EVALUATION

café Feed Your Mind.

The Evaluation Center
Western Michigan
University

The Organizational and Political Landscape for Evidence-Informed Decision Making in Government

Trachtenberg School
of Public Policy
& Public Administration
THE GEORGE WASHINGTON UNIVERSITY

Kathryn Newcomer

November 30, 2016

1

"Evidence-based Policy," "Data-Driven Decision-making"– the New Normal?

Questions to Address Today

- When and why did the "evidence-based policy" imperative become so prevalent in the public and nonprofit sectors?

- How can evaluators help government decision-makers use evidence to **inform** decision-making?

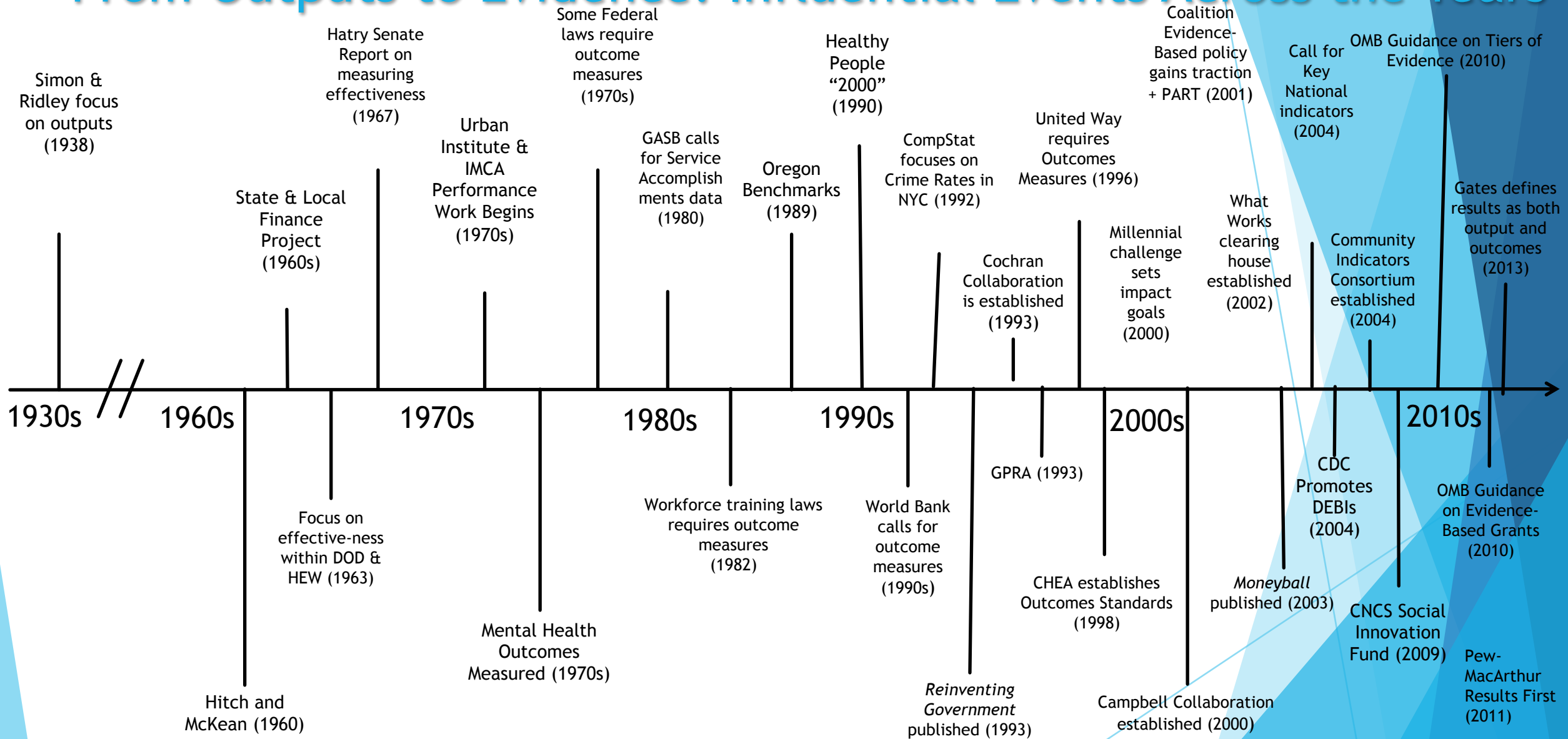- How can we move from generating data for accountability to learning ?

