complexity	13	
International Aid and Development	11	
Public Health	11	
Community Psychology	5	
Social Services	5	
Education	4	
Other	2	
TOTAL	113	9

The process for reviewing and analyzing materials involved sorting excerpts of articles

into large, a priori categories then analyzing each of these categories to develop smaller, a

posteriori sub-categories. A priori categories were based on the analytic framework of six

activities: 1) supporting social problem solving; 2) framing an intervention and its context; 3)

selecting and using methods; 4) engaging in valuing; 5) producing and justifying knowledge; and

6) facilitating use. Each article was read and coded with these categories using NVIVO software.

Each a priori category was further analyzed to identify a posteriori categories of implications of

STCS for each activity.

Six Implications of Systems Thinking and Complexity Science for Evaluation

This section identifies implications of STCS for six activities involved in evaluation practice. Each sub-section begins with a brief overview of how evaluators typically think about and carry out that activity followed by discussion of major ways STCS challenges or expands on this. Specific examples from the reviewed literature are provided.

1. Supporting Social Problem Solving

Evaluations are conducted to contribute to social betterment (Mark, Henry, & Julnes, 2000). Underlying this aspiration is a way of thinking about the nature of social problems, how social problem solving works, and the role evaluation plays in this process. Social problems are typically conceptualized as undesirable conditions or unmet human needs that require change of some kind. Social betterment, as defined by Mark, Henry, & Julnes (2000), refers to the "reduction of social problems and the increased meeting of human needs" (p. 24). Needs assessments typically involve defining the problem, assessing its extent, defining and identifying the targets of interventions (e.g., individuals, families, communities), and describing service needs of these target populations (Rossi, Lipsey, & Freeman, 2004). Social problems are assumed to be separable such that a problem (and its solution) can be studied and addressed independently of other problems; a problem is relatively stable over time and across levels of analysis; and a problem is treatable such that, with the "right" solution, the problem will be reduced or eliminated altogether. Social problem solving is conceived as a relatively straightforward, linear, and rational process in which (1) problems are identified and defined; (2) policies and programs to address these problems are created and implemented; and (3) these policies and programs are evaluated in terms of their efficiency, effectiveness, and impacts on addressing these problems. Moreover, social practices (e.g., teaching, nursing, providing

community services) are also often considered problematic and in need of science-based solutions (Schwandt, 2005), and social betterment is a kind of progress through the application of social scientific and evaluation knowledge to these practices (Schwandt, 1992). The role of evaluation in social problem solving is primarily to provide knowledge about the means of addressing social problems (i.e., policies and programs) with little emphasis on questioning the ends (e.g., constructions of the problem, alternative aims) (Mark, Henry, & Julnes, 2000).

STCS advances quite a different way of thinking about social problem solving and the role of evaluation in supporting this process. Social problems are conceived of as situations that are found to be problematic, undesirable, and requiring change by particular people in a particular place and time; problems are continuously changing and subject to differing perspectives such that any attempt to bound them is temporary. Rittel and Webber (1979) distinguish between tame and wicked problems arguing that all social policy problems are wicked in nature and can be distinguished by ten features[1]. The most urgent, crosscutting, and pervasive social, political, and environmental issues can be described as wicked including poverty, hunger, homelessness, racism, educational inequity, and environmental degradation. Defining these problems is part of what makes them problematic. For each problem, there are pluralistic and contested definitions and any definition implies a solution option and, therefore, a

---

[1] 1) There is no definitive formulation of a wicked problem; 2) wicked problems have no stopping rule; 3) solutions to wicked problems are not true-or-false, but good-or-bad; 4) there is no immediate and no ultimate test of a solution to a wicked problem; 5) every wicked solution to a problem is a "one-shot operation"; because there is no opportunity to learn by trail-and-error, every attempt counts significantly; 6) wicked problems do not have an enumerable (or an exhaustively describable) set of potential solutions, nor is there a well-described set of permissible operations that may be incorporated into the plan; 7) every wicked problem is essentially unique; 8) every wicked problem can be considered a symptom of another problem; 9) the existence of a discrepancy representing a wicked problem can be explained in numerous ways. The choice of explanation determines the nature of the problem's resolution; 10) the planner has no right to be wrong (Rittel & Webber, 1979, p. 161-166).

political and normative stance about what should be done. Further, most social problems cannot be separated, and any policy or program to address one problem inevitably influences others. Therefore, rather than conceiving of social practices as comprising problems in need of science-based solutions, practice requires ongoing evidence and normative reflection. The notion of social betterment thus entails ongoing moral-political inquiry and debate as well as scientific inquiry (Schwandt, 1992, 2005). Social problem solving becomes an ongoing, iterative process of learning about social problems and adaptive management. The role of evaluation in social problem solving broadens from a focus on informing decisions about existing social policies and programs to facilitating learning about and informing continuous adaptive management. Table 3 contrasts these two approaches to social problem solving.

Table 1.2. Contrasting Two Approaches to Social Problem Solving

Dimensions	Traditional Approach	STCS Approach
Studying problems	• Problems as undesirable conditions or unmet needs; problems as gap between current situation and goals • Assess the nature and extent of the problem and target populations	• Problems as gap between current situation and normative ideals about how things should be • Conduct descriptive analyses of the situation, including influential factors and stakeholders' views, and normative analyses of what should be
Envisioning change	• Identify the necessary changes in individuals' attitudes, knowledge and behaviors	• Identify the structures and processes that influence the situation
Designing interventions	• Determine best means to the desired end • Design activities to change individuals' attitudes, knowledge, and behaviors • Ensure that activities can be directly (linearly and rationally) linked to goals and results	• Determine desired ends and multiple means • Design interventions that could affect the structures and processes that influence the situation (e.g., leverage points, social networks, information flows)
Implementing interventions	• Implement intervention with fidelity to pre-determined plans • Interventions are often implemented by single agencies working independently and managed by top-down and centralized expertise and decision-making	• Implement activities with flexibility, ongoing innovation, and adaptation to the changing situation • Interventions often require multiple agencies working together with de-centralized management, expertise, and decision-making
Evaluating interventions	• For learning and improvement of existing interventions • For accountability – to assess whether desired results occurred and, if not, why not	• For learning – to capture what happened, why, and what actions should be taken next • For innovation and adaptation – to inform what should be done next

STCS suggests that questioning assumptions about social problem solving is necessary because it is these assumptions that often lead to intervention failure, policy resistance, and negative unintended consequences. Sterman (2006) makes this point as follows: "Policy resistance arises because we do not understand the full range of feedbacks surrounding—and created by—our decisions... Yesterday's solutions become today's problems" (p. 507). There are numerous examples of policy resistance across all arenas of social intervention including pesticides and herbicides that poison other species up the food chain; flood control efforts that lead to more severe floods by preventing natural dissipation of water; and road building programs to decrease congestion that increase traffic, delays, and pollution (Sterman, 2006, p. 506). In each of these situations, it is not adequate for evaluators to demonstrate the ineffectiveness and negative consequences of the interventions without challenging the underlying logic and disregard for dynamic complexity (e.g., feedback, time delays, stocks and flows) that prevented policymakers from anticipating the consequences of these interventions.

As more intervention-driven fields and agencies (e.g., governments, foundations, international aid and development organizations) shift to adaptive management approaches to social problem solving, evaluators will need to adjust their practices to support this approach. In a review of the value of systems thinking for evaluation, Trochim and colleagues (2006) conclude that a shift to ongoing planning, action, and evaluation is the greatest implication of systems thinking. Scholars have theorized variations on adaptive management approaches to social problem solving. For example, Sanderson (2009) argues for intelligent policymaking, as opposed to evidence-based policy making, "in which we accommodate the complexity surrounding the application of intelligence in policy making, treat our policies as hypotheses to be tested in practice, to be piloted where feasible and appropriate and to be subject to rigorous

evaluation, and in which we learn from these processes and apply the intelligence thus gained to future policy thinking and decisions. We must attend to building capacity to support such a model of policy learning…" (p. 700). Jones (2011) advises agencies implementing interventions to facilitate decentralized action and self-organization through means such as building adaptive capacity, supporting networked governance, and building leadership and facilitation (p. viii-viiii). Eoyang & Holladay (2013) suggest organizations should engage in adaptive action: cycles of observing dynamics and patterns creating uncertainty, better understanding the current situation and options, and taking effective action. In addition to these adaptive management approaches for agencies, particular social interventions are being designed with similar principles such as building networks (Benjamin & Greene, 2009), generating innovations (Gopalkrishnan, Preskill, & Lu, 2013), and directing ongoing actions (Barnes, Matka, & Sullivan, 2003). Evaluating these interventions requires new ways of thinking about and practicing evaluation evident, for example, in the call of Arkesteijn, van Mierlo, and Leeuwis (2015) for a reflexive evaluation and monitoring approach.

2. Framing an Intervention and its Context

At the outset of an evaluation, evaluators develop an understanding of the intervention and the context in which it is situated to inform how they design and conduct the evaluation. This activity has largely been conceptualized as a descriptive activity – giving an account of the relevant characteristics of the intervention and the context in which it is situated ("BetterEvaluation," 2015). Interventions are broadly considered to be activities carried out by people at a particular place and time to achieve some kind of change (Weiss, 1998). Contexts are the settings within which these activities occur including descriptive and demographic characteristics, material and economic features, institutional and organizational climate,

22

interpersonal dimensions, and political dynamics (Greene, 2005). Further, context includes the "combination of factors (including culture) accompanying the implementation and evaluation of a project that might influence its results" (Thomas, 2004, p. 11 in Fitzpatrick, 2012). To construct a description of an intervention and its context, evaluators typically read program documents and past evaluations, talk with stakeholders, directly observe intervention activities, and, potentially, articulate the logic or theory of how the intervention works (Weiss, 1998).

STCS challenges evaluators to consider different ways of framing social interventions and contexts. The literature included multiple, and not necessarily mutually exclusive, conceptualizations of interventions and context. Interventions can be framed as: 1) characterized by their complex aspects, 2) a system (or subsystem), 3) a complex adaptive system, 4) an agent to change a system, 5) a situation of human activity, 6) a space for innovation and adaptation, and 7) a complex causal process. Context can be framed as 1) real-world system(s), 2) the problematic situation addressed by the intervention, and 3) a constructed reference system(s). Each framing draws on different STCS concepts and methods. Tables 4 and 5 provide overviews of each conceptualization, the STCS-informed evaluation concepts this conceptualization draws on, ways to frame the intervention or context using STCS methods, and examples of articles addressing each conceptualization.

Table 1.3. Seven Ways to Conceptualize Interventions Using STCS

Interventions as…	STCS-Informed Evaluation Concepts	Ways to Frame an Intervention Using STCS Methods	Examples
Characterized by complex aspects	• *Simple:* agreement about ends and certainty about means • *Technically complicated:* agreement about ends but uncertainty about means • *Socially complicated:* disagreement about ends but certainty about means • *Complex:* disagreement about ends and uncertainty about means	• Cynefin and other matrices to distinguish simple, complicated, complex, and chaotic aspects	Rogers (2008, 2011) Patton (2011)
Systems nested within systems	• *System*: collection of interdependent parts that, through their interactions, function as a whole • *Purpose*: what the system does • *Boundary*: distinction between system and its environment	• Systems mapping techniques (e.g., causal loop diagrams, influence diagrams) to visualize system	Hargreaves et. al (2010)
Complex adaptive systems	• *Dynamic*: constantly and discontinuously changing • *Massively entangled*: many enmeshed relationships	• Complexity techniques for understanding behavior of CAS over time (e.g., causal diagrams, iterative redesign, shorts and simples, feedback analysis, time series analysis)	Eoyang & Berkas (1999)

Table 1.3 (continued)

Interventions as…	STCS-Informed Evaluation Concepts	Ways to Frame an Intervention Using STCS Methods	Examples
	• *Scale independent*: functions at multiple levels simultaneously		
	• *Transformative*: exhibits transformed and transforming behaviors		
	• *Emergence*: properties that arise from the interconnections of the system		
Agents to change systems	• *Leverage points:* places to intervene in the system • *Target system:* the system an intervention aims to change	• Analysis of the target system behavior, stakeholders perspectives, and leverage points (e.g., soft systems methodology)	Shiell, Hawe & Gold (2008) Wasserman (2010) Foster-Fishman (2007)
Situation of human activity	• Situations of interest or concern: real-world circumstances • *Stakeholding:* conflicts arising from stakeholders (i.e. people involved in or affected by the situation) with differing stakes (i.e. roles and interests) • *Systems:* conceptual ideas used by stakeholders for improving the situation	• Situational mapping and analysis techniques (e.g., rich pictures) • Discourse processes for examining values, assumptions, and perspectives (e.g., critical systems heuristics)	Reynolds (2010)

Table 1.3 (continued)

Interventions as…	STCS-Informed Evaluation Concepts	Ways to Frame an Intervention Using STCS Methods	Examples
Space for innovation and adaptation	• *Innovation:* developing new ways to address a problem situation • *Adaptation*: continuously adjust to the changing environment • *Principles*: guidelines for practice that must be interpreted, applied, and adapted situationally • *Social network*: social structure of connected agents	• Recording what's happening in real-time and decision options (e.g., strategic options development and analysis) • Envisioning future options (e.g., scenario analysis)	Patton (2011) Ling (2012)
Complex causal process	• *Recursive causality:* multi-directional causal relationships governed by feedback loops • *Path dependency:* multiple possible future states for any system • *Negotiated order:* systems as structured and shaped by local action	• Techniques to model the causal processes (e.g., system dynamics, agent-based modeling)	Byrne (2013) Callaghan (2008) Spruill et. al (2001)

Table 1.4. Three Ways to Conceptualize Context Using STCS

Context as…	STCS-Informed Evaluation Concepts	Ways to Frame Context Using STCS Methods	Examples
Real-world system(s) that influence the intervention	*Nested:* vertical connections between systems; system of interest exists within larger systems and contains smaller systems such that a change in any one system may affect the system of interest *Networked:* horizontal connections between systems; system of interest is connected to parallel systems such that a change in one of these systems may affect the system of interest	Systems mapping and computer simulation of system behavior over time	Ling (2008) Hawe et. al (2009) Wasserman (2010)
Problem situation the intervention is embedded in	*Problem situation:* the loosely bound and defined description of a situation that is found in need of change by someone or some group *Human activity system:* activity to serve a purpose within an environment	Surface different stakeholder perspectives on the problem situation (e.g., soft systems methodology)	
Constructed reference system(s)	*Reference system*: interlocking assumptions that determine what facts and values are relevant *Boundary categories*: twelve categories of boundary assumptions that comprise a reference system	Surface and critique different references systems; contrast descriptive ('what is') and normative ('what should be') reference systems (e.g., critical systems heuristics)	

In addition to offering new conceptualizations and ways to frame interventions and contexts, STCS calls attention to the normative dimension of this activity in which evaluators set boundaries within which an intervention and its context can be studied. For example, in any evaluation, evaluators make decisions about issues including time scale, levels of analysis, disciplinary/theoretical lens, relevant problems/needs, intervention purposes, stakeholder perspectives, key activities, and intended outcomes and impact. These issues are called boundary judgments because they determine what is to be included and what is to be excluded in an evaluation (Ulrich, 1988). Boundaries are not given empirically, but rather require making value judgments about what should be considered and these judgments influence the evidence considered relevant in an evaluation. STCS challenges evaluators to set boundaries critically. That involves making implicit boundary decisions explicit, examining alternatives, considering practical and ethical consequences of boundary choices, justifying those chosen, and being open to re-setting boundaries (Ulrich & Reynolds, 2010). The first and foremost boundary judgment has to do with which ways of framing interventions and contexts are relevant in a particular evaluation. For example, should an evaluation of an elementary reading curriculum be examined as a nested system (e.g., within a classroom, school, and community) or as a space for innovation and adaptation (e.g., teachers tailoring the curriculum to groups and individual students differently)? Each framing will highlight subsequent boundary judgments that need to be made such as who or what to include in the community system and what is considered an adaptation of the curriculum versus something other than the curriculum. Therefore, setting boundaries critically requires ongoing surfacing and justifying values advanced in an evaluation.

3. Selecting and Using Methods

For some evaluators, the choice and use of methods are the most important activities in evaluation practice. While there are longstanding debates about methodology, there are three relatively widely held assumptions about the selection and use of methods. The first is that methods are for collecting and analyzing empirical data to describe or explain relatively stable, knowable phenomena such as intervention activities, outcomes, and impacts. The second assumption is that the choice of methods should be based on the characteristics of the intervention and the evaluation circumstances (e.g., purpose, questions, resources, time frame) (referred to as situational responsiveness) (Patton, 2005; Patton, 2012). The third assumption is that methods are neutral tools and any influence of these tools on the object of inquiry (i.e. intervention, stakeholders) is more or less determinable and controllable (Schwandt, 2015). STCS challenges each of these assumptions.

The first assumption, that evaluators use methods to investigate relatively stable and knowable phenomena, is overturned in STCS. Most STCS methods are designed to study unstable, constantly changing, and only partially knowable phenomena. These methods include approaches to modeling the behavior of complex systems by means of causal loop diagrams (Dyehouse, Bennett, Harbor, Childress, & Dark, 2009), system dynamics (Fredericks et. al, 2008; Grove, 2015); and social network analysis (Durland & Fredericks, 2005; Hawe, Bond, & Butler, 2009) and exploring multiple perspectives on a situation or intervention through soft systems methodology (Attenborough, 2007) and critical systems heuristics (Ulrich, 2005; Ulrich & Reynolds, 2010; Reynolds, 2007; Reynolds, 2014).

The second assumption, that method selection should be responsive to the situation of application (i.e., intervention characteristics, evaluation purpose and questions), is challenged by

the multiplicity of ways any situation can be framed. Contrary to widespread thinking in evaluation, STCS contends that an evaluator cannot stand independently from a situation (or intervention) in order to impartially describe it and then select methods based on this description (Mowles, 2014). Many STCS methods are designed to examine situations from multiple perspectives and from each perspective the portrayal of the situation (e.g., key influences, extent of agreement or disagreement, level of uncertainty) may be different. Further, different methods for gathering perspectives elicit different kinds of portrayals, ranging from emphases on causal mechanisms to underlying worldviews. Method selection and the situation of application are interdependent such that selection of a method informs how the situation of application is understood and any understanding of the situation will inform which methods are considered relevant and applicable. The main implication of this interdependence is that evaluators must be thoughtful and cautious about the practical and ethical consequences of methods selected for their influence on the way a situation is understood (Midgley, 1996; Boyd et. al, 2007). This issue has been long-debated in STCS in relation to contingency approaches (e.g., Cynefin) that first classify situations along dimensions (e.g., uncertainty, agreement) and, based on these classifications, suggest methods for investigation or strategies for intervention. Despite these debates, most evaluators using STCS (e.g., Rogers, 2008; Patton, 2011; Parsons, 2007; Hummelbrunner, 2011) continue to advance the assumption that method selection should follow the evaluator's assessment of a situation.

The third assumption, that methods can be used neutrally, is challenged by the assumed influence of STCS methods on the phenomena they are used to investigate. Many STCS methods are designed and used purposefully to have an influence on the phenomena they investigate. For example, computer simulated modeling of complex systems can be used to change the way

people involved in the system perceive and behave in the system (Sterman, 2006) and, similarly, stakeholder-based approaches to deliberating about problem situations are often used to challenge assumptions about and ways of behaving in these situations (Checkland, 1999). While this influence is assumed to be uncontrollable and undeterminable, possibilities for the influence of methods can be anticipated and studying these influences can be part of the process of using the method. Unlike conventional thinking about methods, in STCS considering the ways in which methods actually constitute the objects they seek to investigate are central to method use.

4. Engaging in Valuing

Evaluators make value judgments about social interventions including judgments about process (i.e., formative), merit/worth (i.e. summative), and about what should be done next (i.e., developmental). The main issue with respect to valuing in evaluation has to do with how and what values are selected as the basis of criteria upon which a judgment is rendered (Henry, 2002; Schwandt, 2015). The current literature advises conducting needs assessments and considering a range of criteria based on stakeholder and public values (Scriven, 2007; Davidson, 2013; Mark, Henry, & Julnes, 2000). STCS emphasizes the instability and conditionality of value judgments. Value judgments about interventions are unstable and time bound. For example, in tobacco control, support for banning smoking in public places increases after the policy is implemented (Shiell, Hawe, & Gold, 2008). Value judgments of interventions vary across levels of a complex system such as citywide homeless shelters that provide housing locally but lead to more homelessness in the city over time and fewer shelters statewide as the homeless population relocates to the city with shelters. Value judgments vary depending on the purpose and perspective one takes as in "a soccer game perceived as a game of skill or a means of entertainment. A soccer game played by incompetents could be judged hugely entertaining,

31

whereas a game played skillfully could be judged very dull" (Williams & Rogers, 2010). In short, value judgments differ across time period, level of analysis, purpose, and perspective taken among other factors. STCS suggests another way of making value judgments that can, at least somewhat, account for the instability and conditionality of value judgments.

STCS suggest making value judgments by constructing descriptive (what is) and normative (what should be) models of an intervention and comparing these models. These models can be developed based on social science or systems theory (e.g., viable systems model of organizational functioning), computer simulation (e.g., systems dynamic model of a causal theory of change), or stakeholder deliberation (e.g., conversation about what is happening and what should be happening in an intervention). Models differ from criteria in that they address multiple dimensions in relationship to one another; include ways to weight or prioritize different dimensions; assess an intervention within a specified context; and are open to ongoing revision. However, criteria can be selected based on these models to render a criterion-based judgment. There are some STCS methodological approaches that are particularly useful for constructing and contrasting descriptive and normative models of a system or situation. For example, the viable systems model (VSM) (Hoverstadt, 2010) presents a diagram of the organizational structure and processes necessary for any system to meet the demands of surviving in a changing environment. Viability is the overall normative aim with many dimensions (e.g., variation within an organization matching variation in the environment) that together can be compared with a descriptive analysis of an organization in order to generate a value judgment of the organization and directions for future action. Another example is critical systems heuristics (CSH) that can be used to descriptively and normatively examine four sources of influence on an intervention (i.e.

sources of motivation, control, knowledge, legitimacy) (Reynolds, 2007; Ulrich & Reynolds, 2010; Reynolds & Williams, 2011).

5. Producing and Justifying Knowledge

Evaluations typically produce knowledge of the effects (e.g., outcomes, impact, sustainability) of social interventions. Analyzing and building arguments about causality is key in producing and justifying knowledge of effects. Issues regarding causality are continuously debated in evaluation including what constitutes credible evidence (Donaldson, Christie, & Mark, 2009) and what evaluation designs and methodologies support claims about the impact(s) of interventions (Stern, Stame, Mayne, Forss, Davies, & Befani, 2012; White & Philips, 2012). STCS draws attention to three issues regarding causality that have received less attention in these debates.

The first issue has to do with the emphasis on examining interventions-as-causes and determining how interventions produce effects and what effects. For example, experimental designs investigate interventions-as-causes; theory-based and process tracing approaches analyze causal processes and mechanisms within interventions; and contribution analysis examines the role of interventions within causal packages. The emphasis in STCS is on understanding dynamic complexity – the "often counterintuitive behavior of complex systems that arises from the interactions of the agents over time" (Sterman, 2006, p. 506). Causality is conceptualized as recursive; causes can also be effects and effects can also be causes. Causal processes are comprised of multiple, interdependent factors that influence one another through non-linear, cyclical feedback processes. Over time, these causal processes affect the structure and dynamical behavior of a system or situation. Methodologies for causal analysis analyze the structure and behavior of complex systems (e.g., causal loop diagramming, system dynamics).

The second issue is that models are significant in understanding causal relationships. However, unlike many approaches to generating models of social interventions (e.g., theory-based evaluation, realist evaluation) that view models as representations of the way things really are, models are constructions for a particular purpose and from a particular perspective to generate learning about how things are or ought to be. There is a well-known phrase in STCS – 'the map is not the territory' – reminding us that models are abstractions of reality and should not be confused with reality. This means that models do not aspire to be valid, in the sense of corresponding accurately to the real world, but rather to be useful for generating learning and action for a particular group of people in relation to an actual problem-situation. Causal claims made on the basis of modeling complex systems are uncertain and contingent. Sometimes, models illustrate patterns in the behavior of complex systems (e.g., system archetypes) that may apply to other social interventions and circumstances.

The third issue is the kinds of evidence relevant to producing knowledge of effects and the grounds on which such knowledge is justified. Models of causal processes affecting complex systems or situations often rely on theoretical and hypothetical data as well as empirical data. This is particularly evident in computer simulations in which the point is often to explore unexpected and emergent consequences by testing a variety of situations (Sterman, 2006). The widespread reductionist approach of adding up "detailed pieces of evidence into an accurate account of the costs (or efforts) and the consequences" (Ling 2012, p. 81), such as describing the inputs, processes, outputs, and outcomes followed by an overall evaluative conclusion, does not adequately capture emergent properties that stem from the interconnections of the parts of an interventions and from the interactions between an intervention and its environment. Therefore, STCS suggests that another approach to synthesizing evidence is needed. Viewing a system or

situation from multiple perspectives often enhances the quality of knowledge generated. Justifying this knowledge requires making transparent the boundaries and values that condition it (Midgley, Winstanley, & Foote, 2005). This means that the bases of evaluative arguments as reasoning, evidence and argument (Schwandt, 2015) must be expanded to include boundaries and values.

6. Facilitating Use

The ultimate activity in any evaluation is facilitating use of the evaluation. In the evaluation literature, use is broadly conceptualized as how those involved in or affected by an intervention intend to or actually use the evaluation. Use can be instrumental, conceptual, symbolic, or process oriented (Weiss, 1998; Schwandt, 2015). STCS challenges this way of thinking about use as user-centered and coming near the end or after the end of an evaluation. Rather, in STCS use can be alternatively thought of as a design feature of the intervention. Evaluations and using evaluations are embedded in the ongoing management of the intervention. There are three ways evaluators can design evaluations as a feature of the intervention to be used to guide ongoing management: as a feedback loop to provide real-time information; as social learning to guide ongoing governance; or as normative reflection, critique, and direction setting.

Designing an evaluation as a feedback loop that provides real-time information offers another way of conceptualizing use (Ling, 2012). The role of the evaluator shifts from solely providing information to actually facilitating change and adaptive management: "Evaluators can use their experiences and expertise to focus on learning an as an adaptive mechanism" (Eoyang & Berkas, 1999). This requires evaluators to assess communication and information flows in the intervention and to design aspects of the evaluation to facilitate use of evaluation data in these ongoing exchanges.