"Evidence-based Policy"

▶ The evaluation illuminary affecting governmental decision-makers, foundations, nonprofit boards, intermediaries and --- evaluation practice!

▶ Myth or reality?

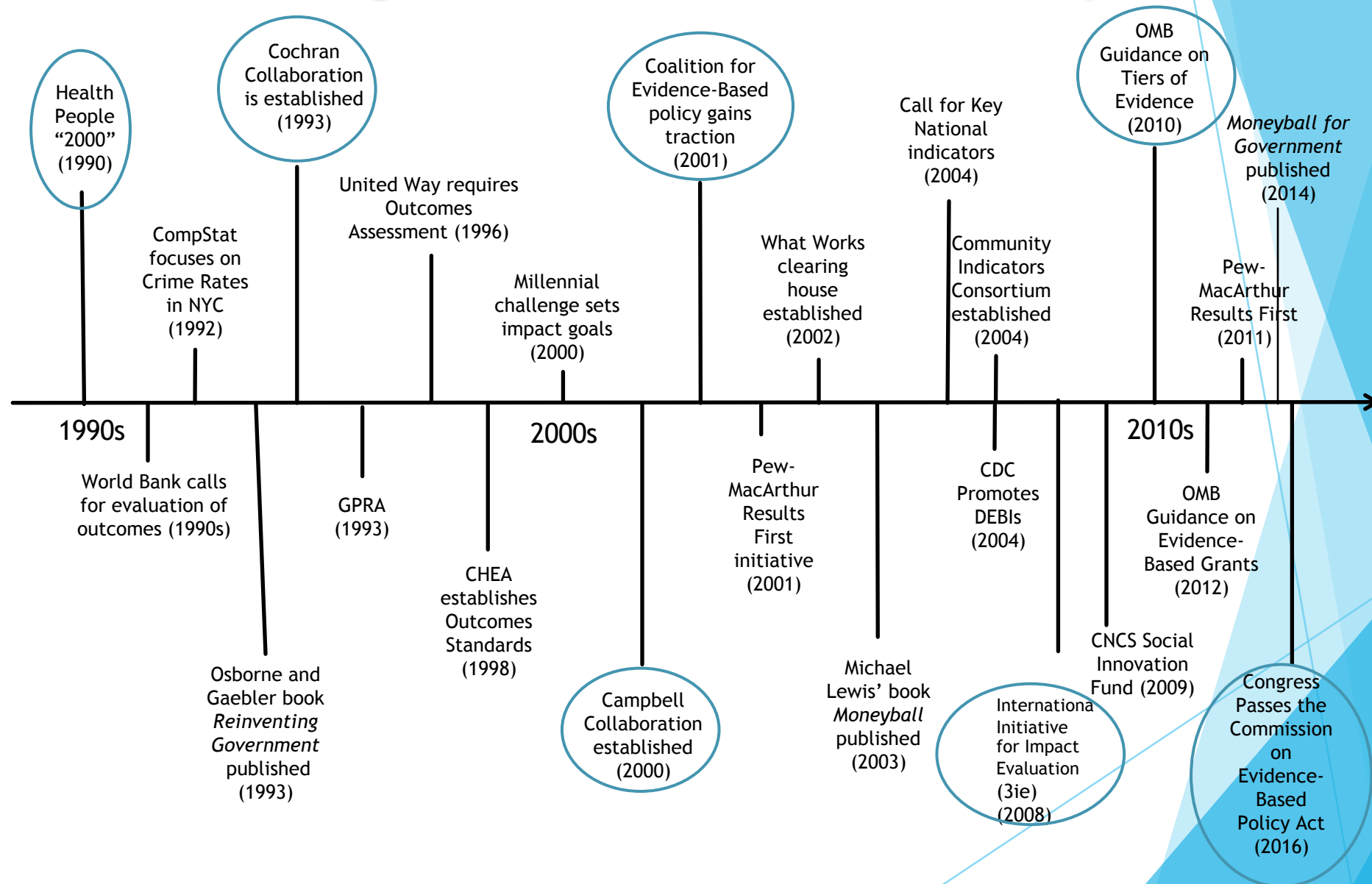▶ Advantages and disadvantages for decision-makers and for evaluators?

Efforts in Several Arenas Have Moved the Dialogue to Embrace "Evidence–Based Policy"

- Since the 1960s in the U.S. dialogue about the target for government's efforts has changed from a focus on effectiveness to outcomes to results to evidence ---- Why?

  - Tracking of diseases in Public Health, e.g., "Healthy People 2000," and The Cochran Collaboration

  - North American and European Social Scientists Established the Campbell Collaboration to mimic The Cochran Collaboration

  - The U.S. Office of Management and Budget's efforts to Assess "evidence of program effectiveness" in both the George W. Bush and Barrack Obama Administrations were highly influenced by the Coalition for Evidence-based Policy

  - Leading Foundations have invested resources to encourage evidence-based decision-making, e.g., Pew, MacArthur, Arnold, and Grant

  - Evaluation and Monitoring of International Development Efforts

5

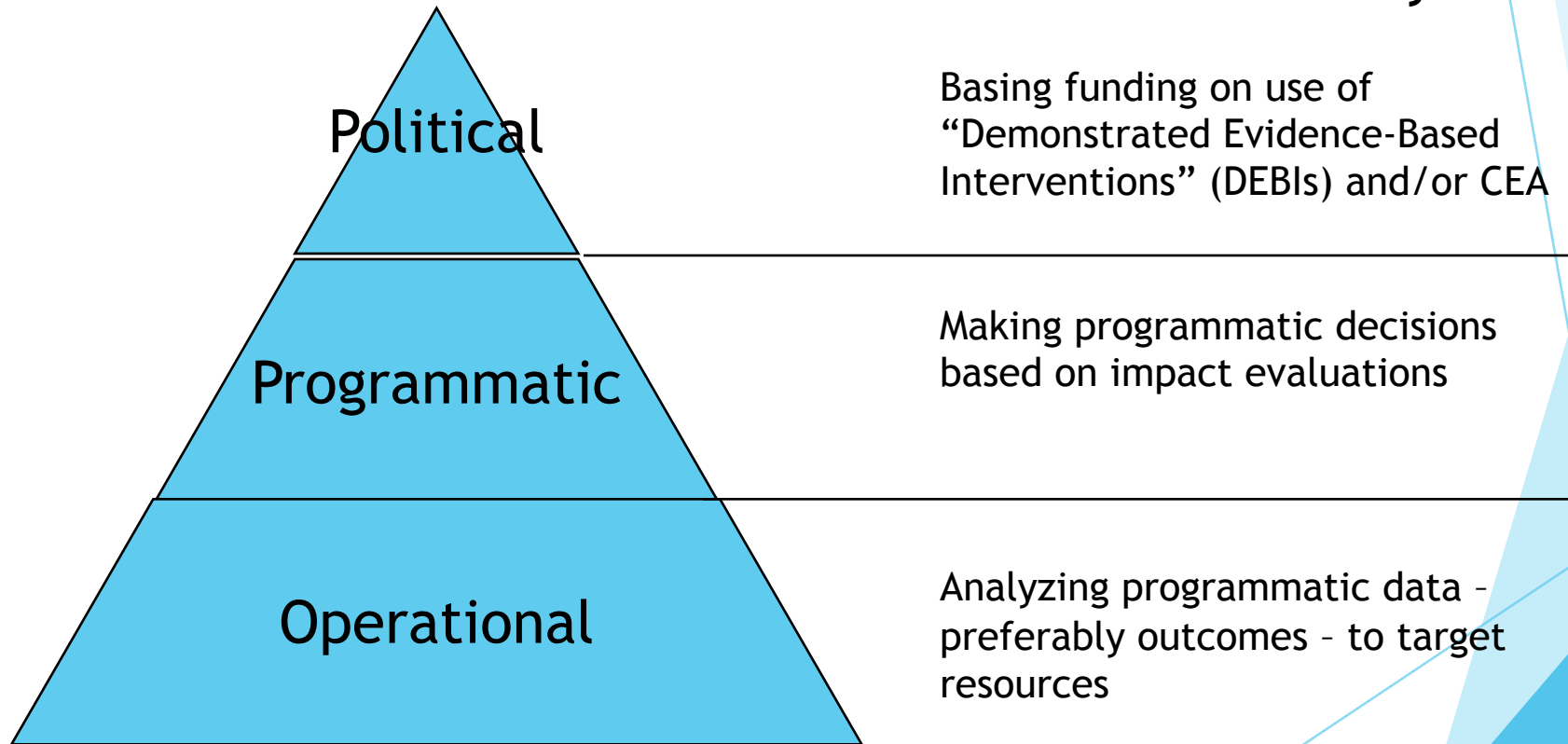# From Outputs to Evidence: Influential Events Across the Years