Designing an evaluation to facilitate social learning offers another way of facilitating use. Social learning is "a complex, time-dependent process that involves, in addition to the action itself (which breaks into the stream of ongoing events to change reality), political strategy and tactics (which tell us how to overcome resistance), theories of reality (which tell us what the world is like), and the values that inspire and direct the action" (Friedman 1987 in Reynolds, 2007, p. 9). Social learning can take place in the evaluation team and among those involved or affected by the intervention. Social learning is similar to models of team science that move "beyond simple linear methods of drawing association, to methods that enable multiple sources of data from different aspects of the scientific enterprise to inform the evaluation. Such multimodality may require modeling such different perspectives in the composition of the evaluation team itself" (Norman, Best, Mortimer, Huerta, & Buchan, 2011, p. 80). Social learning draws on critical pedagogy, participatory action research, soft systems methodology, and other group reflection-learning-action approaches. Some constraints on and risks to learning in groups include defensive routines and group think, suppress dissent, ignore disconfirming evidence, cognitive and group errors, time delays (Sterman, 2006).

A third way to facilitate use suggested by STCS is to design an evaluation to generate normative reflection, critique, and direction setting. Aspiring to this kind of evaluation use requires moving beyond single loop (i.e. how existing activities can be done better) and double loop (i.e. whether those are the right activities to do) questions to triple loop questions (i.e. what makes these the right things to do) (Hummelbrunner & Reynolds, 2013). In addition to raising triple-loop learning issues, evaluators can actually facilitate discussion and negotiation among stakeholders about normative matters.

Conclusion

Agencies aiming to address social problems that are increasingly recognized as wicked (i.e., exhibiting interconnectedness, uncertainty, dynamism) are turning to adaptive management approaches (Stern, Saunders, & Stame, 2015; Forss, 2011; Sanderson, 2000; APS, 2007). Subsequently, these agencies request evaluations that inform ongoing learning about social problems, the extent to which social interventions address these problems, and what actions should be taken next. To conduct such evaluations, evaluators will need to draw on insights from STCS.

Assuming that the importation of ideas and practices from STCS into evaluation continues, the need grows for more systematic and critical examinations of its ideas. This paper took a first step in outlining some of the major implications of STCS for how we think about and practice six activities central to most evaluations. For these implications to make a difference in evaluation practice, more careful attention will likely need to be directed at reconsidering the role of evaluation in governing and managing ongoing change, examining the institutional contexts for conducting evaluations, and revisiting evaluators' training, capacities, and responsibilities. At present, evaluations primarily are oriented around assessing social interventions not learning about social problems or normative direction setting. STCS suggests that this role be expanded. Evaluators often work for specific evaluation commissioners on short-term, program-centered contracts and not ongoing cross-institution and sector, problem-centered change efforts. Evaluators' contracts will need to shift to fully take up implications for supporting social problem solving and facilitating use. Evaluators' training is usually limited to one substantive area (e.g., public health, education, policy analysis) along with extensive social science methodological training. This training needs to be broadened to include more inter-

37

disciplinary substantive training and familiarity with system and complexity theories and

methodologies.

REFERENCES

Alkin, M. C., & Christie, C. A. (2003). An evaluation theory tree. *Evaluation*, 12-65.

Arkesteijn, M., van Mierlo, B., & Leeuwis, C. (2015). The need for reflexive evaluation approaches in development cooperation. *Evaluation, 21*(1), 99-115.

Attenborough, K. (2007). Soft Systems in a Hardening World: Evaluating Urban Regeneration. In B. Williams & I. Imam (Eds.), (pp. 75-88). Point Reyes, California: EdgePress.

Australian Public Service (APS) (2007). Tackling wicked problems: A public policy perspective. *Australian Government*.

Barnes, M., Matka, E., & Sullivan, H. (2003). Evidence, understanding and complexity: evaluation in non-linear systems. *Evaluation, 9*(3), 265-284.

BetterEvaluation (2015). Develop initial description. Retrieved from: http://betterevaluation.org/plan/define/develop_initial_description. [Last accessed: October 23, 2015]

Boyd, A., Geerling, T., Gregory, W. J., Kagan, C., Midgley, G., Murray, P., & Walsh, M. P. (2007). Systemic evaluation: a participative, multi-method approach. *Journal of the Operational Research Society, 58*(10), 1306-1320.

Byrne, D. (2013). Evaluating complex social interventions in a complex world. *Evaluation, 19*(3), 217-228.

Cabrera, D., & Trochim, W. M. K. (2006). A protocol of systems evaluation. In Cabrera, D. (Ed.). (2006). Systems evaluation and evaluation systems whitepaper series. Ithaca, NY: Cornell University Dspace Open Access Repository. National Science Foundation Systems Evaluation Grant No. EREC-0535492.

Cabrera, D., Colosi, L., & Lobdell, C. (2008). Systems thinking. *Evaluation and Program Planning, 31*(3), 299-310. doi: 10.1016/j.evalprogplan.2007.12.001

Callaghan, G. (2008). Evaluation and negotiated order: developing the application of complexity theory. *Evaluation, 14*(4), 399-411.

Checkland, P. (1999). *Soft Systems Methodology: A 30-year Retrospective*. West Sussex, England: John Wiley & Sons, Ltd.

Christie, C. A., & Alkin, M. C. (2013). An Evaluation Theory Tree In M. C. Alkin (Ed.), *Evaluation Roots: A Wider Perspective of Theorists' Views and Influences* (pp. 11-57). Thousand Oaks, California: SAGE Publications, Inc.

Davidson, J. (2013). *Actionable Evaluation Basics*. Auckland, New Zealand: Real Evaluation, Ltd.

Davies, R. (2004). Scale, complexity and the representation of theories of change: Part II. *Evaluation, 11*(2), 133-149.

Durland, M. M., & Fredericks, K. (2005). An introduction to social network analysis. *New Directions for Evaluation*(107), 5-13.

Donaldson, S., Christie, C., & Mark, M. (Eds.) (2009). *What Counts as Credible Evidence in Contemporary Evaluation Practice?* Los Angeles, California: SAGE Publications, Inc.

Dyehouse, M., Bennett, D., Harbor, J., Childress, A., & Dark, M. (2009). A comparison of linear and systems thinking approaches for program evaluation illustrated using the Indiana Interdisciplinary GK-12. *Evaluation and Program Planning, 32*(3), 187-196.

Eoyang, G. H., & Berkas, T. H. (1999). Evaluating Performance in a Complex Adaptive System. In M. Lissack & H. Gunz (Eds.), (pp. 1-21). Westport, Connecticut: Quorum Books.

Eoyang, G. H., & Holladay, R. (2013). *Adaptive Action: Leveraging Uncertainty in Your Organization*. Stanford, California: Stanford University Press.

Fitzpatrick, J. L. (2012). An introduction to context and its role in evaluation practice. In D. J. Rog, J. L. Fitzpatrick, & R. F. Conner (Eds.), *Context: A framework for its influence on evaluation practice. New Directions for Evaluation, 135*, 7-24.

Forss, K., Marra, M., & Schwartz, R. (Eds.) (2011). *Evaluating the Complex: Attribution, Contribution, and Beyond.* Vol. 18. London, United Kingdom: Transaction Publishers.

Forss, K., & Schwartz, R. (2011). Introduction. In K. Forss, M. Marra, & R. Schwartz (Eds.), *Evaluating the Complex: Attribution, Contribution, and Beyond* (Vol. 18). London, United Kingdom: Transaction Publishers.

Foster-Fishman, P. G., Nowell, B., & Yang, H. (2007). Putting the system back into systems change: a framework for understanding and changing organizational and community systems. *American Journal of Community Psychology, 39*(3-4), 197-215.

Foster-Fishman, P. G., & Watson, E. R. (2012). The ABLe change framework: a conceptual and methodological tool for promoting systems change. *American Journal of Community Psychology, 49*(3-4), 503-516.

Fredericks, K., Deegan, M., & Carman, J. (2008). Using system dynamics as an evaluation tool: experience from a demonstration program. *American Journal of Evaluation, 29*(3), 251-267.

Fujita, N. (ed.) (2010). *Beyond logframe; using systems concepts in evaluation*. Tokyo, Japan: Foundation for Advanced Studies on International Development.

GIZ (2011). *Systemic Approaches in Evaluation: Documentation of the Conference on 25-26 January 2011*. Retrieved from: http://www.evaluation-conference.de/en/index.html. [Last accessed: October 23, 2015]

Gopalkrishanan, S., Preskill, H., & Lu, S. (2013). Next Generation Evaluation: Embracing Complexity, Connectivity, and Change: FSG.

Greene, J. C. (2005). Context. In S. Mathison (Ed.), *Encyclopedia of Evaluation* (pp. 82-84). Thousand Oaks, California: SAGE Publications, Inc.

Gregory, A. (1997). Evaluation practice and the tricky issue of coercive contexts. *Systems Practice, 10*(5), 589-609.

Gregory, A., & Jackson, M.C. (1992a). Evaluating organizations: a systems and contingency approach. *Systems Practice, 5*(1), 37-60.

Gregory, A., & Jackson, M. C. (1992b). Evaluation methodologies: a system for use. *Journal of the Operational Society, 43*(1), 19-28.

Grove, J. T. (2015). Aiming for Utility in 'Systems-based Evaluation': A Research-based Framework for Practitioners *IDS Bulletin* (Vol. 46). Oxford, United Kingdom: Institute of Development Studies.

Hargreaves, M. B. (2010). Evaluating System Change: A Planning Guide (pp. 1-20). Princeton, New Jersey.

Hawe, P., Bond, L., & Butler, H. (2009). Knowledge theories can inform evaluation practice: what can a complexity lens add? *New Directions for Evaluation* (124), 89-100. 9

Henry, G. T. (2002). Choosing criteria to judge program success: a values inquiry. *Evaluation, 8*(2), 182-204.

Hoverstadt, P. (2010). The Viable Systems Model. In M. Reynolds & S. Holwell (Eds.), *Systems Approaches to Managing Change: A Practical Guide* (pp. 87-133). London, United Kingdom: Springer.

Hummelbrunner, R. (2010). Beyond logframe: critique, variations, and alternatives. In N. Fujita (Ed.), *Beyond logframe; using systems concepts in evaluation* (pp. 1-33). Japan: Foundation for Advanced Studies on International Development.

Hummelbrunner, R. (2011). Systems thinking and evaluation. *Evaluation, 17*(4), 395-403.

Hummelbrunner, R., & Reynolds, M. (2013). Systems Thinking, Learning and Values in Evaluation (pp. 9-10): European Evaluation Society.

Jones, H. (2011). *Taking responsibility for complexity: How implementation can achieve results in the face of complex problems*. ODI Working Papers. London, United Kingdom. Retrieved from http://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/6485.pdf. [Last accessed: October 23, 2015]

Kusek, J. Z., & Rist, R. C. (2004). Ten Steps to a Results-based Monitoring and Evaluation System. *The International Bank for Reconstruction and Development*. Washington, D.C.: The World Bank.

Leischow, S., Best, A., Trochim, W., Clark, P., Gallagher, R., Marcus, S., & Matthews, E. (2008). Systems thinking to improve the public's health. *American Journal of Preventive Medicine, 35*(2 Suppl), S196-203.

Leischow, S., & Milstein, B. (2006). Systems thinking and modeling for public health practice. *American Journal of Public Health, 96*(3), 403-405.

Levin-Rozalis, M. (2014). *Let's Talk Program Evaluation in Theory and Practice* Tel Aviv, Israel Dekel Academic Press.

Ling, T. (2012). Evaluating complex and unfolding interventions in real time. *Evaluation, 18*(1), 79-91.

Mark, M., Henry, G. T., & Julnes, G. (2000). *Evaluation: An Integrated Framework for Understanding, Guiding, and Improving Public and Nonprofit Policies and Programs*. San Francisco, California: Jossey-Bass, A Wiley Company.

Marra, M. (2011a). Micro, Meso, and Macro Dimensions of Change: A New Agenda for the Evaluation of Structural Policies. *Evaluating the Complex: Attribution, Contribution, and Beyond* (pp. 97-121). New Brunswick, New Jersey Transaction Publishers.

Marra, M. (2011b). Some Insights from Complexity Science for the Evaluation of Complex Policies. In K. Forss, M. Marra, & R. Schwartz (Eds.). New Brunswick, New Jersey: Transaction Publishers.

Midgley, G. (1996). Evaluating services for people with disabilities: a critical systems perspective. *Evaluation, 2*(1), 67-84.

Midgley, G., Winstanley, A., & Foote, J. (2005). Scoping the potential uses of systems thinking in developing policy on illicit drugs: Institute of Environmental Science and Research (ESR).

Milstein, B. (2008). Hygeia's Constellation: Navigating Health Futures in a Dynamic and Democratic World (pp. 1-124).

Morell, J. A. (2010). *Evaluation in the Face of Uncertainty: Anticipating Surprise and Responding to the Inevitable* New York, New York: The Guilford Press.

Morell, J. A., Hilscher, R., Magura, S., & Ford, J. (2010). Integrating Evaluation and Agent-based Modeling: Rationale and an Example for Adopting Evidence-based Practices. *Journal of Multidisciplinary Evaluation, 6*(14), 32-57.

Mowles, C. (2014). Complex, but not quite complex enough: The turn to the complexity sciences in evaluation scholarship. *Evaluation, 20*(2), 160-175.

Norman, C., Best, A., Mortimer, S., Huerta, T., & Buchan, A. (2011). Evaluating the science of discovery in complex health systems. *American Journal of Evaluation, 32*(1), 70-84.

Parsons, B. (2007). Initiative Evaluation Guide: A Systems-Oriented Framework for Evaluating Social Change Efforts. W.K. Kellogg Foundation.

Parsons, B., Keene, M., & Dhillon, L. (2014). *Visionary evaluation for a sustainable, equitable future*. Retrieved from http://www.eval.org/evaluation2014. [Last accessed: October 23, 2015]

Patton, M. Q. (2005). Situational Responsiveness. In S. Mathison (Ed.), *Encyclopedia of Evaluation* (pp. 390). Thousand Oaks, California: SAGE Publications, Inc.

Patton, M. Q. (2011). *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*. New York, New York: The Guilford Press.

Patton, M. Q. (2012). Contextual pragmatics of valuing. *New Directions for Evaluation, 133*. Promoting valuation in the public interest: Informing policies for judging value in evaluation, 97-108.

Preskill, H., & Gopalkrishanan, S. (2014). Evaluating Complexity: Propositions for Improving Practice: FSG.

Ramalingam, B. (2013). *Aid on the Edge of Chaos: Re-thinking International Cooperation in a Complex World*. United Kingdom: Oxford University Press.

Ramalingam, B., Jones, H., Reba, T., & Young, J. (2008). *Exploring the science of complexity: Ideas and implications for development and humanitarian efforts*. Retrieved from http://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/833.pdf [Last accessed: October 23, 2015]

Reynolds, M. (2007). Framing purposeful evaluation through critical systems thinking. *Rural Development for the XXIInd Congress of the European Society for Rural Sociology*.

Reynolds, M. (2014). Equity-focused developmental evaluation using critical systems thinking. *Evaluation, 20*(1), 75-95.

Reynolds, M., Forss, K., Hummelbrunner, R., Marra, M., & Perrin, B. (2012). Complexity, systems thinking and evaluation - an emerging relationship? *Evaluation Connections: Newsletter of the European Evaluation Society*, 7-9.

Reynolds, M., & Williams, B. (2011). Systems Thinking and Equity-focused Evaluation (pp. 115-141): UNICEF.

Rodriguez-Campos, L. (2012). Stakeholder Involvement in Evaluation: Three Decades of the American Journal of Evaluation *Journal of Multidisciplinary Evaluation, 8*(17), 57-79.

Rogers, P. J. (2008). Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions. *Evaluation, 14*(1), 29-48. doi: 10.1177/1356389007084674

Rossi, P. H., Lipsey, M. W., & Freeman, H. W. (2004). Assessing the Need for a Program (Seventh ed., pp. 101-132). Thousand Oaks, California: Sage Publications, Inc.

Sanderson, I. (2000). Evaluation in complex policy systems. *Evaluation, 6*(4), 433-454.

Sanderson, I. (2009). Intelligent Policy Making for a Complex World: Pragmatism, Evidence and Learning. *Political Studies, 57*(4), 699-719.

Schwandt, T.A.(1992). Better living through evaluation? Images of progress shaping evaluation practice. *Evaluation Practice, 13*(2), 135-144.

Schwandt, T. A. (2005). Value Judgment & Values. In S. Mathison (Ed.), *Enccylopedia of Evaluation.* Thousand Oaks, California: SAGE Publications, Inc.

Schwandt, T. A. (2015). *Evaluation Foundations Revisited: Cultivating a Life of the Mind For Practice* Stanford, CA Stanford University Press.

Scriven, M. (2007). Key evaluation checklist. (February). Retrieved from: http://www.michaelscriven.info/images/KEC_3.22.2013.pdf [Last accessed: October 23, 2015]

Shadish, W. R., Cook, T. D., & Leviton, L. C. (1991). *Foundations of Program Evaluation: Theories of Practice*. Newbury Park, California: Sage Publications, Inc.

Shiell, A., Hawe, P., & Gold, L. (2008). Complex interventions or complex systems? implications for health eocnomic evaluation. *BMJ, 336*(1281-1283).

Simons, H. (2004). Utilizing evaluation evidence to enhance professional practice. *Evaluation, 10*(4), 410-429.

Stame, N. (2004). Theory-based evaluation and types of complexity. *Evaluation, 10*(1), 58-76.

Sterman, J. D. (2006). Learning from evidence in a complex world. *American Journal of Public Health, 96*(3), 505-514.

Stern, E., Saunders, M., & Stame, N. (2015). Standing back and looking forward: Editors' reflections on the 20th Anniversary of Evaluation. *Evaluation, 21*(4), 380-390.

Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R., & Befani, B. (2012). *Broadening the Range of Designs and Methods for Impact Evaluations*. Working Paper.

Tapella, E., & Rodriguez-Bilella, P. (2014). Shared learning and participatory evaluation: The sistematizacion approach to assess development interventions. *Evaluation, 20*(1), 115-133.

Trochim, W., Cabrera, D., Milstein, B., GAllagher, R., & Leischow, S. (2006). Practical challenges of systems thinking and modeling in public health. *American Journal of Public Health, 96*(3), 538-546.

Ulrich, W. (1988). Churchman's "Process of unfolding" and Its significance for policy analysis and evaluation. *Systems Practice, 1*(4), 415-428.

Ulrich, W. (2005). A Brief Introduction to Critical Systems Heuristics (CSH). (October). Retrieved from: http://wulrich.com/downloads/ulrich_2005f.pdf [Last accessed: October 23, 2015]

Ulrich, W., & Reynolds, M. (2010). Critical Systems Heuristics. *Systems Approaches to Managing Change: A Practical Guide* (pp. 243-292). London, United Kingdom: Springer London.

Walton, M. (2014). Applying complexity theory: a review to inform evaluation design. *Evaluation and Program Planning, 45*, 119-126.

Wasserman, D. (2010). Using a systems orientation and foundational theory to enhance theory-driven human service program evaluations. *Evaluation and Program Planning, 33*(2), 67-80.

Weiss, C. (1998). *Evaluation* (2nd ed.). Upper Saddle River, New Jersey: Prentice Hall.

White, H., & Phillps, D. (2012). *Addressing attribution of cause and effect in small n impact evaluations: towards an integrated framework*. Working Paper. New Delhi, India. Retrieved from www.3ieimpact.org [Last accessed: October 23, 2015]

Williams, B., & Hummelbrunner, R. (2011). *Systems Concepts in Action: A Practitioner's Toolkit*. Stanford, California: Stanford University Press.

Williams, B. & Imam, I. (Eds.) (2007). *Systems Concepts in Evaluation: An Expert Anthology*. Point Reyes, CA: EdgePress of Inverness.

Williams, B., & Rogers, P. (2010). Using systems concepts in evaluation - a dialogue with Patricia Rogers and Bob Williams. In N. Fujita (Ed.), *Beyond logframe; using systems concepts in evaluation* (pp. 55-74). Japan: Foundation for Advanced Studies on International Development.

Wolf-Branigin, M. (2012). Introduction to the mini special issue on complexity theory: a new research paradigm in social work. *Journal of Social Service Research, 38*(5), 580-581.

Wolf-Branigin, M. (2013). *Using Complexity Theory for Research and Program Evaluation.* Oxford, United Kingdom: Oxford University Press.

PAPER 2

**BORROWING KNOWLEDGE FROM THE SYSTEMS AND COMPLEXITY FIELDS:
IMPLICATIONS FOR EVALUATION PRACTICE**

From its very beginning as a field of practice, evaluation has drawn on ideas, theories, and methodologies from other fields and disciplines. Consider, for example, the influences of sociology, psychology, economics, and educational research on the early development of evaluation (Campbell, 2011; Cronbach & Associates, 1980); the interaction of anthropology and evaluation including use of ethnographic approaches to examine context (Fetterman, 1988; Britan, 1978; Butler & Copeland-Carson, 2005); and contributions of organizational theories of performance and learning with regards to evaluation use (Russ-eft & Preskill, 2009). Other disciplines continue to shape evaluation in important ways (Jacob, 2008).

In the last twenty years, many prominent evaluators have been turning to the systems and complexity fields (Eoyang, 1999; Sanderson, 2000; Barnes et al, 2003; Williams and Imam, 2007; Callaghan, 2008; Rogers, 2008; Morell, 2010; Forss, Marra, & Schwartz, 2011; Westhorp, 2012; Byrne, 2013). The significance and implications of this turn for evaluation practice are not yet clear (Mowles, 2014; Stern, Saunders, & Stame, 2015). Moreover, while interest in systems and complexity concepts (e.g., interrelationships, emergence, non-linearity) and methodologies (e.g., system dynamics, agent-based modeling) has grown rather rapidly and broadly throughout the evaluation community, significant questions remain regarding what systems and complexity-informed evaluations entail and whether and how they differ from what already constitutes good evaluation practice (Datta, 2008; Rogers, 2008).

There are several challenges to understanding the implications of the systems and complexity fields for evaluation practice. Those fields are much broader than a single discipline and are comprised of many research communities and sub-communities (e.g., general systems

theory, operations research, complexity science, cybernetics, systems science) (Midgley, 2007).

Boundaries between these and more traditional academic disciplines and communities (e.g.,

biology, computer science, engineering) are blurry and continuously crossed (Hieronymi, 2013).

The theories, concepts, methodologies, and tools that are central to these fields often have

multiple, and sometimes incompatible, definitions and numerous variations[2]. As evaluators select

and translate theories and methodologies they will inevitably be defined and applied differently

so as to be relevant to and useful for evaluators. Finally, evaluators drawing on the systems and

complexity fields are doing so within different evaluation theories (e.g., theory-based,

responsive, equity-focused), for evaluating different kinds of interventions (e.g., social

innovations, health care policy, regional development), and in different circumstances. For these

reasons and others, it is challenging to make over-arching claims about the implications of the

systems and complexity fields for evaluation practice.

Despite these challenges, there are three significant reasons why there is both a need for

and potential value in examining the implications of the systems and complexity fields for

evaluation practice. First, evaluation commissioners[3] and stakeholders are developing new kinds

of social interventions[4], for example networks, emerging innovations, and systems change, and

requesting evaluations that are compatible with these interventions and the change processes they

advance. This is changing the circumstances in which many evaluators practice to increasingly

require evaluations that use systems and complexity ideas and approaches, particularly in

international development (e.g., Jones 2012; Ramalingham, 2013), public health (e.g., Milstein,

---

[2] For example, there is little agreement on what constitutes systems thinking (Cabrera, 2006; Cabrera, Colosi, & Lobdell, 2008; Henning and Chen, 2012); there is no generally shared systems theory or definition of a system (Hieronymi, 2013); and there are hundreds of approaches to systems-based inquiry (Williams, 2005).
[3] Numerous federal agencies and private foundations in the United States as well as international development agencies have launched and funded initiatives and commissioned evaluations that explicitly use the language of systems and complexity.

2008), social services (e.g., Wolf-Branigin, 2012), and community development (e.g., Foster-Fishman, Nowell, and Yang, 2007). The rapid, international growth of developmental evaluation illustrates this demand for new approaches to evaluation (Patton, 2016). Second, there are many evaluators who claim to be using systems and complexity ideas and methods, such as systems thinking and complexity science, in evaluation practice[5] (Patton. 2016). Yet, there has been very little empirical research examining the character and extent of use of these ideas and methods in evaluation practice. Third, a growing interest in research on evaluation has led to examining how evaluation theorists practice evaluation as a means to better understand how theory informs the practical considerations and choices evaluators face (Chelimsky, 1998; Alkin and Christie, 2005; Fitzpatrick, Christie, and Mark, 2009). Studying the practice experiences of evaluators who are explicitly using ideas, theories, and methodologies from other disciplines, such as the systems and complexity fields, offers another avenue for studying the theory-practice relationship in evaluation[6].

This paper examines what borrowing ideas and techniques from the systems and complexity fields entails and with what implications for evaluation practice. *Borrowing* refers to the translation and application of an array of ideas, assumptions, concepts, theories, and techniques (i.e. methodologies, methods, heuristics, tools) developed in the systems and complexity fields and used in evaluation practice. This paper reports the results of a study of

---

[5] In a speech at the 2014 American Evaluation Association (AEA) meeting, Bob Williams asked those who use systems thinking in their evaluation work to stand and remain standing and for others to be seated. At this event attended by more than 400 evaluators, only a handful of evaluators sat down (Patton, 2016). This is just one indication of the prevalence of systems and complexity in evaluation. Other indications include the rise in evaluation cases addressing systems or complexity published in prominent evaluation journals and the numerous guides to conducting evaluation using various systems and complexity concepts and methods (e.g., Parsons, 2007; Hargreaves, 2010).

[6] Scholars broadly interested in cross-disciplinary exchanges and knowledge production across and between the academic disciplines have made similar examinations in order to re-think what defines a particular discipline or practice (Gibbons, 1994; Kellert, 2008; Chambon, 2012).

eight cases of evaluation practice employing systems and complexity thinking and techniques as defined by particular evaluators. Materials for analyzing each case include semi-structured interviews with each evaluator and documents (e.g., journal articles, book chapters, and evaluation plans and reports) describing evaluations these evaluators conducted. Evaluation practice is defined as how evaluators think about what being an evaluator means and what they are doing when they are practicing evaluation, specifically, the activities evaluators carry out when designing and conducting evaluations in particular circumstances.

Overview of Borrowing from the Systems and Complexity Fields in Evaluation

Borrowing knowledge from the systems and complexity fields is not new in the evaluation field. Systems thinkers including Kurt Lewin, Russell Ackoff, Chris Argyris, and Donald Schön were influential in the origins of the evaluation field and continue to influence the works of evaluation scholars (Ramage & Shipp, 2009; Chelimsky, 1998; Schwandt, 2015). Many evaluators are formally trained in academic disciplines that have been influenced by and influence the systems and complexity fields including sociology, political science, natural sciences, and psychology as well as methodological and practical traditions to which systems theorists and practitioners have contributed. In the 1980's and 1990's, scholars in the systems fields including Werner Ulrich (1988), Amanda Gregory and Michael Jackson (1992ab), and Gerald Midgley (1996) connected their work to evaluation.