- Simon & Ridley focus on outputs (1938)
- Hatry Senate Report on measuring effectiveness (1967)
- Some Federal laws require outcome measures (1970s)
- Healthy People "2000" (1990)
- Coalition Evidence-Based policy gains traction + PART (2001)
- Call for Key National indicators (2004)
- OMB Guidance on Tiers of Evidence (2010)
- State & Local Finance Project (1960s)
- Urban Institute & IMCA Performance Work Begins (1970s)
- GASB calls for Service Accomplishments data (1980)
- Oregon Benchmarks (1989)
- CompStat focuses on Crime Rates in NYC (1992)
- United Way requires Outcomes Measures (1996)
- What Works clearing house established (2002)
- Community Indicators Consortium established (2004)
- Gates defines results as both output and outcomes (2013)
- Cochran Collaboration is established (1993)
- Millennial challenge sets impact goals (2000)
- Focus on effective-ness within DOD & HEW (1963)
- 1930s
- 1960s
- 1970s
- 1980s
- 1990s
- 2000s
- 2010s
- GPRA (1993)
- CDC Promotes DEBIs (2004)
- Hitch and McKean (1960)
- Workforce training laws requires outcome measures (1982)
- World Bank calls for outcome measures (1990s)
- CHEA establishes Outcomes Standards (1998)
- *Moneyball* published (2003)
- OMB Guidance on Evidence-Based Grants (2010)
- Mental Health Outcomes Measured (1970s)
- CNCS Social Innovation Fund (2009)
- *Reinventing Government* published (1993)
- Campbell Collaboration established (2000)
- Pew-MacArthur Results First (2011)

Embracing Evidence-Based Policy

Evidence-Based Policy – Made by Whom?

Decisions to be Informed by Evidence

Political

Basing funding on use of "Demonstrated Evidence-Based Interventions" (DEBIs) and/or CEA

Programmatic

Making programmatic decisions based on impact evaluations

Operational

Analyzing programmatic data – preferably outcomes – to target resources

8

Contrasting Views on Evidence-Based Policy

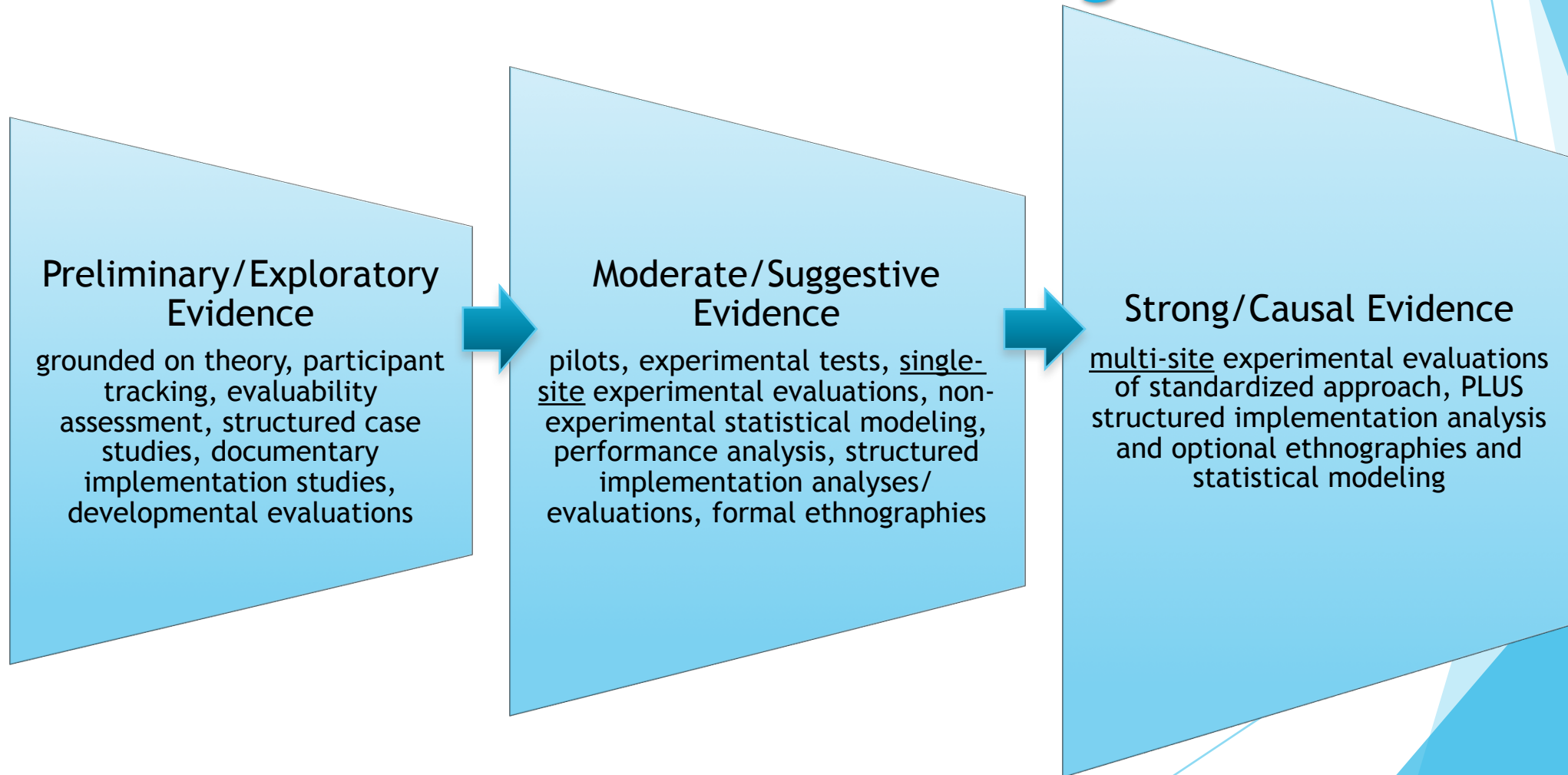| Fixed Mindset	Growth Mindset
1. We need to collect data to test if programs work or do not work.	1. We need to learn which program mechanisms work for whom, where and under what circumstances.
2. Policy should be made at the top and based on evidence.	2. Policy is "made" through implementation processes at multiple levels by multiple actors with different types of data available to them.
3. Program impact can be measured precisely.	3. Measuring program impact is difficult as programs and intended impactees change and evolve.
4. Random Control Trials (RCTs) are the gold standard for research and evaluation design.	4. Research designs must be matched to answer the question raised; RCTs are appropriate for certain impact questions.
5. Proven program models can be replicated in multiple locations as long as they are implemented with fidelity to the original design.	5. Program mechanisms may be replicated in multiple locations as long as they are adapted to meet local conditions.
6. Benefit-cost analysis should be used to compare social programs.	6. Benefit-cost analysis is difficult to use to compare social programs given the challenge of costing out benefits , especially those accruing over time.