However, in many ways, what marks the beginning of evaluators investigating these ideas in the United States was a two-day meeting in Berkeley, California in 2004, which led to the book *Systems Concepts in Action: An Expert Anthology* edited by Bob Williams and Iraj Imam. The Systems in Evaluation TIG was formed at the American Evaluation Association in 2004 with four members and has rapidly grown to include members from over 35 countries

(SETIG website, 2016). Since the publication of the anthology, numerous books have been

published applying systems and complexity ideas to evaluation (Williams & Hummelbrunner,

2011; Morell, 2010; Patton, 2011; Forss, Marra, and Stern, 2011; Levin-Rozalis, 2014; Patton,

McKegg, and Weipeihana, 2016). New methods from the systems and complexity fields are

being used in evaluation practice including causal loop diagrams and system dynamics

(Fredericks, Deegan, & Carman, 2008; Dyehouse, Bennett, Harbor, & Childress, 2009); agent-

based modeling (Morell, Hilscher, Magura, & Ford, 2010); soft systems methodology

(Attenborough, 2007); social network analysis (Durland & Fredericks, 2005); and critical

systems heuristics (Reynolds & Williams, 2011). Evaluators have developed new conceptual

frameworks and guides for practice based on ideas from the systems and complexity fields (e.g.,

Cabrera & Trochim, 2006; Parsons, 2007; Hargreaves, 2010; Gopalkrishnan, Preskill, & Lu,

2013; Preskill & Gopalkrishanan, 2014; Marra, 2011a; Cabrera, Colosi, & Lobdell, 2008;

Wasserman, 2010). In reflecting on the twentieth anniversary of the journal *Evaluation: An*

International Journal of Theory, Research, and Practice, Stern, Saunders, and Stame (2015)

observe that, "over the last 20 years, the acknowledgement of 'complexity' as a way of

understanding contemporary social and economic processes and forms has emerged as probably

the single most intrusive 'elephant in the room' for evaluators" (p. 385).

At present, there has been some research on and discussion of the implications of the

systems and complexity fields for evaluation, but little empirical research investigating the

implications for evaluation practice. Stern, Saunders, and Stame (2015), for example, assert,

"how to take complexity into account and how to accommodate it, defining what we mean by

complexity, distinguishing complexity science from general systems theory and critical systems

heuristics remains open to debate. Some progress is being made at the conceptual level…but the

consequences and subsequent uses of systems and complexity thinking remain opaque" (p. 385).

Significant research efforts thus far include reviews of the literature by Mowles (2014) and

Walton (2014); numerous case studies including but not limited to those in Williams and Imam

(2007) and Patton et. al (2016); numerous conceptual papers examining one or several issues

from the systems and complexity fields in connection with evaluation; briefs promoting

reflection and discussion about these ideas (FASID, 2011; IDS, 2015); and recent issues of

Evaluation Connections. However, a considerable gap in this research is the lack of empirical

investigations of what borrowing entails and its limitations, added value, and challenges for

evaluation practice.

Conceptualizing Evaluation Practice

There are differing views on how to conceptualize evaluation practice. This is the case

for at least three reasons. First, as Schwandt (2015) contends, evaluation practice is inherently a

particular, situated phenomenon that cannot be described as an arena or site in which

practitioners routinely follow clearly defined tasks and associated procedures:

> Practice is indeterminate because choices of the appropriate and effective actions to take
> in dealing with others arise within specific circumstances and are thus contextually
> relative. Practice is concerned with the particular (rather than the general) precisely
> because it is about an evaluator taking the right action in consideration of *this* situation,
> with *these* people, at *this* time and place, in *this* set of conditions. (p. 44).

Therefore, different evaluators working with different evaluation commissioners and

stakeholders, at different times and places, and under different conditions will inevitably

conceive of evaluation practice differently.

Second, while there are over-arching theories of evaluation practice (e.g., theory-based,

responsive, participatory), research on evaluation has found that these theories do not neatly map

onto what evaluators actually do. This is so because evaluation theory is not descriptive or

straightforwardly prescriptive but, rather, as Schwandt (2014) defines it, consists of the

"repertoires of concepts, insights, explanations, and tools that professional practitioners can use as heuristics, tools to think with" (p. 234). There are numerous theories of evaluation practice, as well as ideas, concepts, and tools, that any one evaluator can draw on to inform how he or she carries out a specific evaluation. Further, the same evaluator may draw on different theories or interpret and apply the same theory differently in different circumstances and in a different way than another evaluator might.

Third, while, in general, there are some tasks or activities that evaluators carry out, these are uniquely influenced by the characteristics of evaluators conducting the evaluations and the circumstances in which they practice. For example, the Better Evaluation Rainbow Framework defines evaluation practice in terms of seven tasks as shown in Table 1. However, evaluators may not carry out all of these tasks; they typically do not carry out these tasks as discrete activities or events or in a sequential manner, as they appear in this table; and discerning how evaluators carried out any one of these tasks in a particular evaluation can be difficult because, sometimes, they are rather implicit in an evaluator's practice and case descriptions and evaluation reports may not explicitly discuss how they were carried out.

Table 2.1. Better Evaluation Rainbow Framework: Tasks and Descriptions

Task	Description
1. Manage an evaluation or evaluation system	Manage an evaluation (or a series of evaluations), including deciding who will conduct the evaluation and who will make decisions about it.
2. Define what is to be evaluated	Develop a description (or access an existing version) of what is to be evaluated and how it is understood to work.
3. Frame the boundaries for an evaluation	Set the parameters of the evaluation – its purposes, key evaluation questions and the criteria and standards to be used.
4. Describe activities, outcomes, impact, and context	Collect and retrieve data to answer descriptive questions about the activities of the project/program/ policy, the various results it has had, and the context in which it has been implemented.
5. Understand causes of outcomes and impact	Collect and analyse data to answer causal questions about what has produced outcomes and impacts that have been observed.
6. Synthesize data from one or more evaluations	Combine data to form an overall assessment of the merit or worth of the intervention, or to summarize evidence across several evaluations.
7. Report and support use of findings	Develop and present findings in ways that are useful for the intended users of the evaluation, and support them to make use of them).

Acknowledging these variations in conceptualizing evaluation practice, I began this study conceptualizing evaluation practice using the seven Better Evaluation tasks. I used this as an initial heuristic to inform my interview questions about how evaluators designed and conducted an evaluation and as an initial analytic device when I reviewed the interview data and case materials. However, I found that how evaluators described how they carried out a specific evaluation was informed by and interwoven with what they thought, more generally, about the role and practice of evaluation. Therefore, in what follows, I refer to evaluation practice simultaneously as how evaluators think about the practice of evaluation and what they do when they practice evaluation in particular circumstances.

Study Design

The manner in which this study of eight cases of evaluation practice was designed and conducted is described below. The study involved examining eight cases of evaluation practice employing systems and complexity thinking and techniques as defined by particular evaluators. Three questions guided this study: 1) What ideas and techniques are these evaluators borrowing in their evaluation practices? 2) How do these ideas and techniques influence the way these evaluators design and conduct evaluations?

Selection of Cases

Cases were chosen for inclusion in this study in three ways. First, I identified published evaluations that used systems and complexity thinking and techniques by searching academic journals using EBSCOHost, JSTOR, and other search engines using the terms *systems*, *complexity*, *evaluation*, and *case* and reviewing evaluation cases published in the *Systems Concepts in Evaluation: An Expert Anthology*. Second, I used a maximum variation sampling plan to identify a collection of evaluations that met the following criteria: different systems or complexity approaches (e.g. system dynamics, soft systems) were used; different types of interventions (e.g., public health, higher education) were evaluated; each evaluation involved a theoretically grounded approach or methodology; each evaluation applied the systems or complexity approach in an empirical case and provided explicit description of how the approach was used; and the lead evaluator had at least ten years of experience with systems and complexity thinking and techniques. Third, the lead evaluator of the case had to be willing to participate in an interview about how he/she was applying systems and complexity thinking and techniques in one or several evaluations he/she was involved in conducting. After identifying potential evaluations and evaluators that met the first two criteria, I sent an e-mail invitation to

the eight evaluators including information about the research purpose and questions, my purpose for the interview, interview topics, and procedures regarding participation, audiotaping and transcribing data, and presentation of data in the study report. All eight evaluators agreed to participate in the study.

Data Gathering

Semi-Structured Interviews. Interviews were conducted during the first three weeks of January 2016 over the phone or Skype and lasted between sixty and ninety minutes each. At least one week prior to each interview, respondents were provided with a general overview of the three topics to be discussed in the interview: evaluator's background in evaluation and the systems and complexity fields; examples from practice of using systems thinking and/or complexity science in evaluation(s); and the evaluator's views on the interest in systems thinking and complexity science in the evaluation field. An interview guide of open-ended, descriptive and interpretive questions was used during each interview to make sure key topics were explored with all participants (see Appendix A for Interview Guide). Evaluators were asked to discuss examples from practice that had some publicly accessible materials (e.g., evaluation plan, report, project description, academic journal article, or book chapter). All evaluators were informed of their voluntary consent to participate and, importantly, their option to be identified by name or pseudonym in this study. All evaluators agreed to be identified by name. Any information that an evaluator did not want to be reported identifiably is reported anonymously. Member checks were used periodically during interviews to clarify that the interviewer accurately understood what an evaluator was saying, and all evaluators received brief summaries of the interview and the option to expand on or change information. Interviews were transcribed verbatim using an independent,

professional company. Copies of transcriptions were made available to all evaluators and provided to those who requested one.

       Case Materials. Documents related with one or several evaluations that the evaluator conducted were reviewed as supplementary materials to the interviews. Materials were identified by the researcher using search engines (e.g., EBSCOHost, Google) and, in some instances, were mentioned or provided via email by the evaluator during the interview. These included theoretical articles that informed the approaches that a particular evaluator used in practice; journal articles or book chapters describing evaluations the evaluator conducted or his/her evaluation practice more generally; and original evaluation plans and reports when publicly available. Articles reviewed in this study are only a small part of each evaluator's research and evaluation practice and do not represent the extent or variety of their views or the evaluations they conduct. Additionally, multiple evaluators conducted many of the evaluations reviewed; these other evaluators were not interviewed in this study. Appendix B provides a table of all materials reviewed in this study.

Case Analysis

       Each case was first examined individually in the following way: First, I carefully read interview summaries, interview transcripts, and other materials related with each case. Second, I completed a data display identifying themes and examples of how evaluators in each case carried out the seven tasks outlined in the Better Evaluation Rainbow Framework. For most cases, this included examples from many different evaluations and did not include a complete understanding of how all seven tasks were carried out in one evaluation. Third, I conducted an analysis of interview data to identify themes and issues that were not directly related with the seven tasks of evaluation but were related with the first and third research questions. Data that

were not relevant to the research questions were not analyzed in this study. Fourth, findings from each case were identified typically including about seven to ten findings related with the seven tasks of evaluation and five to ten additional findings for a total of between twelve and twenty findings for each case.

Cross case analysis was then conducted in three steps. First, I created a large data display of findings from all eight cases pertaining to the seven tasks of evaluation. I then developed cross case findings based on patterns of similarities and differences between and across cases. For each cross case finding, I identified the cases that supported the finding and the cases that did not support or challenged the finding. All cross case findings discussed in this paper are supported by at least five cases. Second, for each cross case finding, I identified two or three examples from cases that provide illustrations of the findings. This involved moving between the cross case finding, case findings, and interview data as well as case materials. Researcher judgment determined selection of examples and was guided by the purposes of selecting examples that illustrated the finding and that were detail rich. Third, I identified cross case findings related with the first and third research questions and that were present in findings from at least five cases.

Case Descriptions

For ease of identification of these cases of evaluation practice, I have labeled each with the name of the practitioner. Below, I briefly describe each case in terms of what area(s) of practice the evaluator works in; which systems and complexity thinking and techniques he/she typically draws on; and specific instances of evaluation practice which were reviewed in this study. These limited descriptions do not capture the variety or richness of each practitioner's evaluation practice and exclude other kinds of work (e.g., teaching, research, consulting) that

practitioners' carry out. The descriptions simply highlight a few main features of each practitioner's evaluation practice identified in the interviews and materials reviewed.

Glenda Eoyang. Eoyang's practice uses Human Systems Dynamics (HSD), an approach she developed that integrates a complex adaptive systems worldview with an explanatory and action oriented inquiry process, to evaluate and inform learning and action in a broad range of organizations and professional practices (e.g., business, education, healthcare). Two main HSD ideas, containers-differences-exchanges (CDE) (also referred to as pattern logic) and adaptive action, inform the evaluations she conducts (Interview). Pattern logic is a way to identify patterns of self-organization in complex adaptive systems, and adaptive action is an iterative inquiry process to inform learning and action that is guided by three questions: What? So What? Now What? Eoyang has written about evaluating performance in complex adaptive systems (Eoyang & Berkas, 1999) and using adaptive action in organizations (Eoyang, 2013). I reviewed two instances of Eoyang's practice: an HSD evaluation of a social service department (Eoyang, 2007) and a dynamic evaluation of healthcare communities of practice (Eoyang & Yellowthunder, 2007). During the interview, Eoyang spoke in depth about the latter instance.

Margaret Hargreaves. Hargreaves' practice includes the use of systems-informed approaches, including systems change (Hargreaves, 2010), rapid evaluation (Hargreaves, 2014), and developmental evaluation (Patton, 2011), to evaluate policies, programs, and initiatives in the areas of social welfare, public health, and community development. She frequently draws on systems and complexity concepts including system dynamics, boundaries, relationships, perspectives, and ecological levels (Hargreaves, Cole, Coffee-Borden, Paulsell, and Boller, 2013) and uses techniques such as outcome mapping, social network analysis, and systems mapping in the evaluations she conducts (see Hargreaves, 2010, p. 14 for fuller list). I reviewed

59

one instance of Hargreaves' practice: an evaluation of a home visiting system (Hagreaves et. al, 2013). In the interview, Hargreaves discussed three additional instances: evaluation of Building Healthy Communities commissioned by the California Endowment; evaluation of Change in Mind initiative commissioned by the Alliance for Strong Families and Communities; and Models for Change: The Legacy Initiative funded by the Macarthur Foundation.

Richard Hummelbrunner. Hummelbrunner's practice primarily includes evaluations of regional development supported by European Union structural funds. He has extensive training in systems thinking and techniques and has been influential in translating these for use in evaluations (e.g., Hummelbrunner, 2011; Williams & Hummelbrunner, 2009). He frequently uses the core systems concepts of interrelationships, perspectives, and boundaries in his evaluations and draws on a variety of techniques from the systems fields including circular dialogue, balanced scorecard, process monitoring of impacts, and causal loop diagrams (Hummelbrunner, 2007; Hummelbrunner & Reynolds, 2013; Hummelbrunner, 2015). I reviewed two accounts of Hummelbrunner's practice using systems thinking and techniques[7] (Hummelbrunner, 2007; Hummelbrunner, 2011). During the interview, Hummelbrunner discussed several instances of his practice.

Douglas Luke. Luke's practice involves evaluations of policies and organizations in public health, particularly in tobacco control, and primarily funded by the Centers for Disease Control and Prevention (CDC) and National Institute for Health (NIH). He typically uses social network analysis in the evaluations he conducts and has written more broadly on the use of systems science methods (i.e., system dynamics, social network analysis, and agent-based modeling) in public health (Luke and Stamatakis, 2012); the use of network analysis in evaluations of tobacco control policies (NCI, 2007); and development of an evaluation

---

[7] Evaluation reports and case examples from his practice are written in German and, therefore, not reviewed here.

framework for policy-relevant agent-based models (Wallace & Ogawa, 2015). I reviewed one instance of Luke's practice: a case study of network analysis in tobacco control (NCI, 2007). In the interview, Luke discussed several additional instances.

Gerald Midgley. Midgley's practice draws on his expertise in systems thinking and techniques (Midgley, 2000; Midgley, 2003) and, typically, uses boundary critique, stakeholder participation, and theoretical and methodological pluralism (Boyd et. al, 2007). Techniques he uses include viable systems modeling, soft systems methodology, critical systems heuristics, and interactive planning (Midgley, 2000). He typically purposefully combines aspects of multiple t methods to compensate for their different strengths and weaknesses, adapts techniques, and, sometimes, develops new techniques for particular circumstances. I reviewed three instances of Midgley's practice: evaluation of services for people with disabilities (Midgley, 1996); evaluation of diversion from custody of mentally disordered offenders (Midgley, 2000); and evaluation capacity building for community health services (Boyd et. al, 2007). Midgley discussed these and several additional instances from his practice during the interview.

Robin Lin Miller. Miller's practice includes experience as an internal evaluator at the Gay Men's Health Crisis in New York City and, more recently, evaluations of community-based programs and practices in the areas of sexual health and HIV/AIDS prevention and care. She routinely uses core systems concepts (i.e., relationships, perspectives, boundaries) in the evaluations she conducts and has extensive training in and experience using system dynamics modeling. I reviewed an article discussing her use of system dynamics modeling to examine contextual factors and implementation challenges influencing client recruitment and retention in evidence-based, HIV prevention programs (Miller, Levine, McNall, Khamarko, and Valenti, 2009). This uniquely is not an instance of Miller's evaluation practice, but rather a research

application of system dynamics that drew on data collected during prior evaluations she conducted. I included this example because some applications of systems science modeling techniques, particularly in public health, are first carried out as research studies and then later integrated into evaluation practices. During the interview, Miller spoke about how this study and her other applications of system dynamics modeling have informed her evaluation practice and about specific instances of her evaluation practice.

Michael Quinn Patton. Patton's practice centers on evaluations of social innovations across a variety of areas (e.g., human rights, youth homelessness, international development) using an approach he founded, *Developmental Evaluation* (DE) (Patton, 2011), which explicitly incorporates systems thinking and complexity concepts (e.g., non-linearity, emergence, adaptation, uncertainty, dynamical systems). More recently, he has developed and begun using two spin-offs of DE, Principles-Based Developmental Evaluation and Blue Marble Evaluation. I reviewed one instance of Patton's practice: an evaluation of an inter-agency youth homelessness initiative funded by the Otto Bremer Foundation, which he conducted with Nora Murphy (Homeless Youth Collaborative on DE, 2014; Murphy, 2016). In the interview, Patton discussed this instance in depth.

Deborah Wasserman. Wasserman's practice primarily includes evaluations of human service programs and organizations. She developed a systems orientation to evaluation based on integrating program theory, self-determination theory, and systems thinking as conceptualized by Cabrera, Colosi, and Lobdell (2008) (i.e., distinctions, systems, relationships, perspectives) (Wasserman, 2008, 2010). I reviewed three instances of Wasserman's practice: two evaluations described in Wasserman (2010) of a comprehensive out-of-school program and child mental

health service coordination initiative and an article discussing program evaluations of Africentric rites of passage youth programming (Wasserman & Emery, 2010).

My goal in analyzing cases was to identify how the use of systems and complexity ideas and techniques was influencing the ways *these* evaluators thought about the role and practice of evaluation and how they designed and conducted evaluations in the *particular* instances of evaluation practice I reviewed. I focused on only a few instances of each evaluator's practice. During interviews, some evaluators spoke in depth about these instances and some discussed additional instances for which I did not review secondary materials. Each evaluator's practice is unique and variable, as are the instances of practice I reviewed. I am not claiming that the collection of cases I reviewed are representative of cases of all evaluation practice or that the instances of a particular evaluation practice I reviewed are representative of a particular evaluator's practice. Findings are not generalizable to other cases or instances, although they may provide insights to understanding the broad idea of what borrowing from one field to another entails as well as considerations for borrowing in evaluation practice.

Findings

Through conversations with these practitioners, I came to realize how evaluators' thinking about their own practice and evaluation practice more generally continually evolves, as does how they design and conduct evaluations. I first discuss findings related with how evaluators conceived of and talked about the nature of evaluation practice within which they are borrowing systems and complexity ideas and techniques. Then, I turn to findings related, more specifically, with how evaluators draw on systems and complexity ideas and techniques when they design and conduct evaluations in particular circumstances.

Nature of Evaluation Practice

Each evaluator that I interviewed had her or his own understanding of what evaluation practice means. When they discussed borrowing ideas and techniques from systems and complexity fields they thus did so with this understanding of the practice in mind. What this means is that what they borrowed and how they viewed the import of that borrowing for evaluation was always situated in their particular understandings of evaluation practice. In this section, I discuss findings regarding evaluators' different understandings of evaluation practice and borrowing systems and complexity ideas and techniques within these understandings.

No "standard" evaluation practice. Each evaluator expressed a different idea about what "standard" (i.e., traditional, typical, prevalent) evaluation practice is and how his/her practice using systems and complexity thinking and techniques is different from this standard. Thus, dichotomous comparisons between "standard" evaluation practice and evaluations using systems and complexity thinking and techniques are not feasible or particularly useful, as there is substantial variation amongst evaluators and across evaluation practices. This may well be attributable, in part, to the different fields in which they were primarily trained (e.g., systems science, community psychology, public health); their non-evaluation, professional roles such as a complexity or systems practitioner, consultant, researcher, or professor; the kinds of institutions they work for (e.g., academic universities, consulting agencies, and research organization); the agencies that typically commission the evaluations they conduct (i.e., public government, private sector, and international development institutions); the problem areas and types of policies and programs they evaluate; and the national and international contexts in which they work.

Shared reasons for using systems and complexity thinking and techniques. The evaluators I interviewed expressed a sense of familiarity with and preference for these ideas and

techniques. Many were first exposed to ideas or methods from these fields during graduate school, by a professor's research or course they took, and often in disciplines other than evaluation such as family systems, sociology, or community psychology.

These evaluators also shared a belief in the potential value of systems and complexity thinking and techniques for addressing what they regarded as gaps and limitations in what they regarded as standard evaluation practices. One major gap, mentioned particularly by Patton, Hargreaves, Eoyang, and Hummelbruneer, is the considerable attention paid to pre-determined goals and activities of a program in evaluation with few ways to address emerging and unexpected processes and consequences. Another gap mentioned by Luke, Patton, and Hargreaves is the emphasis in many evaluations on assessing characteristics and changes in individuals, which, sometimes, may be aggregated to assess changes in groups; but, there are considerably fewer methods for assessing changes at more macro units of analysis such as systems, networks, and organizations.

Evaluators mentioned several limitations of evaluation practice, which they felt systems and complexity thinking and techniques were particularly valuable for addressing. These include the emphases on assessing results, outcomes, and impacts over implementation processes and, subsequently the use of methods for causal attribution rather than exploring complex causal relationships. Evaluators have selected and used techniques specifically for addressing this limitation, including Miller's use of system dynamics and Eoyang's use of her HSD CDE framework. Another limitation evaluators discussed is a tendency to privilege the views and values of those commissioning an evaluation rather than the perspectives, interests, and values of those involved (e.g., staff members, participants) and those affected (e.g., families, community

members). Evaluators have also used techniques to address this limitation, such as Midgley's use of boundary critique and Hummelbrunner's use of dialogic techniques.

Lack of agreement on language. The evaluators I interviewed did not share a common language for describing evaluation practice nor did they seem to agree on meanings of specific terms that they borrowed from systems and complexity fields. They fully acknowledged this lack of agreement. Varied meanings were given to *systems thinking* as thinking about systems, the use of systems concepts, particularly interrelationships, perspectives and boundaries and D-S-R-P, as well as the broad array of techniques from the systems fields; *complexity* as specific concepts, complex adaptive systems, and complexity science; *boundaries* as the material border between a system and its environment and a judgment about what is included or excluded in an evaluation or a specific model of a system; and *relationships* as how groups of people (e.g., intended beneficiaries, program leaders) relate to each other and to connections between multiple scales and units of analysis. Most evaluators emphasized that terms were being used too loosely within the broader evaluation field and that there is a need for better understanding of the theoretical basis for terms and their particular meanings within different systems and complexity traditions.

This finding is not surprising given extended and ongoing debates about terminology in the systems and complexity fields as well as, more recently, in the evaluation literature (e.g., Cabrera, 2008; Mowles, 2014). However, lack of a shared language poses a considerable challenge for continued borrowing and for learning from those who are using these ideas. If a practice is carried in part via a shared language then, using terms with different meanings, sometimes using meanings originating in the systems and complexity fields, sometimes using meanings adapted or invented by evaluators, and sometimes using everyday, lay meanings, means that a practice is not easily understood or communicated.

Multiple conceptions of systems- and complexity-informed evaluation practice.

Interviews revealed that these evaluators are taking up different ideas and techniques from the systems and complexity fields and modifying the way they practice as they go. They also expressed different, and sometimes conflicting, views on which ideas and methods are most important for evaluation practice. Variations in their approaches were evident in the following ways:

- Relative emphasis given to systems and complexity thinking versus techniques. Several evaluators suggested the importance of systems and complexity concepts, but that the techniques, methods, or approaches were not necessary. Others contended that the methods provided the deepest insights and learning and were essential.

- Degree of explicit attention paid to the incorporation of systems and complexity ideas in evaluation practice. For example, one evaluator remarked that systems thinking was integral to all of the evaluations he/she conducted. However, the use of systems thinking operated largely in the background, informing how he/she thought about the evaluand and planning the evaluation but there were few tangible examples of this influence when studying evaluations he/she conducted. Others talked about using systems thinking as an explicit process that occurred in the foreground of an evaluation; these ideas were evident throughout the evaluation report as well as the stories the evaluator told about the evaluation.

- Disagreement on which systems and complexity thinking or techniques they viewed as most relevant and significant for evaluation practice. For example, Eoyang conceives of human organizations as complex adaptive systems (CAS) and finds this understanding useful for drawing her attention to particular characteristics. Two other evaluators

expressed doubts about the relevance and value of complexity science for conducting evaluations.

- Disagreement on purposes and situations for which specific systems and complexity thinking or techniques should be used. For example, Patton distinguishes between simple, complicated, and complex situations saying developmental evaluation, and its use of systems and complexity concepts, is most appropriate for complex situations. Hargreaves focused on using systems thinking and techniques when evaluating interventions designed to influence systems change. Others viewed systems and complexity thinking and techniques as generally relevant across different evaluation purposes and situations and emphasized different ideas and methods depending on the particularities of a situation.

- Emphasis given to traditional social science methods versus systems and complexity methods. Some viewed traditional evaluation methods as the primary data collection and analysis techniques and others viewed systems and complexity methods as primary.

These instances of variation suggest the need to be cautious about presuming that evaluations labeled with some version of "systems" or "complexity" refer to common underlying ideas about evaluation practice.

No unifying theory of evaluation practice. Evaluators interviewed in this study drew on different evaluation theories or models for their practice. For example, Patton used his own developmental evaluation model (Patton, 2011), which was also used by Hargreaves; Hummelbrunner (2011) employed a set of ideas about systems thinking in regional development; Hargreaves (2010, 2014) used systems change and dynamic evaluation frameworks; and Luke (2015)'s evaluation framework primarily involved the use of social network analysis within an

adapted logic model. What this variation in theory of practice suggests is that systems and complexity is not another theory or model of evaluation that sits neatly alongside the several dozen existing evaluation models (e.g., improvement and accountability-oriented approaches, social agenda and advocacy approaches) (Stufflebeam & Shinkfield, 2007). Rather, it is a set of ideas and techniques that evaluators can draw on in a myriad of different ways to be relevant to and useful in the particular circumstances in which they practice.

Specific Aspects of Evaluation Practice

The findings discussed in this section are concerned with how evaluators apply and adapt particular systems and complexity ideas and techniques to inform how they carry out specific tasks, activities, or issues in the evaluations they design and conduct. I present several over-arching patterns that were evident across cases and instances of evaluation practice, along with several examples and variations.

Re-structuring the evaluation as itself an intervention. An ongoing consideration in evaluation practice regards whether and the extent to which an evaluation should influence the policy or program being evaluated and the extent to which an evaluator's role should be independent from versus involved in (or even an advocate for) particular changes in the evaluand and broader situation. Most of the practitioners that I interviewed clearly conceived of the evaluation as an intervention and the evaluator's role as an agent that intentionally influences the intervention.

Hummelbrunner explicitly structures evaluations as systems and as interventions. That involves recognizing the mutually influential relationships between an evaluation as a system, a client system (i.e., funders, managers, participants, public), and an evaluator system (i.e. experts conducting the evaluation). Structuring the evaluation system as an intervention means "applying

69

external influence upon a system with an aim of producing change" while acknowledging that "social systems are self-determined" such that "every system decides on its own – and according to its own logic" (Hummelbrunner, 2007, p. 163). In other words, evaluation is not an external intervention trying to change what is being evaluated. Rather, evaluation is a "regulatory mechanism" of the evaluand that provides information and facilitates reflection and discussion about the present state of a system or situation and what should stay the same and what should be changed. In my understanding, re-structuring an evaluation as an intervention has three major influences on the evaluations Hummelbrunner conducts. First, he ensures that an evaluation communicates in ways that are aligned with the language, rules and behavior patterns, predominant issues of concern, and values of those involved in an intervention (Hummelbrunner, 2007). Second, he prioritizes high levels of stakeholder involvement throughout an evaluation and, more importantly, he views stakeholder differences as a "resource rather than an obstacle" (p. 169); this is necessary to foster constructive dialogue to inform decision-making and changes in an intervention. Third, he positions himself as an evaluator as an "external observer" who has additional points of view and skills that can help manage the change processes of an intervention and, also, as someone who intentionally influences the intervention (Hummelbrunner, 2011, p. 271). In his words, an evaluator should "actively based on systemic principles, collect information and feed it back in varied (often surprising) ways to trigger reactions within the evaluated system in order to find solutions or develop new patterns of an interaction." (Hummelbrunner, 2011, p, 271).

Patton, somewhat less explicitly, also re-structured an evaluation as an intervention in the developmental evaluation he conducted with Nora Murphy of an inter-agency youth homelessness initiative. This initiative was comprised of a reflective practice group with leaders

from the funding agency, the Otto Bremer Foundation, and six grantee agencies (three emergency shelters, two youth opportunity drop-in centers, and one street outreach organization). The initiative broadly aimed to build collaboration across these agencies and improve outcomes for homeless youth. Patton and Murphy used a principles-based developmental evaluation approach to inform direction and ongoing development of this collaborative initiative. A significant component of the evaluation involved working with leaders to identify principles of practice that were shared across agencies, which Murphy then empirically verified by conducting fourteen case studies with youth who were successfully served by these agencies (Murphy, 2016). These principles then became the intervention, which leaders used as a framework to align practices across agencies (e.g., hiring practices, staff training) and to guide future inter-agency collaborations. Moreover, according to Patton, these principles have received international attention and have influenced a statewide task force on homelessness in Minnesota. Murphy (2016) describes their role as evaluators as "instruments" in the change process (p. 80). Rather than taking a distant or objective role, they expected that their relationship "would change the process and influence the outcomes" (Murphy, 2016, p. 80).

Patton and Murphy may contend that the evaluation was not an intervention, but merely supported the continual development of the intervention. But, in my view, the evaluation was designed to influence change in the intervention it was evaluating and, indeed, influenced changes and, therefore, illustrates re-structuring evaluation itself as an intervention.

Examples from other evaluators' practices suggest variations on the idea of evaluation as an intervention. Several evaluators explicitly incorporated a design component into an evaluation – a phase in which evaluators worked with stakeholders, sometimes using systems design techniques, to develop plans for changing an intervention. For example, in an evaluation Midgley

conducted of a program to divert individuals with mental health problems from custody in the criminal justice system, he used features of soft systems methodology in the initial phase of the evaluation to build agreement around the intervention's objectives and operational plans; then, in the final phase of the evaluation, he used aspects of critical systems heuristics and interactive planning to facilitate strategy and decision-making about how to use the evaluation findings to inform changes to the intervention (Midgley, 2000). When conducting evaluations using social network analysis, Luke frequently structures the analysis in three phases: discovery, diagnosis, and design (NCI, 2007). The design phase focuses on "helping the community collectively identify strategies to design (tune) a network to accomplish its goals more effectively" (NCI, 2007, p. 179).

Another variation involved structuring meetings for stakeholders, sometimes with evaluators as facilitators, to review evaluation data and discuss what findings mean for the intervention and any potential changes to be made. For example, in an evaluation of the Change in Mind Initiative, Hargreaves used a gallery walk format to post theories of change from different sites and a six-hour process during which grantees reviewed and discussed each other's strategies.

The use of systems and complexity thinking and techniques may mean that the line between an evaluation and the intervention being evaluated is rather ambiguous and flexible. Structuring evaluations as interventions may provide unique opportunities for using systems and complexity ideas and techniques, as many were developed for purposes of problem analysis and intervention design. This finding also supports a conception of the role of evaluators as closer to change agents than neutral observers.

Re-defining the object of evaluation. Typically evaluation commissioners define the evaluand and evaluators work within and around this definition. In many of the evaluations reviewed, evaluators raised significant questions about the way an evaluand and its context were defined and drew on systems and complexity concepts to redefine the object of evaluation.

In her evaluation of the California Endowment Building Healthy Communities, Hargreaves illustrates how systems concepts can be used to create a "lens through which you see the situation" (Interview). The evaluation was of an initiative to embed a health focus into all policies in the state of California within the broader agenda of increasing health equity, reducing health disparities and addressing social determinants of health. Hargreaves and Foster-Fishman were invited to conduct a retrospective systems change evaluation of the initiative. In primarily a qualitative, historical study, the evaluation examined the initiative using the idea of systems leverage from Donella Meadows, which states, "the effectiveness of a strategy depends on the level of systems leverage it has" (Interview). Three dimensions of systems leverage – smart strategy, population reach, and paradigm shifts – informed both how they defined and examined the initiative. They looked at the initiative as a five year process and examined the origins of the underlying concept of health and all policy, what conditions led up to the Endowment's theory of change, and then examined the process that occurred that led to the successful policy change from identifying early adopters to creating enabling conditions. When I asked Hargreaves how what they did would have been different had they not used systems concepts, she emphasized that they could have taken a traditional policy change perspective narrowing the examination of the initiative and treating it as a "campaign to design a policy, advocate for the policy, and get the policy passed" (Interview). The systems-informed lens they used allowed for a much broader

and deeper examination of "the conditions in which the state was able to create the task force" which later went on to design, advocate for, and get the policy passed (Interview).

Eoyang draws on her understanding of the characteristics of complex adaptive systems (CAS) to frame how she defines and examines the human organizations she evaluates. She distinguishes a CAS by several characteristic behaviors: dynamic, massively entangled, scale independent, transformative, and emergent (Eoyang & Berkas, 1998). She developed a framework for identifying, analyzing, and influencing self-organizing change processes of complex adaptive systems based on three ideas: containers, differences, and exchanges (CDE). Eoyang and her colleague used these three ideas, within a broader HSD-informed dynamic evaluation approach, to re-define an inter-professional education and practice for students and professionals in clinical and administrative health care settings (Eoyang and Yellowthunder, 2007). Rather than define the object of evaluation as the community of practice meetings that occurred at different sites and levels of the organization, the evaluators spoke with stakeholders to identify relevant containers (or units of analysis). They then engaged stakeholders in collecting data about what was occurring within these containers, particularly what differences were present and emerging and what exchanges were happening or not happening. This informed what they call a CDE portrait of the self-organizing patterns occurring within the initiative and broader context. They developed multiple CDE portraits at different points in time to examine how these containers, differences, and exchanges were changing or staying the same. Alongside their evaluation, other evaluators conducted a more traditional evaluation focused on observing pre-determined activities (i.e. community of practice meetings) and assessing pre-determined outcomes. Eoyang recalled this latter evaluation being pretty useless to the initiative's leaders, as the initiative had changed significantly over its three years and in different ways at different

sites. Contrarily, the complexity-informed evaluation captured dynamics as they emerged and was useful both for describing what was actually occurring and for identifying opportunities for leaders and staff to modify or capitalize on how the initiative was unfolding.

Other evaluators discussed more modest ways in which they redefined the object of evaluation, for example, by extending the boundaries of an evaluation object beyond those used by evaluation commissioners, designers, or managers. Miller provided an example of this. In an evaluation of a program for people living with HIV/AIDS, a commissioner requested information about knowledge related with sexual health and risky behaviors, issues directly related with the program objectives; after advocating for broader boundaries, she found that victimization was a much more significant issue for the people served by this program. Midgley mentioned numerous examples of re-defining the boundaries of the object of evaluation. One instance was in an evaluation of housing services for older people in which the commissioner wanted to know how currently collected assessment data could be used to improve services. In talking with beneficiaries of these services, Midgley discovered that any suggestions that were made by older people for services that were not currently provided by the agency were not recorded in this assessment data. So while he was initially tasked with using the assessment data to evaluate current services, he had to negotiate to re-define the object of evaluation to focus on ideal or desired services requested by beneficiaries in addition to current services.

Systems and complexity thinking and techniques may challenge evaluators to question and even redefine initial descriptions and ways of framing the policies, programs, and initiatives they are commissioned to evaluate. Such efforts may be particularly useful for commissioners and managers looking to understand their initiatives in new ways and to make continuous improvements and for capturing unexpected and emerging processes and outcomes.

Re-figuring relationships with evaluation commissioners. The relationship between evaluation commissioners and evaluators often is characterized in client-service terms. Commissioners are the clients who have questions about particular policies, programs, or projects, and they hire evaluators to design and conduct an evaluation that answers their questions. Of course, there are other types of commissioner-evaluator relationships but this has been a long-standing model. Evaluators in this study re-figured the relationship with commissioners to a partnership model that involved shared question-setting, negotiated and flexible contracts, commissioner involvement in decisions about the evaluation, high levels of commissioner (and stakeholder) participation in the evaluative inquiry, and learning on the part of both commissioners and evaluators.

The principles-based developmental evaluation conducted by Patton and Murphy highlights one way in which evaluators re-figured the relationship with the evaluation commissioner. Whereas, commissioner-evaluator relationships can often be guided by pre-determined objectives and a clear, specified work plan, evaluators viewed commissioners, and initiative leaders, as social innovators which called for an adaptive, collaborative, and learning-oriented relationship:

> Social innovators working on systems change have to navigate complexity, deal with turbulence and uncertainty, adapt rapidly to changing conditions, be vigilant about what is emerging as they engage, and watch for nonlinear interactions and consequences (Murphy, 2016, p. 63).

The group, including evaluators, members of the foundation staff, and directors of the grantee agencies, met monthly throughout the evaluation in a reflective practice group that collectively made decisions about the evaluation. Together, the group identified shared principles, developed a case study approach to examine the principles in practice, interpreted the case studies, revised the principles, discussed implications, and established ways to disseminate

76

and use the principles. Trust, open and ongoing communication, and shared decision-making were essential in this commissioner-evaluator relationship.

In a recently commissioned evaluation from the MacArthur Foundation of The Legacy Initiative, Hargreaves also discussed the importance of re-figuring the evaluator-commissioner relationship. In her evaluation work, traditionally, an organization sends out a Request for a Proposal with pre-set questions for the evaluation; evaluators submit proposals with relatively static evaluation designs for answering the questions; the commissioner selects the design they want to fund; and the evaluator implements this design. According to Hargreaves, the design often "doesn't really reflect the complexity of the situation or the context in which the evaluation's happening" and "doesn't really allow for engagement with the client about what are their priorities and how their perspectives can influence the evaluation design" (MH interview). Therefore, in this evaluation, the commissioner added on an initial planning period (also called an environmental scan period) during which the evaluators collect additional information, learn more about the particular dynamics of the situation, and get to know the interests of the commissioner as well as other stakeholders. The initial design may be modified or changed entirely to be more responsive to the context and information needs of stakeholders. Furthermore, this process provides time and space to build a more collaborative commissioner-evaluator relationship.

These examples highlight ways in which evaluators have re-figured their relationships with commissioners, and how commissioners have initiated and willingly engaged in re-figuring this relationship. Other evaluators also described small yet substantial ways they have modified their relationships with commissioners. Luke discussed several challenges of using social network analysis in evaluation, which are more easily addressed when evaluators and

commissioners work collaboratively as partners.[8] Hummelbrunner shared a story of a roundtable meeting of commissioners and evaluators, at a time when evaluation was rather new in Austria, in which commissioners were saying, in short, "We need better evaluations. Come on evaluators, do a better job" (Interview). In that meeting, Hummelbrunner spoke out saying, "If you want better evaluations, it has to be a collaborative effort. We have to work on this together. Better evaluations demand something from you and demand something from us…Let us establish a platform and exchange ideas and so on over a longer period of time" (Interview). This led into what they called "ongoing evaluation" which centered on a re-figured evaluator-commissioner relationship, more flexible contracting, and evaluation throughout the course of an intervention (Interview).

Incorporating systems and complexity thinking and techniques in evaluations may call for and even require different relationships between evaluators and commissioners. Such relationships cannot be fostered by evaluators alone, but rather require a process of trust building, open communication, shared commitment to some issue, questions, or situation to facilitate relationship building. Evaluator-commissioner partnership is not merely a means to evaluation use, but rather a necessity for collecting and analyzing data that are relevant, meaningful, and useful. This may require altering contracts and reporting requirements to better support such a re-figured relationship.

Re-purposing and expanding methods. Evaluators typically use an array of methods that, according to Mark, Henry, and Julnes (2000), fall into four inquiry modes each focused on different purposes: description-oriented methods for measuring characteristics of clients,

---

[8] One challenge has to do with the high response rates (90-100%) needed from survey participants to adequately map a social network; this requires commissioners' help, for example by providing staff assistance, to recruit and ensure participation in surveys. Another challenge has to do with interpreting a network analysis, which often requires genuine interest in the evaluation questions and input from commissioners and stakeholders throughout an evaluation

services, outcomes, and contexts (e.g., interviews, focus groups, observations, surveys, performance monitoring); classification techniques for developing categories and/or assigning cases to them (e.g., taxonomies, cluster analysis); causal analysis for identifying mechanisms and estimating effects (e.g., randomized experiments, case studies); and values inquiry for identifying value positions of stakeholders and the public (e.g., surveys, group deliberations). Evaluators in this study reported using methods for two new purposes, not covered in these inquiry modes, and using new systems and complexity methods for all four inquiry modes.

Evaluators, particularly Hummelbrunner and Midgley, reported using methods for surfacing, reflecting on, and deliberating about stakeholder differences. Many evaluations inadvertently reveal differences among stakeholders, including multiple and potentially conflicting perspectives, political interests, cultural norms, and values with regarding a particular situation, policy, or program. Hummelbrunner contends, "systemic evaluation should go beyond merely illustrating stakeholder different opinions (e.g., by visualizing diversity through rating and ranking) and effectively work with them" (Hummelbrunner, 2011, p. 262). One way to work with stakeholder differences is to use methods that make stakeholders' mental maps "explicit and visible" and that "help to build a more complete picture of reality by linking individual mental maps and working towards the emergence of collective mental maps" (Hummelbrunner, 2011, p. 262). Two methods he uses for this purpose include circular dialogue, a technique for facilitating structured conversations in which participants take on different roles and perspectives, and causal loop diagrams, a tool for visualizing complex relationships that interact through circular feedback processes that can be used to surface and discuss different views on and explanations of what is affecting a particular situation. Midgley similarly discussed using methods for working with stakeholder differences. For example, in an evaluation of a diversion from custody program

for mentally disordered offenders, Midgley recognized the lack of agreement within the multi-agency group, with members from criminal justice, mental health, social service, and academic agencies, regarding the objectives and plans for the program. Prior to conducting the evaluation, he carried out preliminary boundary critique to raise awareness of potential controversial issues and unspoken conflicts; team-building exercises with members of the initiative including facilitated discussions about differences in and commonly shared language; and aspects of soft systems methodology to develop a vision for the main activities of the initiative (Midgley, 2000).

Evaluators also reported using methods for creating designs (i.e., plans, models, strategies) to inform practitioners' decision-making and actions taken regarding a particular situation, policy or program. Midgley, in the evaluation of a diversion from custody program, incorporated a strategic planning phase of the evaluation that used aspects of critical systems heuristics and interactive planning during participatory workshops with service users, staff and management to design potential next steps for the program that drew on both the evaluation findings and their differing normative views on an ideal mental health and criminal justice system (Midgley, 2000). Midgley shared other examples of using these methods for design purposes, as well as other methods including viable systems modeling. Luke, when using social network analysis in an evaluation of a tobacco control network, has used a "3D – discovery, diagnosis, and design – model" (NCI, 2007, p. 177) to identify questions that network leaders, policy officials, and community members can use to guide their use of and learning from a specific network analysis. The final phase, design, involves "putting the findings to use" by identifying and enacting steps to "modify network structures and relationships and enhance the effectiveness of the network" (NCI, 2007, p. 179).

Additionally, evaluators in this study discussed using new systems and complexity techniques for the four inquiry modes identified earlier. One example of each is mentioned below:

- *Description*: Luke uses social network analysis for describing networks in tobacco control, which involves defining and identifying a network, collecting information about the network (e.g., funding relationships, frequency of contact, degree of cooperation, perceived importance of network partners), and analyzing patterns in relationships and networks.

- *Classification*: Patton, Hargreaves, and Eoyang mentioned using variations of the landscape diagram for classifying a particular situation, system, or aspects of a program on two axes. Variations include an HSD landscape diagram for classifying how close to or far from agreement and certainty a particular system dynamics are (Eoyang); a system dynamics landscape diagram for classifying high to low agreement in perspectives and high to low predictability in relationships (Hargreaves, 2010); and use of the uncertainty/conflict matrix and cynefin frameworks for distinguishing between simple, complicated, complex, and chaotic situations (Patton, 2011).

- *Causal analysis*: Miller describes system dynamics as a method that "builds on diverse sources of data and group process techniques to develop models that may be used to conduct virtual experiments via computer simulation" (Miller et. al, 2009, p. 3). She has used this method for identifying and analyzing the causal feedback processes and stocks and flows that may affect patterns of recruitment, retention, and graduation in community-based, small group HIV prevention programs (Miller et. al, 2009).

- *Values inquiry*: Midgley often uses some process of boundary critique, which is discussed in his *Systemic Intervention* book, in the initial phase of an evaluation to surface and question different boundary judgments regarding a program and to identify different relevant values as well as potential conflicts and issues: "in all sorts of projects whether they're evaluation or systemic design, is the need to take account of the boundary judgments people are making. The value judgments they make, processes of marginalization, how some stakeholders get demonized in the process, issues of conflict, and how these play out in either allowing or restricting…the emergence of ideas in the process" (Interview).

This finding suggests that borrowing systems and complexity methods in evaluations involves re-considering and potentially adapting the purposes for which methods are used. Moreover, some evaluators discussed the flexibility and creativity required in using some systems and complexity techniques as there are multiple purposes and adaptations for any given method and methods can be combined to compensate for strengths and weaknesses.

Revisiting and problematizing the making of value judgments. Evaluation is often defined as making value judgments using evaluation-specific methodology (e.g., criteria, performance standards, weighting and synthesis procedures). Evaluators that I interviewed found the idea of making value judgments problematic, struggled with making value judgments when using systems and complexity thinking and techniques, and only one discussed using evaluation-specific methodology.

A key challenge Eoyang raised with regards to her evaluation practice based on Human Systems Dynamics is that valuing is contextual and any judgment of value depends on the

purpose(s) or function(s) one values with regards to a particular system of interest. She

contended that her role as an evaluator does not involve making value judgments:

> That's the stance we took…describing what happened, telling what that says about the dynamics and how those might affect it, how that might inform practice… but without judging this is better, this is worse, this is right, this is wrong. There is no judgment in HSD except in relation to fitness (Interview).

"Fitness" for Eoyang means the quality and extent to which a system functions to achieve

particular purposes. In her words, "we value the ability of the system to inform it's own self-

organizing towards whatever end it might want to" (Interview). In the beginning of the

evaluation she conducted of inter-professional education and practice, the project leaders did not

want to focus on the university programs that students attended. They viewed this as separate

from the inter-professional practice and not necessary to include in the evaluation. About half-

way through the project, several project members informed Eoyang that there were major

differences between the curricula taught in these university programs, with one being highly

connected to practice, one being rather disconnected and irrelevant, and one being somewhat

connected. After noticing this, the university programs started working together to develop a

common curriculum. The CDE model helped identify this container, difference, and exchange,

but it was the stakeholders who decided the meaning and quality of this difference and what to

do about it: "Nobody really cared that the system was sub-optimized. Once that difference had

been surfaced and a container created for exchanges for it to be resolved, the system as a whole

is much more fit for function" (Interview). So, rather than making value judgments, Eoynag

identifies emerging CDE patterns and stakeholders determine how to judge the quality of these

patterns and what to do about them.

Wasserman finds valuing challenging when using systems thinking because of the lack of

empirical theories to inform how evaluators make judgments about the quality, health, or

longevity of a system. In her evaluation practice, she often is not in a position to place value on what she is seeing, but she is in a place to point out "alarms going off or flags going up in this area or that area," and she adamantly feels that the basis for these judgments should be transparent and, when possible, based on an empirical theory (Interview). In her area of specialty, family systems, there are concepts like family differentiation that indicate tolerance levels for intimacy and individuality, and these can be used to predict how well a family will deal with a crisis, for example. She uses self-determination theory to identify eight key qualities and relationships (i.e. pulse points) that an evaluation should assess in order to "provide feedback to a program system that supports self-determination" (Wasserman & Emery, 2013).

Luke explained three reasons why valuing, or judging the quality of a social network, is particularly challenging. First, in public health, there is a lack of theory and empirical evidence that explains what constitutes a "good" or "bad" network. Second, there is rarely just one network. Rather, there are multiple, overlapping networks and even in networks with the same members, there can be different kinds of network ties (e.g., friendship, trust). How boundaries are defined and which network ties are highlighted influences what defines a particular network. Third, defining the quality, or strengths and opportunities of a particular network, depends on what the evaluation questions are or the purposes for analyzing a particular network. In the evaluations he conducts, evaluation partners (i.e., commissioners, key stakeholders) inform the interpretation of a social network by what they want to know, what decisions they want to make, or what actions they want to take. For example, in an analysis of an African American childhood obesity research network, one of the aims of the commissioning organization was to identify central agencies that can be used to create new coalitions. Identifying these central agencies and their connections guided the analysis. In another example, Luke's colleague was analyzing the

emergency response networks in a Midwestern state and found that the east and west sides of the network were connected through only one person who was retiring in the coming year. This prompted examination of what new connections could be made.

Evaluation-specific methodology for making value judgments may be less relevant or appropriate for evaluators using systems and complexity thinking and techniques. Instead, some evaluators are turning to empirical or normative theories about how a system ought to function and others are leaving the valuing process up to stakeholders and stakeholder-led discussions. Or perhaps evaluation-specific methodology needs to be revisited and revised in light of systems and complexity ideas.

Renewing an emphasis on instrumental use. Some evaluators favor enlightenment use, arguing that instrumental use (i.e., direct, immediate use of evaluation findings in decision-making or changes) is less common and difficult in practice**.** Most evaluators that I spoke with emphasized instrumental use, specifically learning, with the aim of changing cognition and behavior of people involved in a program, situation, or intervention as well as changing policies, practices, and culture of an organization, inter-organizational collaboration, professional practice, or, even, governance in a region.

In an evaluation of the Change in Mind Initiative, Hargreaves described a high level of evaluator involvement, facilitating learning throughout an evaluation, and building evaluation capacity to collect and use data in an ongoing manner – all strengthening instrumental use of the evaluation. This developmental evaluation focuses on helping leaders "identify the target of what they want to change and document the evolution of their efforts to find allies, change the mindsets of their allies, create collective action with their allies, and shift other local services and institutions and governance structures and funding processes in support of these new brain

science concepts" (Interview). Evaluators are highly involved in the development of the initiative. Early on, they helped develop a systems theory of change and worked with the group on their message; then, they shifted to collecting information about the initiative and provided guidance in the initiative collecting analyzing data; and, next, they plan to develop a knowledge management framework that creates the evidence record of their activities and outcomes. Hargreaves remarked that this evaluation really "turns the tables on the grantees…they are engaged in their own evaluation using it for their own development and then sharing the results with us" (Interview).

Patton's developmental evaluation of the inter-agency homelessness initiative resulted in a series of direct, immediate changes in the way people thought about homelessness and the services for homeless youth as well as how the inter-agency professional practice is carried out. The evaluation's direct contributions were development of nine principles of practice (i.e., journey-oriented, trauma-informed, nonjudgmental, harm reduction, trusting youth-adult relationships, strengths-based, positive youth development, holistic, and collaboration) and a community of practice including members from all six agencies focused on learning and improving services for homeless youth (Murphy, 2016). Leaders developed the principles into an informational packet and distributed these throughout the agencies; they continue to influence hiring, organizational culture, and cross-organization collaboration (Murphy, 2016, p. 74). More indirectly, the evaluation helped call attention to the significance of building and managing relationships both for homeless youth and staff and community members working with these youth. Whereas the official theory of change and funding model focused on providing housing and ensuring youth remain in housing, the evaluation helped uncover the "informal system of support" in which relationships were essential. The principles helped capture what mattered to

youth, and to many staff members, about this informal system and to align the work of these six agencies. Furthermore, Patton used the principles at a statewide task force meeting around homelessness to encourage shifting thinking about homelessness and services for homeless people to focus on these principles.

These two examples, both of developmental evaluations, emphasize instrumental use. Other evaluators also shared stories of tangible and intangible ways that evaluations influenced the stakeholders' thinking and behavior regarding a particular issue, situation, or evaluand. Another renewed emphasis was on knowledge generation regarding specific social issues and types of interventions. For example, Miller's work has focused specifically on using system dynamics modeling to identify implementation challenges faced by particular kinds of evidence-based, HIV/AIDS treatment programs. She has used results from multiple evaluations to inform the development of system dynamics models and simulations of client recruitment and retention, and these models, in turn, inform the evaluations she conducts. Similarly, some of Luke's evaluations have used social network analysis to identify barriers to dissemination and implementation of health practices; over time, he and colleagues are developing a knowledge base that can inform future programming and evaluations.

Systems and complexity thinking and techniques may help evaluators conduct evaluations that are immediately and directly useful to commissioners and stakeholders. However, this instrumental use extends beyond using evaluation findings for decisions or changes that were identified prior to an evaluation. Rather, in these examples, evaluation processes and findings facilitate uncovering problem areas and opportunities for changes, which then may be taken up and addressed by commissioners and stakeholders. Additionally, this

instrumental use includes both the way people think about an issue or initiative and how they behave in relation to this issue or initiative.

Conclusion

This study drew on the practical experience of seasoned evaluators to examine ways in which their practice is being expanded through the influence of systems and complexity thinking and techniques. Based on an analysis of their views, it is apparent that, first, there is no single or predominant way to use systems and complexity thinking and techniques in evaluation practice or a "standard" evaluation practice within which these ideas and techniques are imported and applied. On the contrary, there is great variety in the concepts and techniques evaluators are using and the ways and evaluation circumstances in which they are used. Second, what we see is not necessarily a wholesale change in the way evaluation practice is traditionally conceived but rather significant rethinking of roles, responsibilities, and methods that includes restructuring the evaluation itself as an intervention; redefining the object of evaluation; re-figuring relationships with commissioners; repurposing methods; revisiting and problematizing the making of value judgments; and renewing an emphasis on instrumental use. These findings pose several implications for evaluator training, evaluation funding and commissioning, and research on evaluation.

Evaluator training. There are currently few training opportunities for evaluators interested in the many concepts, theories, and methodologies of the systems and complexity fields. Efforts to develop such training should include introductory-level material in professional development workshops, webinars, and guidebooks as well as more advanced instruction through textbooks, graduate courses and programs and summer institutes. The American Evaluation Association (AEA) and the European Evaluation Society (EES) currently offer

professional development workshops and webinars on topics related to systems and complexity. These provide a broad overview of systems and complexity ideas (e.g., systems thinking, complex adaptive systems) and introduce particular methodologies (e.g., causal loop diagrams). Summer institutes for practicing evaluators, such as those sponsored by foundations and federal agencies, could be developed in connection with existing training programs, such as the System Dynamics Institute at Massachusetts Institute of Technology, the Santa Fe Institute, the Human Systems Dynamics Institute, and the Open University's Systems Thinking in Practice Group.

There are several challenges to developing evaluator training that require further reflection and conversation within the evaluation field. Systems and complexity ideas and techniques are not merely tools or approaches that evaluators can readily apply, but rather are grounded in different ways of thinking about the world, social interventions, and scientific inquiry and different ways of being researchers and practitioners. The core assumptions and processes of these ways of thinking and being are still being identified and refined within the evaluation field, as are how these ways of thinking and being are distinct from and even challenge common ways of thinking and being among evaluators[9]. Training evaluators in a different way of thinking and being is not an easy or straightforward task, as it requires evaluators re-examining, challenging, and replacing long-held assumptions. Further consideration of pedagogical techniques and activities that can support such training are needed, such as Sweeney and Meadows' (2010) *Systems Thinking Playbook* and Cabrera and Cabrera's (2015) *Systems Thinking Made Simple*.

---

[9] For example, some evaluators make the distinction between systems and complexity thinking (Reynolds, Forss, Hummelbrunner, Marra, & Perrin, 2012), but there is little evidence of evaluators' using this distinction in the evaluation literature. Additionally, there are at least two, differing views on what constitutes systems thinking such as three core concepts (i.e., interrelationships, perspectives, boundaries (e.g., Williams, 2003; Williams, 2005; Williams & Imam, 2007) and a cognitive process (i.e., distinctions, systems, relationships, perspectives) (Cabrera et. al, 2008).

Another challenge has to do with who should be considered the experts in systems- and complexity-informed evaluation and what constitutes the basis for their expertise. At present, there are many evaluators claiming to use systems and complexity thinking in their evaluation practices. But, very few evaluators have studied these ideas in-depth or are versed in methodologies posing a risk that these ideas and techniques become misinterpreted, confused, or diluted. Yet, even evaluators who may be considered experts in these ideas and approaches, such as those interviewed in this study, vary widely in their interpretations and applications of these ideas and techniques. This raises the question of what should constitute the basis of expertise in systems- and complexity-informed evaluation. Identifying experts and bases for expertise is important as these are the evaluators who will presumably lead development of evaluation training materials as well as lead trainings.

Another related challenge involves envisioning what high quality systems- and complexity informed evaluation looks like or should look like and what are the best approaches to fostering such quality evaluation practice. Currently, there are many efforts to guide evaluators in when and how to use systems and complexity ideas and techniques in practice in the form of evaluation guides, evaluation approaches, and, most recently, principles for implementing specific evaluation approaches. But some of these efforts presume one or several standard ways of practicing evaluation. The varieties of ways evaluators use systems and complexity ideas and techniques, as found in this study, suggest that there is not one or several standard ways. Evaluation training may be better off promoting creative experimentation and trial-and-error. However, this poses a challenge for discerning and improving the quality of systems- and complexity-informed evaluation. One way to begin addressing this challenge could involve producing guides that carefully explain the strengths and weaknesses of different

concepts and methodologies and the variety of circumstances in which they may be useful to inform evaluators' decision-making. These guides can be informed by the extensive systems literature on theoretical and methodological pluralism.

Evaluation commissioning and contracting. Evaluators are constrained in their use of systems and complexity and ideas and techniques by the evaluation circumstances in which they work and by those who commission the evaluations they conduct. To move towards more fully taking up some of these ideas and techniques, there may need to be changes to the way evaluations are commissioned and contracted. One major consideration involves the kinds of social problems or issues commissioning agencies aim to address and the nature of the social change process the agencies aspire to facilitate to address such problems or issues. There are several trends in the types of initiatives for which evaluations are being commissioned that are particularly suited to the use of systems and complexity ideas and techniques. For example, initiatives designed to address wicked social issues (i.e., persistent, difficult, dynamic, inter-disciplinary) are not based on static definitions of the problem, pre-determined objectives or desired results, or even a model for the change process; such initiatives call for developmental evaluations. Additional examples, particularly evident in public health, include evaluations of evidence-based practices that examine contextual factors influencing implementation that benefit from systems methodologies and initiatives designed as networks, that benefit from social network analysis. Beyond individual initiatives, evaluations commissioned to examine multiple initiatives or to build the capacity of an agency for ongoing problem solving, innovation, or professional practice offer additional opportunities for using systems and complexity ideas and techniques.

There may also need to be new ways of constructing an evaluation contract that allow for flexibility and adaptability in evaluation design. Many evaluation contracts are established through a Request-for-Proposal process in which an evaluator submits an evaluation design, which becomes the basis for the evaluation if awarded. This may need to be changed to incorporate an initial period in which evaluators better understand the circumstances surrounding a project and evaluation and the information needs and perspectives of stakeholders. Such a shift poses a considerable challenge to contracting, as typically evaluators are selected based on the designs they submit. Embracing flexibility in proposed designs raises the question of what other considerations commissioners should make when selecting an evaluator, evaluation team, or evaluation proposal. Additionally, there is a need for more flexibility and adaptability in the evaluation design in order to make adjustments to changes in the circumstances and changes in knowledge. Such shifts in evaluation design pose a challenge for evaluators' accountability to commissioners. Whereas typically evaluators are held accountable largely to pursuing pre-determined evaluation questions and carrying out the evaluation plan as proposed, acknowledging flexibility and adaptability in designs may require higher levels of trust and closer relationships between evaluators an commissioners. Planning for such a flexible and adaptive design also creates numerous financial and logistical challenges, as evaluators need to budget and allocate resources in new ways as well as potentially coordinate data collection, analysis, and reporting in shorter, more iterative cycles.

Research on evaluation. Examining systems and complexity ideas and techniques in connection with evaluation is an emerging area of research with numerous potential new directions. Thus far, research has focused largely on conceptual papers examining the relevance of a particular theory or idea to the evaluation field and case-based methodological papers that

describe the application of a particular technique to a particular evaluation. Additionally, there have been numerous books and several literature reviews that more broadly examine connections between the systems and complexity fields and the evaluation field. One area in need of research, building on this study, centers on examination of how evaluators select, translate, and use specific ideas and techniques in particular evaluation circumstances as a means to begin developing an evidence base for systems- and complexity-informed evaluation practice. Similarly, there is a need for empirical research examining how evaluation commissioners and stakeholders view and experience the use of systems and complexity thinking and techniques in an evaluation. Another area needed is for exploratory research that identifies overlooked areas of the systems and complexity fields that may be relevant to the practice of evaluation, such as second-order cybernetics and systemic design. More specifically, this research could focus on issues or challenges that many evaluators face and explore how systems and complexity practitioners deal with this challenge. A final area of needed research focuses more specifically on refining and advancing the use of systems methodologies by describing and evaluating the use and added value of specific methodologies.

While the systems and complexity fields hold much promise for the field of evaluation, the implications for evaluation practice are only beginning to emerge. As more evaluators and evaluation commissioners take up systems and complexity ideas and techniques, further research should continue to examine their implications. Such examinations are essential to revealing how the evaluation field is continually developing and to setting out future directions so that evaluation practice can continue to be relevant and useful to commissioners amidst changing circumstances.

REFERENCES

Alkin, M.C. & Christie, C.A. (2005). Unraveling Theorists' Evaluation Reality. *New Directions for Evaluation*, 106, 111-128.

Barnes, M., Matka, E., & Sullivan, H. (2003). Evidence, Understanding and Complexity: Evaluation in Non-Linear Systems. *Evaluation, 9*(3), 265-284.

Better Evaluation. (2014). Retrieved from: http://betterevaluation.org

Boyd, A., Geerling, T., Gregory, W. J., Kagan, C., Midgley, G., Murray, P., & Walsh, M. P. (2007). Systemic evaluation: a participative, multi-method approach. *Journal of the Operational Research Society, 58*(10), 1306-1320.

Britan, G. M. (1978). The Place of Anthropology in Program Evaluation *Anthropological Quarterly, 51*(2), 119-128.

Butler, M.O. & Copeland-Carson, J. (Eds.) (2005). *Creating Evaluation Anthropology: Introducing an Emerging Subfield*. Wiley-Blackwell.

Byrne, D. (2013). Evaluating complex social interventions in a complex world. *Evaluation, 19*(3), 217-228.

Cabrera, D. & Cabrera, L. (2015). *Systems Thinking Made Simple New Hope for Solving Wicked Problems*. Ithaca, NY: Odyssean.

Cabrera, D. (2006). Systems Thinking (Doctoral dissertation). Retrieved from https://ecommons.cornell.edu/bitstream/handle/1813/2860/DerekCabreraDissertation.pdf?sequence=1.

Cabrera, D., Colosi, L., & Lobdell, C. (2008). Systems thinking. *Evaluation and Program Planning, 31*(3), 299-310.

Cabrera, D., & Trochim, W. M. (2006). *A Theory of Systems Evaluation*. Ithaca, NY.

Cabrera, D., & Trochim, W. M. (2006). *A Protocol of Systems Evaluation*. Ithaca, NY.

Callaghan, G. (2008). Evaluation and negotiated order: developing the application of complexity theory. *Evaluation, 14*(4), 399-411.

Campbell, D. (2011). Assessing the impact of planned social change. *Journal of Multidisciplinary Evaluation*, 7(15), 3-43.

Chambon, A. (2012). Disciplinary Borders and Borrowings: Social Work Knowledge and its Social Reach, a Historical Perspective. *Social Work and Society, 10(2), 1-12*

Chelimsky, E. (1998). The role of experience in formulating theories of evaluation practice. *American Journal of Evaluation*, 19(1), 35-55.

Cronbach, L. J. (1980). *Toward Reform of Program Evaluation*. San Francisco, CA: Jossey-Bass, Inc.

Datta, L.E. (2008). Response to paper "Systems thinking" by D. Cabrera et al.: systems thinking: an evaluation practitioner's perspective. *Evaluation and Program Planning, 31*(3), 321-322.

Durland, M. M., & Fredericks, K. (2005). An introduction to social network analysis *New Directions for Evaluation*(107), 5-13.

Dyehouse, M., Bennett, D., Harbor, J., Childress, A., & Dark, M. (2009). A comparison of linear and systems thinking approaches for program evaluation illustrated using the Indiana Interdisciplinary GK-12. *Evaluation and Program Planning, 32*(3), 187-196.

Eoyang, G. & Yellowthunder, L. (2007). Dynamic Evaluation Report of Interprofessional Education and Practice. Creating an Interprofessional Learning Environment through Communities of Practice: An Alternative to Traditional Preceptorship. Retrieved from:

Eoyang, G. H. (2007). Human Systems Dynamics: Complexity-based Approach to a Complex Evalaution. In B. I. I. Williams (Ed.), (pp. 123-140). Point Reyes, CA: EdgePress of Inverness.

Eoyang, G. H., & Berkas, T. H. (1999). Evaluating Performance in a Complex Adaptive System. In M. Lissack & H. Gunz (Eds.), (pp. 1-21). Westport, Connecticut: Quorum Books.

FASID. (2010). In N. Fujita (Ed.), *Beyond logframe; using systems concepts in evaluation*. Japan: Foundation for Advanced Studies on International Development.

Fetterman, D. (1988). *Qualitative Approaches to Evaluation in Education: The Silent Scientific Revolution*. New York: Praeger Publications.

Fitzpatrick, J., Christie, C., and Mark, M. (Eds.) (2009). *Evaluation in Action*. Thousand Oaks, CA: SAGE Publications, Ltd.

Fredericks, K. a., Deegan, M., & Carman, J. G. (2008). Using System Dynamics as an Evaluation Tool: Experience From a Demonstration Program. *American Journal of Evaluation, 29*(3), 251-267.

Forss, K., Marra, M., and Schwartz, R. (Eds.) (2011). Evaluating the Complex: Attribution, Contribution, and Beyond (2011). (Vol. 18). London, U.K.: Transaction Publishers

Foster-Fishman, P. G., Nowell, B., & Yang, H. (2007). Putting the system back into systems change: a framework for understanding and changing organizational and community systems. *American journal of community psychology, 39*(3-4), 197-215.

Gample, J.A. (2008). A developmental evaluation primer. The J.W. McConnell Family Foundation. Retrieved from: http://www.mcconnellfoundation.ca/de/resources/publication/a-developmental-evaluation-primer

Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, and Trow, M. (1994). *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies*. London, UK: SAGE Publications Ltd.

Gopalkrishanan, S., Preskill, H., & Lu, S. (2013). Next Generation Evaluation: Embracing Complexity, Connectivity, and Change: FSG.

Gregory, A. J., & Jackson, M. C. (1992a). Evaluation methodologies: a system for use. *Journal of the Operational Research Society, 43*(1),19-28.

Gregory, A., & Jackson, M. (1992b). Evaluating organizations: a systems and contingency approach. *Systems Practice, 5*(1), 37-60.

Hargreaves, M., Cole, R., Coffee-Borden, B., Paulsell, D., & Boller, K. (2013). Evaluating Infrastructure Development in Complex Home Visiting Systems. *American Journal of Evaluation, 34*(2), 147-169. doi: 10.1177/1098214012469271

Hargreaves, M. B. (2010). Evaluating System Change: A Planning Guide (pp. 1-20). Princeton New Jersey.

Hargreaves, M. B., & Paulsell, D. (2009). Evaluating Systems Change Efforts to Support Evidence-based Home Visiting: Concepts and Methods. Princeton New Jersey.

Hargreaves, M., & Podems, D. (2012). Advancing systems thinking in evaluation: a review of four publications. *American Journal of Evaluation, 33*(3), 462-470.

Henning, P.B. & Chen, W. (2012). Systems Thinking: Common Ground or Untapped Territory? *Systems Research and Behavioral Science. 29, 470-483.*

Hieronymi, A. (2013). Understanding Systems Science: A Visual and Integrative Approach. *Systems Research and Behavioral Science, 30,* 580-595.

Hirsch, G. B., Levine, R., & Miller, R. L. (2007). Using system dynamics modeling to understand the impact of social change initiatives. *American journal of community psychology, 39*(3-4), 239-253.

Hummelbrunner, R. (2008). Response to paper "systems thinking" by D. Cabrera et al.: a tool for implementing DSRP in programme evaluation. *Evaluation and Program Planning, 31*(3), 331-333.

Hummelbrunner, R. (2010). Beyond logframe: critique, variations, and alternatives. In N. Fujita (Ed.), *Beyond logframe; using systems concepts in evaluation* (pp. 1-33). Japan: Foundation for Advanced Studies on International Development.

Hummelbrunner, R. (2011). Systems thinking and evaluation. *Evaluation,* 17(4) 395–403

Hummelbrunner, R., & Reynolds, M. (2013). Systems thinking, learning and values in evaluation (pp. 9-10): European Evaluation Society.

Hummelbrunner, R. (2015). Learning, Systems Concepts, and Values in Evaluation: Proposal for an Exploratory Framework to Improve Coherence (Vol. 46): Institute of Development Studies.

Hummelbrunner, R. (2011). Applying systems thinking to evaluation of structural funds. In Olejniczak, K., Kozak, M., and Bienias, S. (Eds). *Evaluating the effects of regional interventions: A look beyond current structural funds' practice* (pp. 255-275). Ministry of Regional Development: Warsaw, Poland.

Institute of Development Studies Bulletin (IDS) (2015). Towards systemic approaches to evaluation and impact. Retrieved from: http://www.ids.ac.uk/publication/towards-systemic-approaches-to-evaluation-and-impact

Jacob, S. (2008). Cross-disciplinarization: a new talisman for evaluation? *American Journal of Evaluation*, 29(2), 171-94.

Jones, H. (2012). *Taking responsibility for complexity: How implementation can achieve results in the face of complex problems*. ODI Working Papers. London, UK. Retrieved from http://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/6485.pdf

Kellert, S. (2008). *Borrowed Knowledge: Chaos Theory and the Challenge of Learning Across Disciplines*. Chicago, IL: University of Chicago Press.

Levin-Rozalis, M. (2014). *Let's Talk Program Evaluation in Theory and Practice.* Tel Aviv, Israel Dekel Academic Press.

Marra, M. (2011). Some Insights from Complexity Science for the Evaluation of Complex Policies. In K. Forss, M. Marra, & R. Schwartz (Eds.). New Brunswick, NJ: Transaction Publishers.

Midgley, G. (1996). Evaluating Services for People with Disabilities: A Critical Systems Perspective. *Evaluation, 2*(1), 67-84.

Midgley, G. (2000). Planning and Evaluating Diversion from Custody for Mentally Disordered Offenders *Systemic Intervention: Philosophy, Methodology, and Practice*. New York, New York Plenum Publishers.

Midgley, G., Cavana, R. Y., Brocklesby, J., Foote, J. L., Wood, D. R. R., & Ahuriri-Driscoll, A. (2013). Towards a new framework for evaluating systemic problem structuring methods. *European Journal of Operational Research, 229*(1), 143-154.

Midley, G. (2000). Planning and Evaluating Diversion from Custody for Mentally Disordered Offenders (pp. 333-365). New York, New York: Plenum Publishers.

Midgley, G. (2006). Systemic intervention for public health. *American Journal of Public Health, 96*(3), 466-472.

Midgley, G. (2008). Response to paper "Systems thinking" by D. Cabrera et al.: the unification of systems thinking: is there gold at the end of the rainbow? *Evaluation and Program Planning, 31*(3), 317-321.

Midgley, G., & Richardson, K. (2007). Systems thinking for community involvement in policy analysis. *E: CO, 9*(1-2), 167-183.

Midgley, G. (2007). Systems Thinking for Evaluation In I. Imam & B. Williams (Eds.), *Systems Concepts in Evaluation: An Expert Anthology*. Point Reyes, CA: EdgePress of Inverness.

Midgley, G., Winstanley, A., & Foote, J. (2005). Scoping the potential uses of systems thinking in developing policy on illicit drugs: Institute of Environmental Science and Research (ESR).

Miller, R. L., Levine, R. L., McNall, M.A., Khemarko, K., & Valenti, M. T. (2009). A dynamic model of client recruitment and retention in community-based hiv prevention programs. *Health Promotion Practice, 12*(1),135-46.

Milstein, B. (2008). Hygeia's Constellation: Navigating Health Futures in a Dynamic and Democratic World (pp. 1-124).

Morell, J., Hilscher, R., Magura, S., & Ford, J. (2010). Integrating Evaluation and Agent-based Modeling: Rationale and an Example for Adopting Evidence-based Practices. *Journal of Multidisciplinary Evaluation, 6*(14), 32-57.

Morell, J. A. (2010). *Evaluation in the Face of Uncertainty: Anticipating Surprise and Responding to the Inevitable* New York, NY: The Guilford Press.

Mowles, C. (2014). Complex, but not quite complex enough: The turn to the complexity sciences in evaluation scholarship. *Evaluation, 20*(2), 160-175.

Murpy, N. F. (2016). Nine guiding principles to help youth overcome homelessness: A principles-focused developmental evaluation. In Patton, M.Q., McKegg, K., & Wehipeihana, N. (Eds.) *Developmental Evaluation Exemplars: Principles in Practice.* New York, New York: The Guilford Press.

National Cancer Institute (2007). *Greater Than the Sum: Systems Thinking in Tobacco Control.* Tobacco Control Monograph No. 18. Bethesda, MD: U.S. Department of Health and Human Services, National Institutes of Health, National Cancer Institute. NIH Pub. No. 06-6085.

Parsons, B. (2007). Initiative Evaluation Guide: A Systems-Oriented Framework for Evaluating Social Change Efforts W.K. Kellogg Foundation

Patton, M.Q. (1994). Developmental evaluation. *Evaluation Practice*, *15*(3), 311-319.

Patton, M.Q. (2016). What is Essential in Developmental Evaluation? On Integrity, Fidelity, Adultery, Abstinence, Impotence, Long-term Commitment, Integrity, and Sensitivity in Implementing Evaluation Models. *American Journal of Evaluation*.

Patton, M.Q., McKegg, K., and Weipeihana, N. (Eds.) (2016). *Developmental Evaluation Exemplars: Principles in Practice*. New York, NY: The Guilford Press.

Patton, M. Q. (2015). Systems Theory and Complexity Theory. *Qualitative Research and Evaluation Methods* (pp. 139-151). Thousand Oaks, CA: SAGE Publications, Inc.

Patton, M. Q. (2011). *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*. New York, New York: The Guilford Press.

Preskill, H., & Gopalkrishanan, S. (2014). Evaluating Complexity: Propositions for Improving Practice: FSG.

Ramage, M., & Shipp, K. (2009). *Systems Thinkers*. United Kingdom: Springer London.

Ramalingam, B. (2013). *Aid on the Edge of Chaos: Re-thinking International Cooperation in a Complex World*. United Kingdom: Oxford University Press.

Reynolds, M., Forss, K., Hummelbrunner, R., Marra, M., & Perrin, B. (2012). Complexity, systems thinking and evaluation - an emerging relationship? *Evaluation Connections: Newsletter of the European Evaluation Society*, 7-9.

Reyonlds, M., & Williams, B. (2011). Systems Thinking and Equity-focused Evaluation (pp. 115-141): UNICEF.

Rogers, P. J. (2008). Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions. *Evaluation, 14*(1), 29-48.

Russ-eft, D., & Preskill, H. (2009). *Evaluation in Organizations: A Systematic Approach to Enhancing Learning, Performance, and Change* (Second ed.). New York, New York: Basic Books.

Sanderson, I. (2000). Evaluation in complex policy systems. *Evaluation, 6*(4), 433-454.

Schwandt, T. (2014). On the mutually informing relationship between practice and theory in evaluation. *American Journal of Evaluation, 35(2), 231-236.*

Systems in Evaluation TIG (2016) Website (no date). Retreived from:http://www.systemsinevaluation.com

Stake, R. (2005). *Multiple Case Study Analysis*. New York, New York: Guilford Publications.

Stern, E., Saunders, M., & Stame, N. (2015). Standing back and looking forward: Editors' reflections on the 20th Anniversary of Evaluation. *Evaluation, 21*(4), 380-390.

Stufflebeam, D. L. & Shinkfield, A.J. (2007). Evaluation theory, models, and applications. San Francisco, CA: John Wiley & Sons, Inc.

Sweeney, L.B., & Meadows, D. (2010). *The Systems Thinking Playbook: Exercises to Stretch and Build Learning and Systems Thinking Capabilities.* White River Junction, VT: Chelsea Green Publishing.

Ulrich, W. (1988). Churchman's "Process of unfolding" and Its significance for policy analysis and evaluation. *Systems Practice, 1*(4), 415-428.

Wallace, R., Geller, A., & Ogawa, V.A. (Eds.) (2015). Assessing the use of agent-based models for tobacco regulation. The National Academies Press. Retrieved from: http://www.nap.edu/catalog.php?record_id=19018

Walton, M. (2014). Applying complexity theory: a review to inform evaluation design. *Evaluation and Program Planning, 45*, 119-126.

Wasserman, D. (2008). A response to paper "Systems thinking" by D. Cabrera et al.: next steps, a human service program system exemplar. *Evaluation and Program Planning, 31*(3), 327-329.

Wasserman, D. (2010). Using a systems orientation and foundational theory to enhance theory-driven human service program evaluations. *Evaluation and Program Planning, 33*(2), 67-80.

Westhorp, G. (2012). Using complexity-consistent theory for evaluating complex systems. *Evaluation, 18*(4), 405-420.

Williams, B. (2013). Systemic Thinking in Evaluation: Three Core Concepts (pp. 7-9): European Evaluation Society.

Williams, B., & Hummelbrunner, R. (2009). *Systems Concepts in Action: A Practitioner's Toolkit*. Stanford, California: Stanford University Press.

Williams, B. & Imam, I (Eds.) (2007). *Systems Concepts in Evaluation: An Expert Anthology*. Point Reyes, CA: EdgePress of Inverness.

Williams, B. (2005). Systems thinking. In S. Mathison (Ed.), *Encyclopedia of Evaluation* Thousand Oaks, CA: SAGE Publications, Inc.

Wolf-Branigin, M. (2013). *Using Complexity Theory for Research and Program Evaluation* Oxford, United Kingdom: Oxford University Press.

PAPER 3

VALUING IN EVALUATION WITH CRITICAL SYSTEMS HEURISTICS

Evaluation is defined by its central task of valuing, defined as the "process of

determining the merit, worth, or value of something, or the product of that process" (Scriven,

1991, p. 139). However, ironically, there is little agreement in the evaluation field about whether

and how evaluators should make judgments of the merit, worth, or significance of social policies

and programs (Schwandt, 2005). For much of the history of the field, valuing received little

attention due to the dominant value-neutral view of social science that endorsed a fact-value

dichotomy and a conviction that value judgments cannot be examined or justified rationally

(House & Howe, 1999). Many evaluators have since moved away from this view and

acknowledged that valuing is central to evaluation and that value judgments can and should be

made rationally (Davidson, 2005; Scriven, 2006, 2012; House & Howe, 1999)[1]. Even so, there

remains considerable unease and differing views on whether and how evaluators should make

value judgments.

Conversation about valuing in evaluation revolves around a variety of significant issues.

Evaluators take different stances on the broad issue of what moral-political interests valuing and

evaluation more generally should advance[2] (Schwandt, 2003; Datta, 2011). There are different

views on the way value should be determined and represented with the most widespread view

being Scriven's (1991) four-step logic of evaluation and an alternative, more experiential

approach proposed by Stake and others (1997; 2004; Stake & Schwandt, 2006). There is little

---

[1] Some evaluators still endorse an implicit fact-value dichotomy and limit evaluations to providing factual claims without value claims or factual claims with value claims that are directly associated with stakeholder values (e.g, Shadish, Cook, & Leviton, 1991) (House & Howe, 1999).

[2] Some evaluators argue for an impartial role of providing unbiased, fair information related to the public interest and others argue for evaluators to endorse particular values and interests in their work (e.g., participation, social justice) and ensure that the voices and experiences of underrepresented and disenfranchised groups are prioritized (Datta, 2011).

agreement on how criteria should be selected and justified and whether and how to reach a synthesis judgment across multiple criteria (Hall, Ahn, & Greene, 2011; Mabry, 2010; Henry, 2002; Julnes, 2012a; Schwandt, 2015). Another challenge evaluators discuss is how to select appropriate and responsive methods of valuing for particular evaluation contexts (Patton, 2012; Julnes, 2012b). There is also disagreement about the roles and responsibilities of evaluators and stakeholders in valuing (Schwandt, 2015; Alkin, Vo, & Christie, 2012). These are just a sampling of the contested issues regarding valuing illustrating that much remains unsettled regarding the making of value judgments in evaluation.

In evaluation practice, valuing often occurs implicitly with little reflection and explicit justification on the part of evaluators[3]. For example, Hall, Ahn, and Greene (2011) note that, "evaluators still grapple with acknowledging and actively engaging values in their work" (p. 195), and Julnes (2012a) observes that evaluators have been "unreflective, and even sloppy, in their approaches to valuing" (p. 4). One reason for this lack of explicit attention to valuing in evaluation practice may be that evaluation theories have yet to reconcile or even adequately address the numerous disagreements that persist regarding how evaluators should attend to values and make value judgments (Shadish, Cook, and Leviton, 1991; Greene, 2011). Other reasons include the emphasis on non-partisanship amongst some evaluation commissioners that constrains evaluators from making synthesis value judgments (Shipman, 2012); the lack of evaluation policies on valuing (Julnes, 2012b); and the absence of professional standards, guidelines, and training materials to inform valuing in evaluation practice (Morris, 2012). Regardless of the reasons, there is an evident need for further discussion of how evaluators can pay explicit attention to valuing in evaluation practice.

---

[3] Exceptions include use of rubrics in evaluation (Davidson, Wehipeihana, and McKegg, 2011; Oakden, 2013) and values-engaged approaches to evaluation (Greene, Destefano, Burgon, & Hall, 2006; Greene, 2011).

Furthermore, there is some urgency in advancing the conversation about valuing within the evaluation field. As Julnes (2012a) argues, "the lack of consensus on methods of valuing is becoming more problematic now that (a) evaluation is becoming more central to public-sector decision making and (b) the increasing pressure for evidence-based governance is pushing for more evidence-based, and hence systematic, policies on the methods of valuing appropriate for evaluation, often privileging specific approaches to assessing performance and economic impacts"[4] (p. 4). If evaluation commissioners dictate preferred methods of valuing with little input from the professional evaluation community, then the professional independence and quality of evaluation may be at risk. Several evaluation scholars are urging the evaluation community to "be proactive in articulating its own working consensus" (Julnes, 2012a, p. 4) on valuing and, for the American Evaluation Association and other professional associations, to "elevate the quality of discourse concerning public-sector valuation" (Morris, 2012, p. 87).

This paper aims to inform the conversation about valuing by explaining what critical systems heuristics (CSH) can contribute to how evaluators think about and practice valuing. The paper does not address CSH as another method of valuing[5], but as a way of thinking about values and valuing that can be used by a variety of evaluators and in a range of evaluation contexts to inform explicit and responsible choices throughout an evaluation regardless of which methods are used to render value judgments. CSH is a strand of critical systems thinking, within the field of applied systems thinking, concerned with the normative core of professional practice and how professionals can (and should) reflect on, critique, and justify the values and assumptions that

---

[4] Examples of evaluation commissioners setting constraints on valuing include the priority given to randomized controlled trials and cost benefit analysis by some federal agencies in the United States and the required (or strongly encouraged) use of five criteria (i.e. relevance, efficiency, effectiveness, impact, sustainability) developed by the Organization for Economic Cooperation and Development (OECD)'s Development Assistance Committee (DAC) as the basis for value judgments in some international evaluations.
[5] See Reynolds (2007), Ulrich and Reynolds (2010), and Reynolds (2014) for discussion of how CSH can be used as a framework and method for valuing.

influence their work (Ulrich, 2012). Many evaluators have expressed interest in systems and complexity thinking (Reynolds, Forss, Hummelbrunner, Marra, & Perrin, 2012), but connections between CSH and evaluation have remained largely on the margins of the evaluation field, particularly in the United States (Ulrich, 1988; Gregory & Jackson, 1992a; Gregory & Jackson, 1992b; Gregory, 1997; Ulrich & Reynolds, 2010; Reynolds & Williams, 2011; Reynolds, 2014). There has been attention on how critical systems methodological approaches can be applied in evaluation contexts (Cabrera, Colosi, and Lobdell, 2008; Midgley, 1996; Boyd et. al, 2006; Ulrich and Reynolds, 2010; Reynolds and Williams, 2011), but little exploration of how the ideas of critical systems can inform valuing in evaluation.

The central argument advanced here is that CSH can help evaluators be more critically reflective about valuing. Being critically reflective means identifying, questioning, and justifying the values and assumptions that influence and should influence an evaluation and, particularly, a judgment of the merit, worth, or significance of an evaluand[6]. After introducing CSH and three dimensions that comprise the foundation of its contributions to valuing in evaluation, four ways in which CSH can help evaluators be more critically reflective about valuing are examined: 1) in considering the social value of evaluation; 2) in framing the evaluand and the evaluation; 3) in selecting and justifying criteria with consideration for exclusion and marginalization; and 4) in developing a witness role. Being critically reflective in these four ways can help evaluators pay explicit attention to the choices they make regarding the influence of values and the process of valuing in an evaluation and foster responsibility for making and justifying these choices in light of potential consequences. Moreover, addressing valuing more systematically and explicitly, as discussed here, can enhance the credibility, defensibility, and legitimacy of an evaluation. The

---

[6] Evaluand refers to the program, policy, process or intervention being evaluated.

paper concludes with considerations of several hindrances to taking up a more critically reflective approach to valuing.

Critical Systems Heuristics (CSH)

While many evaluators may have heard of systems thinking, it is likely that fewer are familiar with critical systems thinking and, more specifically, the strand called critical systems heuristics. "Critical" systems thinking[7] is a particular version of systems thinking that pays attention to power relations and, specifically, how decisions are (and should be) made about which interrelationships and perspectives are considered relevant and irrelevant in a particular inquiry or intervention. CSH[8] can generally be defined as a development of critical systems thinking that aims to bring a critically reflective perspective to applied systems thinking and professional intervention broadly as well as to civil society (Ulrich, 2012; Ulrich, 2004). CSH focuses on the normative core of professional practice and how professionals can systematically identify, criticize, and justify their assumptions and the resultant selectivity and partiality of professional claims (Ulrich, 2012). CSH, as examined here, is represented in the works of C. West Churchman (1968, 1971, 1979)[9,] Werner Ulrich (1983), Gerald Midgley (2000), and Martin Reynolds (Ulrich & Reynolds, 2010).

Three dimensions of CSH comprise its key contributions to valuing in evaluation, including a critically reflective orientation to: 1) practicing as a professional; 2) thinking about social problems, interventions, and inquiry; and 3) using practical heuristics of boundary critique

---

[7] There are three distinct traditions associated with systems thinking. The first, often referred to as "hard" systems thinking, focuses on studying the interrelationships of systems, sub-systems, and their environments, such as in general systems theory, cybernetics, and complexity science. The second, called "soft" systems thinking, emphasizes exploring different perspectives on problem-situations and designing actions to address these situations, as in soft systems methodology (Checkland, 1991) and interactive planning (Ackoff, 2001). See Midgley (2007) for an introduction to these three traditions.

[8] Since the emergence of critical systems thinking in the early 1980's, it has split into at least two strands, total systems intervention and critical systems heuristics (Midgley, 2007; Ulrich, 2012).

[9] Churchman, a student of Edgar A. Singer, Jr., was concerned with the ethics of social system design and analysis (Ramage & Shipp, 2009a) and contributed to the early ideas of CSH.

and critical systems heuristics. Each of these dimensions is briefly introduced here followed by discussion of its implications for valuing in evaluation.

Professional Practice

CSH is based on an orientation to professional practice informed by critical social theory and practical philosophy (Midgley, 2000; Ulrich & Reynolds, 2010). This orientation challenges the view that social inquirers can and should be impartial—a fundamental assumption of disinterested social science[10]. Alternatively, CSH argues that social inquiry is unavoidably selective and partial, and that social inquirers have a professional and moral responsibility to handle this selectivity and partiality critically. Ulrich (2011a) contends that the aspiration for professionals to be disinterested is replaced by a particular kind of professional integrity that involves:

> The professional's awareness (or alertness, reflective stance, critical distance, etc.) regarding the institutional patterns at work, along with the worldviews, values, and interests that shape them and which put pressure on professionals to adapt to them, whether consciously or not (p. 10).

Professionals face practical choices in their work. These choices include considerations such as the purpose(s) that should guide a social inquiry or intervention; the level(s) of analysis at which a problem or intervention should be examined; the disciplinary and theoretical viewpoints that should be considered relevant to the inquiry or intervention; the methodologies or methods that should be used to investigate the problem or intervention; who or what groups are involved in and affected by the inquiry and intervention; and how, if at all, these groups should participate in the inquiry or intervention. Making these choices requires practical reasoning and deliberation, involving simultaneous consideration of the empirical (i.e. "what is")

---

[10] Disinterested social science is the idea generally endorsed by Western social science that "the enterprise of social science and the individual social scientist should be value free – that is, neutral with respect to how we should live or act as humans in society" (Schwandt, 2015, p. 76).

(e.g., facts, information, observations) and the normative ("what should be") (e.g., values, interests, perspectives). These choices are unavoidably selective –in other words, one or several options based on some facts and values are taken and other options based on other facts and values are not taken. This selectivity is always partial to some groups and interests and not to others. Ulrich and Reynolds (2010) describe this as "(i) representing only a section rather than the whole of the total universe of relevant considerations, and (ii) serving some parties better that others" (p. 247). The selectivity and partiality conditions a particular social inquiry or intervention and the claims professionals make, and this has real-world consequences for different groups and interests concerned with and/or affected by the inquiry or intervention. Therefore, choices should be made transparent and subject to reflection and critique regarding their selectivity and partiality.

New Ways of Thinking

CSH is characterized by a way of thinking about social problems, interventions, and inquiry that is distinct from other systems and complexity traditions as well as from the predominant ways of thinking in the evaluation field. CSH distinguishes between a situation of interest and a particular framing of a situation. Situations of interest are the real-world[11] circumstances that someone or some group(s) wants to understand or change (Reynolds, 2008). Situations are generally characterized by continually changing influences including complicatedness (i.e., multiple, interconnected factors), complexity (i.e., multiple perspectives and values), and conflict (i.e., differences in power and control) (Reynolds, 2015a).

---

[11] CSH views both situations and framings as epistemological (i.e. relating to knowledge and knowing about the world) and not ontological (i.e. regarding the nature of reality).

Framings are the socially and cognitively constructed contexts, also called reference

systems[12], people use to simplify situations so that they can practically understand, make claims

about, and act in these situations. A framing is defined as a way of understanding a situation

based on descriptive assumptions and claims about what is (i.e. facts) and normative views and

assertions about what should be (i.e. values). Any particular framing of a situation is inevitably

selective – it is not feasible or useful to study the entirety of the situation from all possible points

of view. Decisions must be made about which facts and values to consider relevant and not

relevant. These decisions are called boundaries because they conceptually bound or limit how the

situation of interest is framed. Boundaries can be implicit or explicit and are defined as the

assumptions, decisions, or judgments made about what empirical (e.g., observations, facts,

information) and normative (e.g., values, norms, perspectives, interests, worldviews)

considerations are relevant and not relevant regarding a situation of interest.

Boundaries constitute and condition a framing as well as influence who or what is

included, excluded (i.e. ignored, overlooked, invisible), and marginalized (i.e. neither fully

included or excluded) (Midgley, 1992). Any situation of interest can be understood through

multiple framings depending on what boundaries are used and, subsequently, which facts and

values are considered relevant and irrelevant. Stakeholders – people or groups with a stake (or

interest) in the situation – frame the situation differently and, conversely, different framings

identify different groups and interests as stakeholders. Because each framing will be selective

and partial, those involved in designing and managing social interventions and conducting social

inquiry should be critically reflective about how they frame a situation of interest and the

selectivity and partiality of any framing. Examining multiple boundaries and framings offers a

---

[12] Framing is used here, rather than reference system (a term used by Ulrich (1983) and Ulrich & Reynolds (2010), in order to reserve the latter term to refer to a specific framing developed by making the twelve boundary judgments identified in critical systems heuristics.

means for continually learning about a situation of interest from different points of view, which can facilitate mutual understanding[13] and responsible social action (Reynolds, 2010; Ulrich and Reynolds, 2010).

Practical Heuristics

CSH offers two practical heuristics to guide professionals in being critically reflective about how they bound and frame situations of interest (e.g., problems, policies, programs) and how they address the selectivity and partiality of framings and claims made about situations. The first heuristic, boundary critique, is based on a process of questioning, debating, and justifying decisions about what facts and values are and should be considered relevant and irrelevant with regards to a particular way of framing a situation of interest. The process involves 1) identifying boundaries that influence or should influence a framing; 2) examining potential practical, political, and ethical consequences or issues; 3) considering alternative boundaries; and 4) ultimately making transparent and justifying the boundaries used while remaining open to contestation and revision (Ulrich, 2002a). Professionals can carry out this process reflectively to surface and consider the consequences of boundaries they use or are considering using; stakeholders can employ this process to question implicit and consider alternative boundaries and what consequences they may pose for different views, concerns, and interests; and citizens can use this process to challenge the boundaries professionals use, highlight consequences, and recommend alternatives (Ulrich, 2002a). Boundary critique can be conducted in two modes as depicted in Table 1, an actual mapping mode to identify what facts and values are considered relevant and irrelevant or the empirical selectivity of a framing and an ideal mapping mode to

---

[13] Ulrich and Reynolds (2010) point out that, "when people talk about situations, it often happens that their views differ simply because they frame situations differently…by examining the underlying boundary judgments, we can better understand people's differences and handle them more constructively" (p. 245).

identify what facts and values could or should be considered relevant and irrelevant or the

normative selectivity (Ulrich, 2005, p. 8).

Table 3.1. Four perspectives for examining selectivity (Ulrich, 2005,p. 8)

Perspective	Empirical selectivity ('is' mode)	Normative selectivity ('Ought' mode)
'Facts'	Actual mapping: What 'facts' are considered relevant and which ones are left out?	Ideal mapping: What 'facts' ought to be considered relevant and which ones should be left out?
'Values'	Actual mapping: What 'values' are considered relevant and which ones are left out?	Ideal mapping: What 'values' ought to be considered relevant and which ones should be left out?

Boundary critique is the basic process by which professionals can be more critically

reflective about the assumptions that influence their work and the claims they make. However, it

can be challenging to know which boundaries to surface and question. Therefore, Ulrich (1983),

based on the work of Churchman (1971, 1979)[14] developed a heuristic that identifies categories

and kinds of boundaries that typically influence the way situations are framed.

As displayed in Table 2, a second heuristic device comes in the form of 12 essential

questions, the collective answers to which demonstrate a particular way of bounding and framing

a situation of interest whether an intervention or evaluation. The questions are organized into

four sources of influence on a framing, and each question identifies a boundary judgment to be

made: basis of motivation (e.g. where does a sense of purposefulness and value come from?);

basis of control (e.g. who is in control of what is going on and is needed for success?); basis of

knowledge (e.g. what experience and expertise support the claim?); and basis of legitimacy (e.g.

where does legitimacy lie? (Ulrich, 2002b). The first three sources of influence (i.e. motivation,

---

[14] Churchman developed these categories based on those identified in *Immanuel Kant's Critique of Pure Reason* (Kant & Smith, 1992).

control, and knowledge) refer to who or what is involved in a situation and the fourth source (i.e.

legitimacy) refers to who or what is affected by the situation[15].

Table 3.2. Boundary categories and questions of CSH (adapted from Ulrich & Reynolds, 2010, p. 244)

Relevant Groups	Sources of influence	Boundary judgment	Questions
The involved	Motivation	Beneficiary	Who ought to be/is the intended beneficiary?
		Purpose	What ought to be is the purpose?
		Measure of improvement	What ought to be/is the measure of success?
	Control	Decision maker	Who ought to be/is in control of the conditions of success?
		Resources	What conditions of success ought to be/are under the control?
		Decision environment	What conditions of success ought to be/are outside the control of the decision maker?
	Knowledge	Expert	Who ought to be/is providing relevant knowledge and skills?
		Expertise	What ought to be/are relevant new knowledge and skills?
		Guarantor	What ought to be/are regarded as assurances of successful implementation?

---

[15] Ulrich and Reynolds (2010) further identify a particular social role (i.e. stakeholder) associated with each source of influence and specific concerns that those in this role may have with regards to the situation (i.e. stakes) as well as key problems regarding this role and concern(s) (i.e. stakeholding issues) (see full table on p. 244). These additional distinctions were not included here to simplify the table for evaluators unfamiliar with CSH.

Table 3.2 (continued)

Relevant Groups	Sources of influence	Boundary judgment	Questions
The affected	Legitimacy	Witness	Who ought to be/is representing the interests of those negatively affected but not involved?
		Worldview	What ought to be/are the opportunities for the interests of those negatively affected to have expression from freedom from the worldview?
		Emancipation	What space ought to be/is available for reconciling differing worldviews among those involved and affected?

CSH can be used in a descriptive mode to surface implicit ways of understanding and framing a situation of interest and consider alternative descriptive framings or a normative mode to surface implicit assumptions and views on how things should be with regards to the situation of interest and consider alternative normative framings. Similar to boundary critique, CSH can be used reflectively by professionals, dialogically by or with stakeholders, or questioningly by citizens. Answering the twelve questions, in either or both modes, facilitates a process by which the values and motivations, decision-making structures, knowledge basis, and moral basis that influence a situation of interest can be systematically examined from different points of view (Ulrich & Reynolds, 2010).

Contributions of CSH to Valuing in Evaluation

Evaluators can develop this critically reflective orientation and way of thinking and make use of these practical heuristics in four specific ways, including determining the social value of evaluation as a practice, framing the object of evaluation as well as the evaluation itself, selecting and justifying criteria used in determining the value of the evaluand, and in performing what is called a "witness role" in CSH.

Considering the Social Value of Evaluation

Evaluators hold different views with regards to whether evaluations should advance particular social and political values and, if so, which and whose values, alternatively discussed as regulative frameworks (Greene, 1997), ethical aims (Schwandt, 1997), and the role of politics in evaluation (Schwandt, 2003; Datta, 2011). Evaluators also wrestle with the related issue of whether and how the socio-political values they personally hold should influence the evaluations they conduct (Mabry, 2010; Levin-Rozalis, 2014). CSH challenges evaluators to reject dichotomous thinking—either impartiality or partiality towards particular socio-political values and interests—in favor of careful, critical examination and subsequent justification of which and whose socio-political values are advanced and not advanced in the conduct of a given evaluation.

Evaluation is premised on a conviction that evaluating social policies and programs can contribute to the betterment of society, understood as the reduction of social problems and increased meeting of human needs (Mark, Henry, and Julnes, 2000). The idea that continual social progress or improvement is possible and desirable is embedded in most theories of evaluation practice. As Mark, Henry, and Julnes (2000) point out, "without the possibility of social betterment, evaluation would be at worst an empty exercise, at best a fulfillment of

114

curiosity. That the possibility of social betterment exists is both a personal motivation for evaluators and a critical part of the rationale for the field" (p. 24).

While the possibility of betterment is foundational to evaluation, there is general agreement that evaluators should not be involved in defining what "betterment" means. In pluralistic, democratic societies, there are inevitably numerous and, often conflicting, viewpoints and values regarding what constitutes the "good society", what qualifies as a social problem and what social policies and programs are necessary to address social problems. Establishing the normative ends of society as reflected in the choice of social policies and programs is viewed as a political task best left to politicians and citizens (Weiss, 1998; Shipman, 2012; Chelimsky, 2014). For the most part, evaluators do not question the normative ends assumed in social policies and programs[16.] Instead, they aim to assess things like the efficiency, effectiveness, and impacts of social policies and programs in attaining these normative ends. Still, evaluators wrestle with the issue of how and which socio-political values to advance in an evaluation.

There are two different, widespread ways in which evaluators view the social value of evaluation. The first is a public interest orientation[17], characterized by a "belief in the public interest or common good that transcends the diversity, a highest common denominator, and a role for evaluators as sources of unbiased, fair information relating to this interest" (Datta, 2011, p. 279). The public-interest oriented evaluator remains nonpartisan and noncommittal with respect to favored social goals and values: "evaluators can and should disqualify themselves if there are conflicts of interests in values or anything else. Objectivity and impartiality are possible and essential (e.g., GAO standards)" (Datta, 2011, p. 279).

---

[16] Exceptions include transformative and culturally responsive approaches to evaluation practice (Mertens, 2006; Hood, Hopson, & Kirkhart, 2015).

[17] This orientation was first articulated and continues to be endorsed by evaluators that tend to be concerned with "evaluations of high-stakes, controversial, large-scale national policy issues" (Datta, 2011, p. 273).

115

The second way, a populist orientation (Datta, 2011), does not strive for political neutrality, but, instead, aims to ensure that diverse viewpoints and values are included in the evaluation, particularly, the voices and experiences of historically and contemporarily underrepresented and disenfranchised groups in society (Datta, 2011). This stance is grounded in the conviction that "evaluators are saturated with their own values and thus inherently subjective. Evaluation can only seek to reveal different realities. Objectivity is an illusion (e.g., Guba and Lincoln)" (Datta, 2011, p. 279). Evaluators' commitments to ideas of empowerment, citizen participation, and social justice, tend to heavily influence the evaluation purpose, interactions with stakeholders, and the evaluation design, process, analysis, and interpretation (Datta, 2011). Schwandt's (1997) notion of a value committed framework for evaluation and Greene's (1997, 2006ab, 2011) view of what it means to be values-engaged bear some similarity to a populist orientation. Most culturally responsive and social justice-oriented approaches to evaluation are populist in orientation as well (e.g., Hood, Hopson, and Kirkhart, 2015; Mertens, 2006).

CSH does not share the long-standing conviction that evaluators should not question the normative assumptions underlying social policies or programs. CSH argues, for both practical and ethical reasons, that evaluators should question why, for example, some state of affairs should be considered a social problem, who or what should be considered as having this problem, and what betterment with regards to this problem should mean. Churchman (1971, 1979) argued that conceptualizations of problems and betterment tend to exclude examination of the wider situation in which the problem is embedded and other perspectives on this situation; he refers to this as the "environmental fallacy" (Churchman, 1979, p. 4). This exclusion is practically and ethically problematic as what appears to be an improvement through one level of analysis or from one perspective often is worsening of the problem or another problem altogether

from another level of analysis or perspective. One pervasive example of this is the tendency of humans to define improvement without attending to the consequences for the natural environment; from the point of view of the natural environment, many so-called improvements would be considered destructions.

Aspiring for impartiality, as in a public interest orientation, or partiality towards particular perspectives, values, and interests, as in a populist orientation are problematic from a CSH point of view because neither adequately addresses the selectivity and partiality of evaluation as well as potential consequences. In CSH, the aim is to identify, critique, and justify the socio-political values and normative assumptions that influence and should influence a particular way of framing an intervention or evaluation and ensure the resultant selectivity and partiality is handled as transparently and responsibly as possible[18]. Hummelbrunner and Reynolds (2013)[19] associate the critical systems concept of boundaries with triple loop learning. Whereas single loop learning asks, "Are we doing things right?" and double loop learning asks, "Are we doing the right things?" triple loop learning asks, "What makes this the right thing to do?" (p. 1). Examining this question requires "critically reflecting on the rules and relations of power that affect behaviour and cognition patterns" including those of "evaluation commissioners and evaluators themselves" (Hummelbrunner and Reynolds, 2013, p. 2). Such a critically reflective orientation fits well with what Schwandt (1997) calls a value-critical framework for evaluation that informs "improving praxis by enabling practitioners to refine the rationalities of their practices… this can only be achieved by helping practitioners develop a kind of educative, critically reflective self-knowledge that enables them to question the beliefs and unstated assumptions that sustain a particular practice of education, management, health care,

---

[18] Reynolds (2014) discusses a similar aim of using CSH to question the ethical basis of an intervention.
[19] The distinction between single, double, and triple loop learning has been widely used in the organizational literature (Hummelbrunner and Reynolds, 2013).

and so forth" (p. 35). This role also bears similarity to Schwandt's (2008) conception of evaluation as a social conscience:

This is a society in which we ask serious and important questions about what kind of society we should have and what directions we should take... Evaluation in such an environment is a kind of social conscience; it involves serious questioning of social direction; and it is a risky undertaking in which we endeavor to find out not simply whether what we are doing is a good thing but also what we do not know about what we are doing (p. 143).

Adopting a more critically reflective orientation towards the social value of evaluation requires evaluators to take on new roles and relationships with commissioners and stakeholders and new ways of reflecting on and justifying the values that influence their work. Rather than aspiring to be neutral (as in a public interest orientation) or engaged (as in a populist orientation), CSH calls for evaluators to be critical – continuously calling attention to whose and which socio-political values are and should be influencing an evaluation and potential consequences of this for other groups and interests.

Framing the Evaluand and Evaluation

When evaluators first learn about an evaluand (e.g., a particular intervention such as a policy or program) and begin planning an evaluation, the evaluand and evaluation can be considered situations of interest – messy, not yet bound, continuously changing, influenced by many factors, and subject to differing perspectives as well as differences in power and control. The primary task for evaluators is to carefully and responsibly choose boundaries within which an evaluand can be examined and an evaluation can be designed and carried out. Evaluators generally lack practical frameworks for making these choices and, typically, these choices are primarily descriptive (regarding what is) and not normative (regarding what could or should be).

The practical heuristics of CSH (i.e. boundary critique and the twelve questions of critical systems heuristics) can help evaluators make these choices more explicitly and systematically and with consideration of different points of view.

While framing an evaluand and evaluation are distinct tasks (perhaps carried out at different times), the basic process – selecting and justifying boundaries while remaining open to revision – is the same. Engaging in the act of framing does not require substantial resources or time, as it can be a reflective activity[20] carried out by an evaluator in conjunction with other means of learning about and describing an evaluand (e.g., document review, stakeholder interviews, observations, literature review). Similarly, examining and justifying the way an evaluation is framed is already a routine task, as discussed in the Better Evaluation Rainbow Framework. CSH simply contributes a heuristic and process to inform this task. The three steps of this iterative process are briefly introduced followed by benefits of this process.

Surface. Framing first requires surfacing existing assumptions and ways of understanding an evaluand and evaluation. Descriptions of evaluands provided by evaluation commissioners or program or policy designers, in grant proposals, websites, or other documents discussing the goals, operations, benefits and other aspects of an evaluand, provide a useful starting point for identifying descriptive and normative assumptions. From the point of view of those designing the evaluand (and possibly also those commissioning an evaluation), evaluators can systematically identify implicit framings by answering the twelve CSH questions in both descriptive and normative modes. To surface implicit framings of the evaluation, evaluators carry out a similar process, although questions take on slightly different meanings. Evaluation proposals, theories of evaluation practice (e.g., utilization-focused, culturally responsive), and

---

[20] Evaluators can also use a participatory and deliberative process for facilitating stakeholder discussion for purposes of critically framing an evaluand. This process would require more time and resources (e.g., funding, space, staff).

the evaluator's own experience and preferences can inform surfacing an initial framing of an evaluation.

Consider consequences and alternatives. Potential consequences of each framing are identified followed by consideration of alternative framings from different points of view. Three types of consequences are useful to consider, although, in practice, consequences typically do not fall neatly into one type. Practical consequences include things that may influence the logistics of implementing a policy or program, conducting an evaluation, or the wider situation of interest. These may come from groups and interests that are not considered relevant within a particular framing; "if they are marginalized there is a risk that they will oppose the intervention and hinder its execution" (Williams, 2015, p. 14). Another source of practical consequences comes from limiting an understanding of an evaluand by some boundaries (e.g., time frame, level of analysis, theoretical lens), which can limit identification of other (potentially unanticipated) consequences. Political consequences are what may influence who or what has and does not have power, control, voice, and agency in an evaluand or evaluation. These come from some groups and interests being advanced or better served and others being ignored or negatively affected such that the way an evaluand or evaluation is framed reinforce or challenge the power and privilege of some groups and interests and the exclusion and marginalization of others. Ethical consequences are related with good/right and bad/wrong and stem from the fact that all framing is influenced by ethical beliefs, worldviews, or normative assumptions about how things should be in society. Promoting some ethical views inevitably (but perhaps inadvertently) excludes and marginalizes others, which may be problematic for various groups and interests involved in and/or affected by an evaluand or evaluation (Williams, 2015). Each of these types of consequences, and perhaps others, should be considered in terms of who or what is included,

excluded, and marginalized and what evaluators (and others) might do to potentially mitigate these consequences.

After considering potential consequences of initial framings, evaluators should develop alternative framings from different points of view. For an evaluand, potential points of view can be identified based on people and interests that are actually present in the particular situation of interest or the judgment and experience of the evaluator. For an evaluation, this could involve developing a framing from the point of view of different theories of evaluation practice (e.g., culturally responsive, empowerment, theory based, accountability-driven) or from different stakeholders involved in or affected by the evaluand (e.g., commissioners, parents, students, teachers, administrators, local citizens). Along with developing alternative framings, evaluators should reflect on and discuss who or what is included, excluded, and marginalized in each of these framings and potential practical, political, ethical, and other consequences as well as what might be necessary to mitigate theses consequences.

Consider the example discussed by Levin-Rozalis of a program funded by the OECD to improve the welfare of a migrant community by training young people in government leadership. While this is a worthwhile endeavor from the point of view of those designing the initiative and commissioning the evaluation, other points of view reveal potential significant consequences: traditional leaders could view the initiative as making the community "easier for the authorities to live with" and "shunting traditional leadership" (p. 272); the majority of the community, who is traditional and not well educated, may view the initiative as futile as they will not be able to follow these new young leaders and, even further marginalizing as these members will be isolated and with fewer youth; for social workers and police, the initiative may generate "endless

work" (p. 273); and, for those with radicalized interests, the initiative may support opposition to current government authorities.

The evaluation of such an initiative could also be framed in a variety of ways, each with potential consequences for different groups and interests: a theory-driven approach examining the processes and mechanisms by which youth develop (or do not develop) leadership skills may focus on the points of view of youth and community members who are involved in the initiative but exclude others who question the value of the initiative; a democratic focus facilitating participation of a range of community members in the evaluation could overlook other groups, such as social workers, police, and those with radicalized interests, as few would identify them as members of the community; or a developmental focus that helps traditional leaders develop innovative ways to support the welfare of the migrant community, which may have consequences for the authority of the OECD, local government, as well as social workers and police in the community. This brief example illustrates that there is no definitive way to frame an evaluand or evaluation and that all framings have potential consequences; the point is to consider multiple framings and identify (so as to anticipate and mitigate) potential consequences.

Select, justify and revise as needed. Most importantly, evaluators must select and justify the framing(s) used while remaining open to contestation and revision. In CSH, framings can be justified on the grounds that exclusion, marginalization, and other potential consequences have been considered and mitigated where necessary or appropriate; or that there is some consensus among a group of stakeholders regarding which framing(s) to use; or that a moral theory or framework providing support for a particular framing was endorsed by those involved in and affected by a particular situation. These or other bases for justifying framings may be necessary in evaluation. Whichever framing(s) are selected and justified, these should be made

122

transparent in planning, reporting, and communicating the evaluation[21] and reexamined as well as revised to address changes in the actual situation of interest, knowledge about this situation, and emerging issues.

Framing evaluands and evaluations more explicitly and systematically has three major benefits. First, this process helps to identify and anticipate potential practical, political, and ethical consequences of an evaluand on different groups and interests. As Williams (2015) points out, "single-perspective approaches don't acknowledge the possibility that 'unintended consequences' may have been intended by someone who perceived the intervention through a different lens" (p. 12). Second, it offers a way to systematically consider the contexts of an evaluand and evaluation by surfacing and examining various sources of influence from different points of view. Scholars advancing realist (e.g., Julnes 2012b) and constructivist (e.g., Dahler-Larson and Schwandt, 2012) conceptions of context have argued for evaluators to be more explicit and systematic with regards to how context influences an evaluation. Third, and perhaps most importantly, this process can strengthen the credibility, defensibility, and legitimacy of an evaluation. Credibility is enhanced by ensuring that evaluators systematically consider multiple points of view and purposes and do not automatically (i.e., unreflectively, uncritically) reinforce the assumptions of those designing evaluands and/or commissioning evaluations. Defensibility is enhanced by supporting the independence of an evaluation from the political and normative assumptions associated with an evaluand, a concern raised by Chelimsky (2012), and by helping to make assumptions transparent and subject to examination. Legitimacy is enhanced by considering other normative assumptions, perspectives, and worldviews and potential practical,

---

[21] Many evaluators include descriptions of the evaluand and the evaluation context (e.g., purpose, audience, key questions, methodology) in an evaluation plan and in reports to stakeholders. These are places that particular ways of framing can be explicitly justified and communicated.

political and ethical issues, addressing a concern raised by Julnes (2012b) with regards to valuing that evaluations are limited by "unexamined worldviews" (p. 122).

Selecting and Justifying Criteria With Consideration for Exclusion and Marginalization

Evaluative claims about the merit, worth, and significance of social policies and programs are made on the basis of value-laden criteria[22] and factual claims (House, 2001). There are numerous sources of potential criteria including stated objectives, effectiveness, relevance, equity focus, efficiency, social impact, sustainability, cultural relevance and responsiveness, established requirements, expert opinion, and needs assessment (Schwandt, 2015). Evaluators face the challenge of choosing criteria and justifying that choice to multiple audiences (Hall, Ahn, & Greene, 2011; Mabry, 2010; Henry, 2002; Julnes, 2012a; Schwandt, 2015). CSH urges evaluators to select and justify criteria with consideration of who or what may be excluded and/or marginalized by whichever criteria are chosen. Taking up such a process in evaluation requires revisiting the ways in which evaluators determine and represent value and expanding prescriptive valuing to explicitly incorporate bases for justifying criteria selected.

CSH suggests a way of determining and representing value referred to here as quality-as-mapped. This involves comparing a descriptive map of what is with a normative (or ideal) map of what should be to generate an evaluative assessment, judgment, critique, or conversation. Determining value in quality-as-mapped involves rendering multiple evaluation judgments based on different boundaries and framings in order to develop a more comprehensive and pluralistic understanding of the value of an evaluand[23] while explicitly acknowledging the boundaries and framings used and who or what is and might be excluded and marginalized by these boundaries

---

[22] Criteria, as used here, refers to principles or standards on which a judgment is made and includes those used in both quantitative (e.g., performance indicators, cut scores) and qualitative (e.g., issues, perceptions) methodologies.
[23] This is distinct from taking a descriptive approach to valuing that involves identifying criteria of interest to relevant stakeholders and reporting performance on each of these criteria.

and framings. The 12-question heuristic could be employed to inform rendering evaluative judgments, as discussed by Ulrich & Reynolds (2010) and Reynolds & Williams (2011). Additionally, there are numerous other methodologies and mapping techniques from the systems fields (e.g., system dynamics modeling, viable systems modeling, soft systems methodology) that could be used by evaluators to construct descriptive and normative maps of an evaluand and render comparisons between these maps. Representing value in quality-as-mapped would generally involve the actual maps, in the form of tables, illustrations, simulation models, etc., and discussion of evaluative conclusions and implications based on the comparison between the maps.

Alternatively, using boundary critique to consider exclusion and marginalization could be incorporated into two distinct ways of determining and representing value[24] discussed in the evaluation literature, referred to by Stake and Schwandt (2006) as "quality-as-measured" and "quality-as-experienced" (p. 407-408). The first and widespread[25] way conceives of value as something that can be systematically determined and explicitly stated using the logic of evaluation developed by Scriven (1991):

1. Establish the criteria of merit. On what dimensions must the evaluand do well?

2. Constructing standards. How well should the evaluand perform?

3. Measuring performance and comparing with standards. How well did the evaluand perform?

---

[24] While these two ways of determining and representing value are presented here as opposing choices, both ways of approaching value may inform a particular evaluation. Stake and Schwandt (2006) contend that, "the relative suitability of quality-as-experienced and quality-as-measured depends, in part, on whether or not the evaluand can be intellectually and practically embraced by a single evaluator (or a small evaluation team)" (p. 410). In many evaluations, there is a need for both standardized and measured accounts of value as well as storied and experiential accounts (Stake & Schwandt, 2006).

[25] Conceiving of value as a conclusion reached on the basis of criteria and standards is widespread in evaluation (Scriven, 1991; Weiss, 1998; Davidson, 2005)

4. Synthesizing and integrating data into a judgment of merit or worth. What is the merit or worth of the evaluand? (Julnes, 2012a, p. 6).

Incorporating boundary critique into quality-as-measured would require identification of influential boundary judgments in this approach such as the choice of criteria, indicators, weights, standards, and measurement scales and reflecting on alternative boundary judgments that could be made and what consequences these judgments would have for the kinds of evaluative claims made

The second way, expressed in arguments put forth by Stake and his students (1997, 2004) and Stake and Schwandt (2006) conceives of value as something that is experienced and tacitly understood throughout an evaluation by participants in a given program or project and, only later, made explicit. In the words of Stake and Schwandt (2006): "the notion of quality-as-experienced draws attention to both the subjective and inter-subjective meanings we attach to events, personal encounters, and places" (p. 408). Value judgments are presented not as statements or propositions but in the form of narratives, of the evaluand as it was experienced and often highlighting multiple perspectives, vicarious experience, and issues for consideration. The evaluator's experience and assessment of value shapes the portrayal of this evaluative description, but, usually, there is room for other interpretations and even challenges to this depiction. An evaluator(s) could use boundary critique principles to consider alternative observations, experiences, stories, perspectives, and issues that could inform evaluative narratives about quality and how these alternatives would influence these narratives differently. This approach could also incorporate a participatory mode of boundary critique in which those involved in and affected by a policy or program discuss alternative boundary judgments and underlying rationales for why they should or should not be considered relevant.

Regardless of whether evaluators conceive of quality as mapped, measured, or experienced, they must select and justify criteria upon which evaluative judgments are made. There are two approaches to valuing that treat criteria differently. In descriptive approaches, an effort is made to address all relevant criteria of major stakeholder groups by demonstrating how the evaluand performed on each of those criteria. In a prescriptive approach, one or several criteria are considered superior or more important than all others. CSH rejects the notion that a solely descriptive approach is possible, supports taking a prescriptive approach, and challenges those taking a prescriptive approach to attend to how they justify these criteria.

Those taking a descriptive approach to valuing aspire to communicate the values and interests of all relevant stakeholders to decision-makers as is evident in Mark, Henry, & Julnes' (2000) notion of values inquiry. In this approach, criteria are selected based on stakeholder input (Julnes, 2012a) or input from the wider public, such as the nineteen values identified by Chelimsky (2014) associated with the public interest. The use of in-depth observation and discussion with stakeholders in responsive evaluation to identify values (Stake, 2004, 2006) is a descriptive approach to valuing. In theory, there is no consideration of whether any particular criteria should have priority over others and, therefore, a descriptive approach typically does not include making an overall or synthesis value judgment. From a CSH point of view, a descriptive approach does require making decisions about which groups and interests are relevant stakeholders and which criteria, identified by these stakeholders, should be included and which should be excluded – an issue which is not straightforward and usually rather uncertain and difficult in evaluation practice (Henry, 2002). Therefore, a strictly descriptive approach (i.e. one that does not privilege some criteria over others) is not possible.

From a CSH perspective, all approaches to valuing are fundamentally prescriptive. Those taking a prescriptive approach to valuing select criteria typically for reasons other than they were recommended through stakeholder and public input and, includes prioritization of some criteria over others. Sources of prescriptive criteria include the objectives of the policy or program being evaluated (Julnes, 2012a), commonly used criteria in a particular arena such as the DAC criteria (Kaplan, 2014), an evaluator's choice of criteria based on some theory of evaluation practice (e.g., Greene, 2011), or on an ethical theory such as Rawls' theory of social justice as used in democratic deliberative evaluation (House & Howe, 1999). Consideration of exclusion and marginalization in this process can strengthen the justification of criteria included and prioritized in an evaluation.

CSH urges those taking a prescriptive approach to valuing to attend to how evaluators justify choices about which criteria to select and not select to multiple audiences. Reflectively considering exclusion and marginalization is a necessary but not sufficient basis for justifying choices. One basis for justifying criteria is to ground selection in a participatory process in which stakeholders deliberate about which criteria to select, in part, by considering potential limitations and consequences. Ulrich (1983) argued that values should be normatively acceptable to those involved in and affected by an intervention and this occurs through a participatory, dialogic process. Midgley (2000) asserts that considering multiple values and reflecting on any furthering of exclusion and marginalization offer an ethical basis for selecting and justifying values. In addition to being potentially logistically difficult to arrange, participatory processes still include some stakeholders and exclude others so evaluators would still need to be responsible for continuously bring attention to potential consequences for exclusion and marginalization (Midgley, 2000).

128

Developing a Witness Role

CSH urges evaluators to incorporate a witness role for groups, interests, and worldviews potentially affected by the situation of interest an evaluand (and evaluation) addresses to participate directly or be represented in the valuing process. Witnessing means identifying and calling attention to who or what may be negatively affected and fostering responsibility in an evaluation for mitigating further exclusion and marginalization. This role expands on a suggestion made by House and Howe (1999) that "sometimes evaluators must represent the views of missing stakeholders" (p. 118) by broadening those represented beyond stakeholders (as people) and integrating this role specifically into valuing. Those affected may include stakeholders, but typically there are others that are affected and not included in or represented by involved stakeholders. When those affected cannot directly participate in valuing processes, such as when this group refers to non-human species, the natural environment, future generations, or others unable to speak on their behalf, evaluators themselves or selected stakeholders should serve in a witness role to ensure consideration of these groups and interests. This role has origins in Churchman's (1979) notion of "enemies"; he argues that those involved in social interventions or inquiry should enter into conversation with the enemies of their ideas, values, or arguments in order to learn about their limitations and make adjustments as needed.

Evaluators can incorporate a witness role into valuing regardless of whether they or stakeholders are primarily responsible for rendering evaluative judgments. Alkin, Vo, and Christie (2012) distinguish between three roles for the evaluator in making judgments of the value of a policy or program. In the first mode, stakeholders rather than the evaluator have the primary responsibility for such judgments. This typically involves stakeholders establishing the standards, evaluators providing data, and the primary intended stakeholders or other groups

carrying out the valuing. A witness role in this mode could involve select stakeholders being

responsible for reflecting on and raising concerns about potential consequences of the valuing

process and conclusions for those affected. In the second mode, stakeholders together with the

evaluator are jointly responsible for rendering value judgments. This typically involves

evaluators providing the data and helping establish a framework for valuing or providing the data

and guiding others in the valuing. Evaluators sometimes participate in the valuing with

stakeholders in this mode. Evaluators could incorporate a witness role by explicitly incorporating

consideration of those affected by a valuing into the valuing process. In the third mode, the

evaluator alone is responsible for rendering value judgments. This typically involves valuing

based on evaluator values, evaluation expertise, program expertise, or scientific appraisal. While

this mode may prove most difficult for incorporating a witness role, evaluators could create some

mechanism by which the views and interests of those affected are considered in valuing such as

through an external review process, a stakeholder or public forum, or the use of critical theory or

some other framework to critique one's value conclusions.

A witness role is important to incorporate into valuing as a way to identify, address, and

foster responsibility for addressing potential consequences of an evaluation as well as developing

the moral legitimacy of an evaluation. In CSH, those groups and interests affected but not

involved in a social intervention, inquiry, or evaluation are sources of legitimacy. Moral

legitimacy is the idea that some claim or action is justifiable on the basis that it's potential

consequences have been considered and deemed acceptable by those making the claim or taking

the action as well as those groups or interests who could be considered "enemies" of this claim

or action. Legitimacy cannot be generated or claimed by those involved, but must come from

those affected and not involved. Ulrich (2000) argues that citizens should be considered sources of legitimacy:

> By a civil society I understand a society in which the basic source of legitimacy lies with the individual citizen. A civil society is a democracy in which ordinary people can effectively participate in decisions on matters of collective or public concern (p. 2).

Incorporating citizens into an evaluation in a witness role would require them to speak not only on in their own interests but more specifically on behalf who or what may be negatively affected.

However, a witness role challenges the widely held assumption that evaluation is an intrinsically good or legitimate activity and requires that evaluations themselves need to be evaluated – as in meta-evaluation. But meta-evaluation must go beyond the dimensions of quality commonly addressed, as, for example, when applying the Program Evaluation Standards or Scriven's Key Evaluation Checklist to an evaluation (Scriven, 2005), to include scrutiny of the ethical and political assumptions of the evaluation itself. As Reynolds (2015b) argues, evaluations may be sources of deception if they do not subject their own political and ethical assumptions to scrutiny. And even when these assumptions are surfaced, examined, and justified, "there is always a built-in risk about the value of the evaluation guarantee" (Reynolds, 2015b, p. 2). Evaluators can inevitably be deceived about the world and therefore need ways to address this risk and call attention to potential deception (Ulrich, 2009). A critically reflective meta-evaluation would surface and question the political and ethical assumptions of an evaluation and raise questions about groups and interests potentially negatively affected. Such an approach is supported by long-standing arguments made by House, MacDonald, Kushner, Schwandt and others that meta-evaluation should address political and ethical assumption of the practice of evaluation itself.

Conclusion

The lack of agreement in the evaluation field about whether and how evaluators should make judgments of the merit, worth, or significance of social policies and programs poses a considerable challenge for evaluators. Given the diversity of evaluators, ways to conduct evaluations, and evaluation contexts there is not likely to be consensus regarding how evaluators attend to values or which methods evaluators should use to render value judgments (Julnes, 2012b). However, there is a need for evaluators to pay more explicit attention to values and valuing in evaluation practice.

The argument advanced here is that CSH can help evaluators identify, question, and justify the values and assumptions that influence and should influence an evaluation and, particularly, judgments of the merit, worth, or significance of an evaluand. CSH suggests four major contributions to existing ways of thinking about values and practicing valuing in evaluation. The first contribution invites evaluators to consider the particular values and interests that are advanced (and not advanced) in an evaluation. The second contribution draws explicit attention on the way boundaries are set in the tasks of framing an evaluand and an evaluation and proposes a practical process to surface implicit boundaries, consider alternatives and potential consequences, and select, justify, and revise boundaries used in an evaluation. The third contribution provides an additional way of conceiving of quality (i.e. quality-as-mapped), which could be used alongside or as an alternative to current conceptions of quality in evaluation (i.e. quality-as-measured and quality-as-experienced) and suggests strengthening prescriptive approaches to selecting criteria by considering bases for justifying criteria such as exclusion and marginalization. The fourth contribution invites incorporation of a witness role that specifically addresses the values, interests, and worldviews of groups affected by an intervention into

132

existing roles for evaluators and stakeholders in the valuing process. Being critically reflective in these ways can enhance the credibility, defensibility, and legitimacy of an evaluation.

Three potential hindrances to moving towards more critically reflective valuing in evaluation are worth considering. First, many evaluators work for specific evaluation commissioners, under the requirements of a contract that limits the scope of work, and in short time frames with scarce resources. For political and logistical reasons, commissioners may initially resist (or even not hire) evaluators who consider multiple ways to frame a policy or program or ensure that the perspectives, interests, and worldviews of groups potentially negatively affected are considered. However, commissioners should not diminish the independence and authority of evaluators. Most, if not all, commissioners are motivated by an ideal of social betterment; explaining how these processes can reveal unexpected consequences which may be harmful to those involved or affected in the short or long-term may help to overcome initial resistance. Furthermore, as CSH was presented here as primarily a critical orientation and way of thinking, supported by the use of practical heuristics, it does not add on considerable time to an evaluation or take additional resources. The more evaluators gain familiarity and experience with CSH, the more it can be integrated as routine within evaluation.

A second potential hindrance has to do with the lack of empirical evidence regarding whether the incorporation of CSH into an evaluation enhances the latter's credibility, defensibility, and legitimacy. There is need for empirical research investigating uses and influences of CSH in evaluations and that examines the benefits and limitations of CSH from multiple points of view including evaluators as well as commissioners, involved stakeholders, and other groups affected. Evaluators using CSH, whether informally as a heuristic or more

formally as a methodology, are encouraged to write up descriptive case studies of their evaluations as a way to begin building an evidence base.

A third potential hindrance has to do with the fact that attending to values is outside evaluators' expertise, which, in evaluation, is typically based on some combination of methodological training and professional practical experience conducting evaluations (Schwandt, 2015). The kind of critically reflective role for professionals advanced in CSH is outside most evaluators' training and the standards and competencies endorsed in the evaluation community (AEA, 2004). However, it should be within the bounds of evaluators' responsibilities as professionals; "as a professional one does what one does on behalf of others and thus involves making sense of what it means to serve the client, other stakeholders, and the public" (Schwandt, 2015, p. 140). CSH argues that reflecting on, making decisions about, and taking responsibility for whose and what interests are served in a social inquiry or intervention should be a core responsibility of all professionals (Ulrich, 2004; Levin-Rozalis, 2014).

C. West Churchman (1968) famously said, "a systems approach begins when first you see the world through the eyes of another" (p. 231). But, perhaps less well known, he also pointed out, "the systems approach goes on to discovering that every world-view is terribly restricted" (Churchman, 1970 quoted in Jackson, 1991, p. 137). While evaluators must continuously examine the value of policies and programs through different points of view, they must also continuously examine and justify the assumptions that influence evaluations in light of potential consequences.

REFERENCES

Ackoff, R. (2001). A brief guide to interactive planning and idealized design. Retrieved from: http://www.ida.liu.se/~steho87/und/htdd01/AckoffGuidetoIdealizedRedesign.pdf

Alkin, M. C., Vo, A. T., & Christie, C. a. (2012). The evaluator's role in valuing: Who and with whom. New Directions for Evaluation, 2012(133), 29-41.

American Evaluation Association (2004). American Evaluation Association Guiding Principles for Evaluators. Retrieved from: http://www.eval.org/p/cm/ld/fid=51

Boyd, a., Geerling, T., Gregory, W. J., Kagan, C., Midgley, G., Murray, P., & Walsh, M. P. (2006). Systemic evaluation: a participative, multi-method approach. Journal of the Operational Research Society, 58(10), 1306-1320.

Cabrera, D., Colosi, L., & Lobdell, C. (2008). Systems thinking. Evaluation and Program Planning, 31(3), 299-310.

Checkland, P. (1999). Soft Systems Methodology: A 30-year Retrospective. West Sussex, England: John Wiley & Sons, Ltd.

Chelimsky, E. (2014). Public-interest values and program sustainability: Some implications for evaluation practice. American Journal of Evaluation, 35(4), 527-542.

Chelimsky, E. (2012). Valuing, evaluation methods, and the politicization of the evaluation process. New Directions for Evaluation, 133, 77-84.

Churchman, C.W. (1979). The Systems Approach and Its Enemies. New York, New York: Basic Books, Inc., Publishers

Churchman, C. W. (1971). The Design of Inquiring Systems: Basic Concepts of Systems and Organizations. New York, New York: Basic Books, Inc., Publishers.

Churchman, C. W. (1968). The Systems Approach. New York, New York: Basic Books, Inc., Publishers

Dahler-Larson, P., & Schwandt, T. (2012). Political culture as context for evaluation. New Directions for Evaluation, 135, 75-87.

Davidson, J., Wehipeihana, N. & McKegg, K. (2011). The Rubric Revolution. Retrieved from: http://betterevaluation.org/resources/guides/rubric_revolution

Davidson, J. (2005). Evaluation Methodology Basics: The Nuts and Bolts of Sound Evaluation. Thousand Oaks, CA: SAGE Publications, Inc.

Datta, L. E. (2011). Politics and evaluation: more than methodology. American Journal of Evaluation, 32(2), 273-294.

Greene, J. (2011). Values-engaged Evaluations. In M. Segone (Ed.), (pp. 192-206): UNICEF.

Greene, J. C. (2006). Making the world a better place through evaluation. In M. C. Alkin (Ed.), Evaluation Roots: A Wider Perspective of Theorists' Views and Influences. Thousand Oaks, CA: SAGE Publications, Inc.

Greene, J. C., Destefano, L., Burgon, H., & Hall, J. (2006). An educative, values-engaged approach to evaluating stem educational programs. New Directions for Evaluation(109), 53-71.

Greene, J. C. (1997). Evaluation as advocacy. American Journal of Evaluation, 18(1).

Gregory, A. (1997). Evaluation practice and the tricky issue of coercive contexts. Systems Practice, 10(5), 589-609.

Gregory, A., & Jackson, M. (1992a). Evaluating organizations: a systems and contingency approach. Systems Practice, 5(1), 37-60.

Gregory, A., & Jackson, M. C. (1992b). Evaluation methodologies: a system for use. Journal of the Operational Society, 43(1), 19-28.

Hall, J. N., Ahn, J., & Greene, J. C. (2011). Values engagement in evaluation: Ideas, illustrations, and implications. American Journal of Evaluation, 33(2), 195-207.

Henry, G. T. (2002). Choosing criteria to judge program success: a values inquiry. Evaluation, 8(2), 182-204.

Hood, S. Hopson, R., and Kirkhart, K. (2015). Culturally responsive evaluation: theory, practice, and future implications. In Newcomer, Hatry, and Wholey (Eds.), Handbook of Practical Program Evaluation. San Francisco, CA: Jossey-Bass.

House, E. R. (2006). Work memoir - ideas and influences. In M. C. Alkin (Ed.), Evaluation Roots: A Wider Perspective of Theorists' Views and Influences. Thousand Oaks, CA SAGE Publications, Inc.

House, E. R. (2001). Unfinished business: Causes and values. American Journal of Evaluation, 22(3), 309-315.

House, E., & Howe, K. (1999). Values in Evaluation and Social Research. Thousand Oaks, CA: Sage Publications, Inc.

Hummelbrunner, R., & Reynolds, M. (2013). Systems thinking, learning and values in evaluation (pp. 9-10): European Evaluation Society.

Jackson, M. (1991). Systems Methodology for the Management Sciences. New York, New York: Plenum Press.

Julnes, G. (2012a). Managing valuation. New Directions for Evaluation, 2012(133), 3-15. doi: 10.1002/ev.20002

Julnes, G. (2012b). Developing policies to support valuing in the public interest. New Directions for Evaluation, 133, 109-129.

Kant, I., & Smith, N. K. (1929). Immanuel Kant's Critique of Pure Reason. Boston: Bedford.

Kaplan, J. (2014). DAC Criteria. Retrieved from: http://betterevaluation.org/evaluation-options/dac_criteria

Levin-Rozalis, M. (2014). Let's Talk Program Evaluation in Theory and Practice Tel Aviv, Israel Dekel Academic Press.

Mabry, L. (2010). Critical social theory evaluation: Slaying the dragon. In M. Freeman (Ed.) Critical social theory and evaluation practice. New Directions for Evaluation, 127, 83-98.

Mark, M. M., Henry, G. T., & Julnes, G. (2000). Evaluation: An Integrated Framework for Understanding, Guiding, and Improving Public and Nonprofit Policies and Programs. San Francisco, CA: Jossey-Bass.

Mertens, D. M. (2006). Social transformation and evaluation. In M. C. Alkin (Ed.), Evaluation Roots: A Wider Perspective of Theorists' Views and Influences. Thousand Oaks, CA: SAGE Publications, Inc.

Midgley, G. (2007). Systems thinking for evaluation. In B. Williams & I. Imam (Eds.), Systems Concepts in Evaluation: An Expert Anthology (pp. 11–34). Point Reyes, CA: EdgePress of Inverness.

Midgley, G. (2000). Systemic Intervention: Philosophy, Methodology, and Practice. New York, New York: Kluwer Academic/Plenum Publishers.

Midgley, G. (1996). Evaluating services for people with disabilities: A critical systems perspective. Evaluation, 2(1), 67-84.

Midgley, G. (1992). The sacred and profane in critical systems thinking. Systems Practice, 5(1), 5-16.

Morris, M. (2012). Valuation and the american evaluation association: Helping 100 flowers bloom, or at least be understood? New Directions for Evaluation, 133, 85-90.

Oakden, J. (2013). Evaluation rubrics: how to ensure transparent and clear assessment that respects diverse lines of evidence. Retrieved from: http://betterevaluation.org/sites/default/files/Evaluation%20rubrics.pdf

Patton, M. Q. (2012). Contextual pragmatics of valuing. Promoting valuation in the public interest: Informing policies for judging value in evaluation. New Directions for Evaluation, (133), 97-108.

Ramage, M., & Shipp, K. (2009). C. West Churchman. In Systems Thinkers (pp. 131–138). United Kingdom: Springer London.

Reynolds, M. (2015a). (Breaking) The Iron Triangle of Evaluation. IDS Bulletin (Vol. 46): Institute of Development Studies.

Reynolds, M. (2015b). Rigour (-mortis) in Evaluation. Evaluation Connections, 2-4.

Reynolds, M. (2014). Equity-focused developmental evaluation using critical systems thinking. Evaluation, 20(1), 75-95.

Reynolds, M., Forss, K., Hummelbrunner, R., Marra, M., & Perrin, B. (2012). Complexity, systems thinking and evaluation - an emerging relationship? Evaluation Connections: Newsletter of the European Evaluation Society, 7-9.

Reyonlds, M., & Williams, B. (2011). Systems Thinking and Equity-focused Evaluation (pp. 115-141): UNICEF

Reynolds, M. (2010). Evaluation and stakeholding development. Paper presented at the The 9th European Evaluation Society International Conference, Prague, Czech Republic.

Reynolds, M. (2008). Response to paper "Systems thinking" by D. Cabrera et al.: systems thinking from a critical systems perspective. Evaluation and Program Planning, 31(3), 323-325.

Reynolds, M. (2007). Framing purposeful evaluation through critical systems thinking. Rural Development for the XXIInd Congress of the European Society for Rural Sociology.

Reynolds, M. (2007). Evaluation Based on Critical Systems Heuristics. In B. Williams & I. Imam (Eds.), (pp. 101-122). Point Reyes, CA: EdgePress of Inverness.

Schwandt, T. A. (2015). Evaluation Foundations Revisited: Cultivating a Life of the Mind for Practice. Stanford, CA: Stanford University Press

Schwandt, T. a. (2008). Educating for intelligent belief in evaluation. American Journal of Evaluation, 29(2), 139-150. doi: 10.1177/1098214008316889

Schwandt, T. A. (2005). Value judgment & values. In S. Mathison (Ed.), Encyclopedia of Evaluation Thousand Oaks, CA: SAGE Publications, Inc.

Schwandt, T. A. (2003). In search of the political morality of evaluation practice. Studies in Educational Policy and Educational Philosophy, 1-6.

Schwandt, T. A. (1997). The landscape of values in evaluation: Charted terrain and unexplored territory. New Directions for Evaluation, 3(76).

Scriven, M. (2012). The logic of valuing. New Directions for Evaluation, 133, 17-28.

Scriven, M. (2006). Conceptual revolutions in evaluation: past, present, and future. In M. C. Alkin (Ed.), Evaluation Roots: A Wider Perspective of Theorists' Views and Influences. Thousand Oaks, CA: SAGE Publications, Inc.

Scriven, M. (2005). Metaevaluation. In In S. Mathison (Ed.), Encylopedia of Evaluation Thousand Oaks, CA: SAGE Publications, Inc.

Scriven, M. (1994). The final synthesis. Evaluation Practice, 15(3), 367-382.

Scriven, M. (1991). Evaluation Thesaurus. Newbury Park, CA: SAGE Publications, Inc.

Shadish, W. R., Cook, T. D., & Leviton, L. C. (1991). Foundations of Program Evaluation: Theories of Practice. Newbury Park, CA: Sage Publications, Inc.

Shipman, S. (2012). The role of context in valuing federal programs. New Directions for Evaluation, 133, 53-63.

Stake, R. E. (2006). Responsive evaluation IV. In M. C. Alkin (Ed.), Evaluation Roots: A Wider Perspective of Theorists' Views and Influences. Thousand Oaks, CA: SAGE Publications, Inc.

Stake, R. E., & Schwandt, T. A. (2006). On discerning quality in evaluation. In I. F. Shaw, J. C. Greene, & M. M. Mark (Eds.), (pp. 404-418). London: SAGE Publications Ltd.

Stake, R. E. (2004). Standards-Based & Responsive Evaluation Thousand Oaks, CA: Sage Publications, Inc.

Stake, R., Migotsky, C., Davis, R., Cisneros, E. J., Depaul, G., Dunbar, C., . . . Chaves, I. (1997). The Evolving Syntheses of Program Value Evaluation Practice, 18(2), 89-103.

Ulrich, W. (2012). CST's Two Ways: A Concise Account of Critical Systems Thinking. Retrieved from http://wulrich.com/downloads.html

Ulrich, W. (2011a). What is Good Professional Practice?: Part 1. Werner Ulrich's Webpage. Retrieved from http://wulrich.com/downloads.html

Ulrich, W. (2011b). What is Good Professional Practice?: Part 2. Werner Ulrich's Webpage. Retrieved from http://wulrich.com/downloads.html

Ulrich, W., & Reynolds, M. (2010). Critical Systems Heuristics. In M. Reynolds & S. Holwell (Eds.), Systems Approaches to Managing Change: A Practical Guide (pp. 243-292). Milton Keynes, U.K.: Springer

Ulrich, W. (2009). An Appreciation of C. W. Churchman. Werner Ulrich's Homepage. Retrieved from http://wulrich.com/downloads.html

Ulrich, W. (2005). A Brief Introduction to Critical Systems Heuristics (CSH). Retrieved from http://wulrich.com/downloads.html

Ulrich, W. (2004). A Brief Introduction to "Critical Systems Thinking for Professionals & CItizens". Werner Ulrich's Home Page. Retrieved from http://wulrich.com/downloads.html

Ulrich, W. (2002a). Boundary Critique. In H. G. Daellenbach & R. L. Flood (Eds.), The Informed Student Guide to Management Science. London: Thomson Learning.

Ulrich, W. (2002b). Critical Systems Heuristics. In H. G. Daellenbach & R. L. Flood (Eds.), The Informed Student Guide to Management Science (p. 72). London: Thomson Learning.

Ulrich, W. (2000). Reflective practice in the civil society: The contribution of critically systemic thinking. Reflective Practice, 1(2), 247–268.

Ulrich, W. (1988). Churchman's "Process of unfolding" and its significance for policy analysis and evaluation. Systems Practice, 1(4), 415-428.

Ulrich, W. (1983). Critical Heuristics of Social Planning: A New Approach to Practical Philosophy. Switzerland: Swiss National Science Foundation.

Weiss, C. (1998). Evaluation (Second ed.). Upper Saddle River, New Jersey: Prentice Hall.

Williams, B. (2015). Prosaic or Profound? The Adoption of Systems Ideas by Impact Evaluation IDS Bulletin (Vol. 46). Oxford, UK: Institute of Development Studies.

CONCLUSION

Numerous prominent evaluators contend that systems and complexity ideas and techniques are transforming evaluation practice and, moreover, that the development of the evaluation field should continue to turn to the systems and complexity fields (Williams, 2015; Stern et. al, 2015; Patton, 2016). Other evaluators have and continue to question the significance of this turn and whether much of what appears to be new has already been integrated into good evaluation practice (Datta, 2008; Rogers, 2008). This thesis conceptually and empirically examined the central question of whether, how, and to what extent the systems and complexity fields are indeed making a dramatic change in how evaluators think about and practice evaluation. This project, somewhat unexpectedly, supports both of these views – that the systems and complexity fields have transformative implications for evaluation practice and that some of these implications have been integrated into existing conceptions of and ways of practicing evaluation such that transformations are not readily apparent.

On one hand, there are some significant ways in which systems and complexity ideas are challenging evaluators to re-think long-standing and widespread assumptions and envision, create, and carry out new ways of practicing evaluation. The first paper identifies six activities of evaluation practice (i.e., social problem solving, framing an intervention and its context, selecting and using methods, engaging in valuing, producing and justifying knowledge, and facilitating use) that are being challenged and transformed in rather significant ways. The third paper draws on a relatively overlooked area of systems thinking, critical systems heuristics, and argues that it both challenges assumptions about valuing and enhances how evaluators can explicitly acknowledge the influence of values on evaluations.

On the other hand, the process of selecting, translating, and applying these ideas and techniques occurs within evaluators' existing conceptions of evaluation practice and is heavily

influenced by the circumstances in which they work. The second paper did not reveal wholesale transformations in evaluation practice. Evaluators spoke about ways systems and complexity ideas and techniques were shaping their evaluation practices, but many of these ways varied by the unique ideas and techniques they drew on, how they interpreted and applied these, and the circumstances in which they work. Moreover, the same ideas and techniques used by the same evaluators often had different implications in different instances of evaluation practice. That said, there were some common shifts in ways of thinking about and practicing evaluation: re-structuring the evaluation itself as an intervention; re-figuring relationships with commissioners; re-defining the object of evaluation; re-purposing and expanding methods; revisiting and problematizing the making of value judgments; and renewing an emphasis on instrumental use. Further research is needed to examine whether these implications extend to other cases of evaluation practice.

These papers can be connected in several ways to provide future directions for evaluation practice and for research on evaluation. One major consideration has to do with the role of evaluation practice in social problem solving. This includes factors such as the institutional contexts and governance structures within which evaluations are conducted, and, more specifically, the procedures for evaluation contracting and structuring evaluator-commissioner relationships. Papers 1 and 2 suggest that these factors can significantly constrain as well as provide opportunities for the use of systems and complexity ideas and techniques. Moreover, these factors play a role in the extent to which evaluators can aspire for and expect new kinds of evaluation use (e.g., feedback loop, social learning, normative critique, ongoing instrumental use). The evaluation field may want to facilitate reflection and discussion about influencing some of these circumstances, for example, through evaluation policy. Additionally, there is a need for

research investigating how these factors influence, for better or worse, the use of systems and complexity ideas and techniques in evaluation practice.

Another major consideration has to do with the kinds of knowledge and skills evaluators need to use systems and complexity ideas and techniques and what training opportunities are needed to support their development. Papers 1, 2, and 3 highlighted the definition and framing of the objects of evaluation as one way these ideas and techniques can be used. This includes the use of particular systems and complexity concepts and theory-informed lenses for defining the objects of evaluation and attention on boundary judgments and different perspectives to consider multiple framings. Such skills are not usually taught in evaluation curricula. These three papers also emphasized the breadth of systems and complexity methods, underlying assumptions, and purposes for which they can be used. Evaluators are not typically trained in such methods and, further, using methods rigorously, appropriately, and creatively requires more substantial learning opportunities than professional development workshops.

Valuing was examined in all three papers, yet is perhaps least conclusive. While papers 1 and 3 identified new ways for making value judgments by comparing descriptive and normative models, paper 2 found that evaluators using systems and complexity ideas struggled with valuing in their practices. Further research on valuing using systems and complexity techniques is needed. Two focuses for this research include developing a theoretical and practical basis for this model of valuing and better understanding why evaluators using systems and complexity ideas and techniques struggle with valuing and emerging ways in which they are addressing this challenge.

REFERENCES

Datta, L. E. (2008). Response to paper "Systems thinking" by D. Cabrera et al.: systems thinking: an evaluation practitioner's perspective. *Evaluation and Program Planning, 31*(3), 321-322.

Patton, M.Q. (2016, forthcoming). What is essential in developmental evaluation? On integrity, fidelity, adultery, abstinence, impotence, long-term commitment, and sensitivity in implementing evaluation models. American Journal of Evaluation

Rogers, P. (2008). Response to paper "Systems thinking" by D. Cabrera et al.: Is it systems thinking or just good practice in evaluation? *Evaluation Program Planning, 31*(3), 325-326.

Stern, E., Saunders, M., & Stame, N. (2015). Standing back and looking forward: Editors' reflections on the 20th Anniversary of Evaluation. *Evaluation, 21*(4), 380-390.

Williams, B. (2015). Prosaic or Profound? The Adoption of Systems Ideas by Impact Evaluation *IDS Bulletin* (Vol. 46). Oxford, UK: Institute of Development Studies.

APPENDICES

Appendix A: Interview Guide

Introduction to Study and Review Consent Information

<u>*PART 1*</u>: *I'd like to start by talking about* <u>*your background:*</u> *what brought you to evaluation and then to systems thinking and complexity science.*

1. Could you tell me about your background as an evaluator?

   a. What is your disciplinary background and training?
   b. What first brought you to conduct evaluations?
   c. What do you typically evaluate? In what arena of evaluation practice (e.g., public health, international development, social services)?

Now, I'd like to hear about your interest and involvement in systems thinking and complexity science.

2. What were the circumstances surrounding your initial interest in systems thinking and complexity science?

   a. What sparked your interest?
   b. What connections did you make or want to make between systems thinking and complexity science and evaluation?

3. Which systems and complexity ideas, theories, methods, or approaches are you most interested in with regards to evaluation?

<u>*PART 2*</u>: *Next, I'd like to focus on* <u>*particular evaluation case(s)*</u> *in which you drew on or applied systems thinking and/or complexity science in some way. I'd like to focus our conversation on _____ case, if that is okay with you. If examples from other evaluations you work on come to mind, please feel free to discuss those as well. [*Request written materials about other cases*]*

4. Could you briefly describe the case(s) in which you used systems thinking and/or complexity science?

   a. What was being evaluated?
   b. What circumstances or purposes led you to use systems thinking and/or complexity science in this case?

5. Now, I'd like to hear the story of how you used systems thinking and/or complexity science in this evaluation. In particular, what did you use and for what tasks or activities of the evaluation?

   a. Which aspects of systems thinking and complexity science did you use?
   b. For which evaluation tasks and how?

       i. Defining the evaluand
      ii. Planning the evaluation
    iii. Collecting and analyzing data
    iv. Judging value
     v. Reporting and supporting use of findings

6. How successful or unsuccessful would you say using systems thinking and/or complexity science in this case was?

   a. What, if any, challenges did you face?
   b. What would you say was the greatest value of using these ideas for the evaluation?

PART 3: Now I'd like to step back from this particular case and talk more generally about <u>your reflections and views</u> on the interest in the evaluation field of interpreting and applying concepts, theories, methods and approaches from systems thinking and complexity science.

7. How would you characterize the rather recent interest in using systems thinking and complexity science in the evaluation field?

   a. What would you say is driving this interest in systems and complexity amongst evaluators?

8. What is influencing this interest most and how so?

   a. What is impeding the use of systems and complexity in evaluation?
   b. What is helping the use of systems and complexity in evaluation?

9. What excites you most about this interest?

10. What concerns you most about this interest?

11. What would you like to see next with regards to this the interest in systems thinking and complexity science in the evaluation field?

   a. Research on evaluation?
   b. Evaluator professional development?

Part 4: We are nearing the end of our interview. I would like to invite you to add any additional thoughts you might have.

12. Is there anything you would like to add or elaborate on to what you have said during this interview?

Closing Thoughts

Appendix B: Case Materials

Table B.1. Case Materials Listed By Evaluator Name

Evaluator	Materials
Glenda Eoyang	Eoyang, G. H. (2007). Human Systems Dynamics: Complexity-based Approach to a Complex Evaluation. In I. Williams, Bob, Imam (Ed.), *Systems Concepts in Evaluation: An Expert Anthology* (pp. 123–140). Point Reyes, CA: EdgePress of Inverness. Eoyang, G. & Holladay, R.J. (2013). Adaptive Action: Leveraging Uncertainty in Your Organization. Stanford, CA: Stanford University Press. Eoyang, G. H., & Berkas, T. H. (1999). Evaluating Performance in a Complex Adaptive System. In M. Lissack & H. Gunz (Eds.), (pp. 1-21). Westport, Connecticut: Quorum Books. Eoyang, G. & Yellowthunder, L. (2007). Dynamic Evaluation Report. Interprofessional Education and Practice: Creating an Interprofessional Learning Environment through Communities of Practice: An Alternative to Traditional Preceptorship. Circle Pines, MN: Human Systems Dynamics Institute.
Margaret Hargreaves	Hargreaves, M., Cole, R., Coffee-Borden, B., Paulsell, D., & Boller, K. (2013). Evaluating Infrastructure Development in Complex Home Visiting Systems. *American Journal of Evaluation*, *34*(2), 147–169. Hargreaves, M. (2014). Rapid evaluation approaches for complex initiatives. Mathematica Policy Research. Hargreaves, M. B. (2010). Evaluating System Change: A Planning Guide (pp. 1-20). Princeton New Jersey.
Richard Hummelbrunner	Hummelbrunner, R. (2007). Systemic Evaluation in the Field of Regional Development. In Williams, B., Imam, I. (Eds.), *Systems Concepts in Evaluation: An Expert Anthology* (pp. 161–180). Point Reyes, CA: EdgePress of Inverness. Hummelbrunner, R. (2008). Response to paper "systems thinking" by D. Cabrera et al.: a tool for implementing DSRP in programme evaluation. *Evaluation and Program Planning, 31*(3), 331-333. Hummelbrunner, R. (2011). Applying systems thinking to evaluation of structural funds. Evaluating the effects of regional interventions: A look beyond current structural funds' practices. Warsaw: Ministry of Regional Development.

Evaluator	Materials
Richard Hummelbrunner (continued)	Hummelbrunner, R. (2011). Systems thinking and evaluation. *Evaluation, 17*(4), 395-403. Hummelbrunner, R. & Reynolds, M. (2013). Systems thinking, learning and values in evaluation. *Evaluation Connections: Newsletter of the European Evaluation Society*, 9-10. Williams, B. & Hummelbrunner, R. (2011). Using Systems Concepts in Action: A Practitioner's Toolkit. Stanford, CA: Stanford University Press. Hummelbrunner, R. (2015). Learning, Systems Concepts, and Values in Evaluation: Proposal for an Exploratory Framework to Improve Coherence (Vol. 46): Institute of Development Studies. Hummelbrunner, R., Williams, B., Rogers, P., Fujita, N., & Takahashi, E. (2010). *Beyond Logframe: Using Systems Concepts in Evaluation*. Issues and Prospects of Evaluation for International Development - Series IV. Tokyo, Japan.
Doug Luke	National Cancer Institute (2007). *Greater Than the Sum: Systems Thinking in Tobacco Control.* Tobacco Control Monograph No. 18. Bethesda, MD: U.S. Department of Health and Human Services, National Institutes of Health, National Cancer Institute. NIH Pub. No. 06-6085. Luke, D. A. & Stamatakis, K.A. (2012). Systems Science Methods in Public Health: Dynamics, Networks, and Agents. Annual Review of Public Health, 33, 357-376. Wallace, R., Geller, A., & Ogawa, V.A. (Eds.) (2015). Assessing the use of agent-based models for tobacco regulation. The National Academies Press. Retrieved from: http://www.nap.edu/catalog.php?record_id=19018 Luke, D., Harris, J.K, Shelton, S., Allen, P., Carothers, B.J., & Mueller, N. B. (2010). Systems Analysis of Collaboration in 5 National Tobacco Control Networks. American Journal of Public Health. 100(7), 1290-1297.
Gerald Midgley	Midgley, G. (2000). *Systemic Intervention: Philosophy, Methodology, and Practice*. New York, New York: Plenum Publishers. Boyd, a, Geerling, T., Gregory, W. J., Kagan, C., Midgley, G., Murray, P., & Walsh, M. P. (2006). Systemic evaluation: a participative, multi-method approach. *Journal of the Operational Research Society*, *58*(10), 1306–1320. Midgley, G. (1996). Evaluating Services for People with Disabilities: A Critical Systems Perspective. *Evaluation, 2*(1), 67-84.

Evaluator	Materials
Robin Lin Miller	Miller, R. L., Levine, R. L., McNall, M.A., Khamarko, K, Valenti, M.T. (2009). A Dynamic Model of Client Recruitment and Retention in Community-Based HIV Prevention Programs. *Health Promotion Practice.* Hirsch, G. B., Levine, R., & Miller, R. L. (2007). Using system dynamics modeling to understand the impact of social change initiatives. *American journal of community psychology, 39*(3-4), 239-253. Miller, R.L. (2015). Community psychology, evaluation, and social critique. American Journal of Evaluation. 36, 89-99. Trickett, E.J. Et. Al (2011). Advancing the science of community-level interventions. *American Journal of Public Health*. 101(8), 1410-1419.
Michael Quinn Patton	Patton, M.Q. (2011). Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use. New York, NY: The Guilford Press. Patton, M. Q. (2015). Systems Theory and Complexity Theory *Qualitative Research and Evaluation Methods* (pp. 139-151). Thousand Oaks, CA: SAGE Publications, Inc. Homeless Youth Collaborative on Developmental Evaluation (2014). Nine evidence-based guiding principles to help youth overcome homelessness. Otto Bremer Foundation. Gamble, J. A. (2008). A Developmental Evaluation Primer. The J.W. McConnell Family Foundation Patton, M.Q. (2015). Global Systems Evaluation Course: International Program in Development Evaluation Training (IPDET). Retrieved from: **http://www.ipdet.org/** Patton, M. Q. (2016). What is essential in developmental evaluation? On integrity, fidelity, adultery, abstinence, impotence, long-term commitment, integrity, and sensitivity in implementing evaluation models. American Journal of Evaluation. Patton, M.Q. (1994). Developmental Evaluation. *Evaluation Practice*. 15(3), 311-219.

Evaluator	Materials
Deborah Wasserman	Wasserman, D. L. (2008). A response to paper "Systems thinking" by D. Cabrera et al.: next steps, a human service program system exemplar. *Evaluation and Program Planning, 31*(3), 327-329. Wasserman, D. L. (2010). Using a systems orientation and foundational theory to enhance theory-driven human service program evaluations. *Evaluation and Program Planning*, *33*(2), 67–80. Cabrera, D., Colosi, L., & Lobdell, C. (2008). Systems thinking. *Evaluation and Program Planning*, *31*(3), 299–310 Cabrera, D., & Colosi, L. (2008). Distinctions, systems, relationships, and perspectives (DSRP): a theory of thinking and of things. *Evaluation and Program Planning, 31*(3), 311-317. Wasserman, D.L. & Emery, J. R. (2013). Valuing Kujichagulia: A Self-Determination Theory-Based Approach to Evaluating Africentric Rites of Passage Programming. *Black Child Journal Special Edition: Rites of Passage Foundations and Practices. 24-59.*