Note: I expanded upon the notion of mindset in *Mindset* by Carol Dweck.

Obama Administration: Explicit Emphasis on Producing and Acting on Evidence

- A series of office Memoranda from OMB between 2009 and 2013 signaled that performance measurement and evaluation were to be used to produce "evidence on what works"

- Starting in 2015 OMB Circular A-11 defines evidence for the federal government:

  - **"For purpose of A-11 Part 6, evidence is the available body of facts or information indicating whether a belief or proposition is true or valid. Evidence can be quantitative or qualitative and may come from a variety of sources, including performance measurement, evaluations, statistical series, retrospective reviews and other data analytics and research. Evidence has varying degrees of credibility, and the strongest evidence generally comes from a portfolio of high-quality evidence rather than a single study."**

The Obama Administration Recognizes Tiers of Evidence Based on Perceived Rigour

Preliminary/Exploratory Evidence

grounded on theory, participant tracking, evaluability assessment, structured case studies, documentary implementation studies, developmental evaluations

Moderate/Suggestive Evidence

pilots, experimental tests, single-site experimental evaluations, non-experimental statistical modeling, performance analysis, structured implementation analyses/ evaluations, formal ethnographies

Strong/Causal Evidence

multi-site experimental evaluations of standardized approach, PLUS structured implementation analysis and optional ethnographies and statistical modeling

What are Challenges for Evidence to Inform Policymaking?

▶ Expectations regarding:

  ▶ What constitutes evidence?

  ▶ How transferable is evidence?

  ▶ When and where do we underestimate the role played by the "impactees?"

  ▶ Where is the capacity to support both the demand and supply of evidence?

We Overstate the Certainty of the Evidence we Can Collect

> Perceptions of the certainty of "evidence" have changed.

What are the Opportunities for Evidence to Inform Decision-making?

- ▶ Analyses of "performance" data collected by agencies (or delegated service delivery agents such as grantees)

- ▶ Implementation, Outcome and Impact evaluations typically performed by other agents for government

- ▶ Manipulations of services in experiments by agencies – "behavioral economics"

- ▶ Syntheses or systematic reviews of impact evaluations by external agents, e.g. websites like "What Works"

14

Why isn't There Agreement About the Quality of Evidence?

- Differing professional standards and "rules" or criteria for evidence, e.g., lawyers, accountants, engineers, economists

- Disagreements about methodologies within professional groups, e.g., RCTs

- The constancy of change in problems and the characteristics of the targeted impactees

15

"Evidence-Based" Grant Making

- Grants comprise over $600 billion in the US federal Budget

- OMB started urging agencies to use evidence based grant making starting in 2010 but with little guidance

- Where are we now?

To what extent is there consensus on what constitutes evidence in the grants environment?

	To a great extent/ A lot	A moderate amount	A little/ Not at all	Number of Respondents
Within your Agency	50%	30%	20%	132
With your legislative branch	29%	29%	32%	113
With other funders in your field	30%	31%	39%	112
With academia	34%	29%	39%	98
Within your grantee network	51%	33%	27%	113

Source: Dawes and Newcomer Survey, November 2016)

We Underestimate the Evolving Sources of Complexity Affecting the Production of Relevant Evidence

▶ Change in the nature of problems to be addressed by government, e.g., the nature of natural security threats, the use of the internet in crime

▶ Change in the context in which programs and policies are implemented, e.g., increasingly complicated service delivery networks, PPPs

▶ Changing priorities of political leaders – and under Trump?
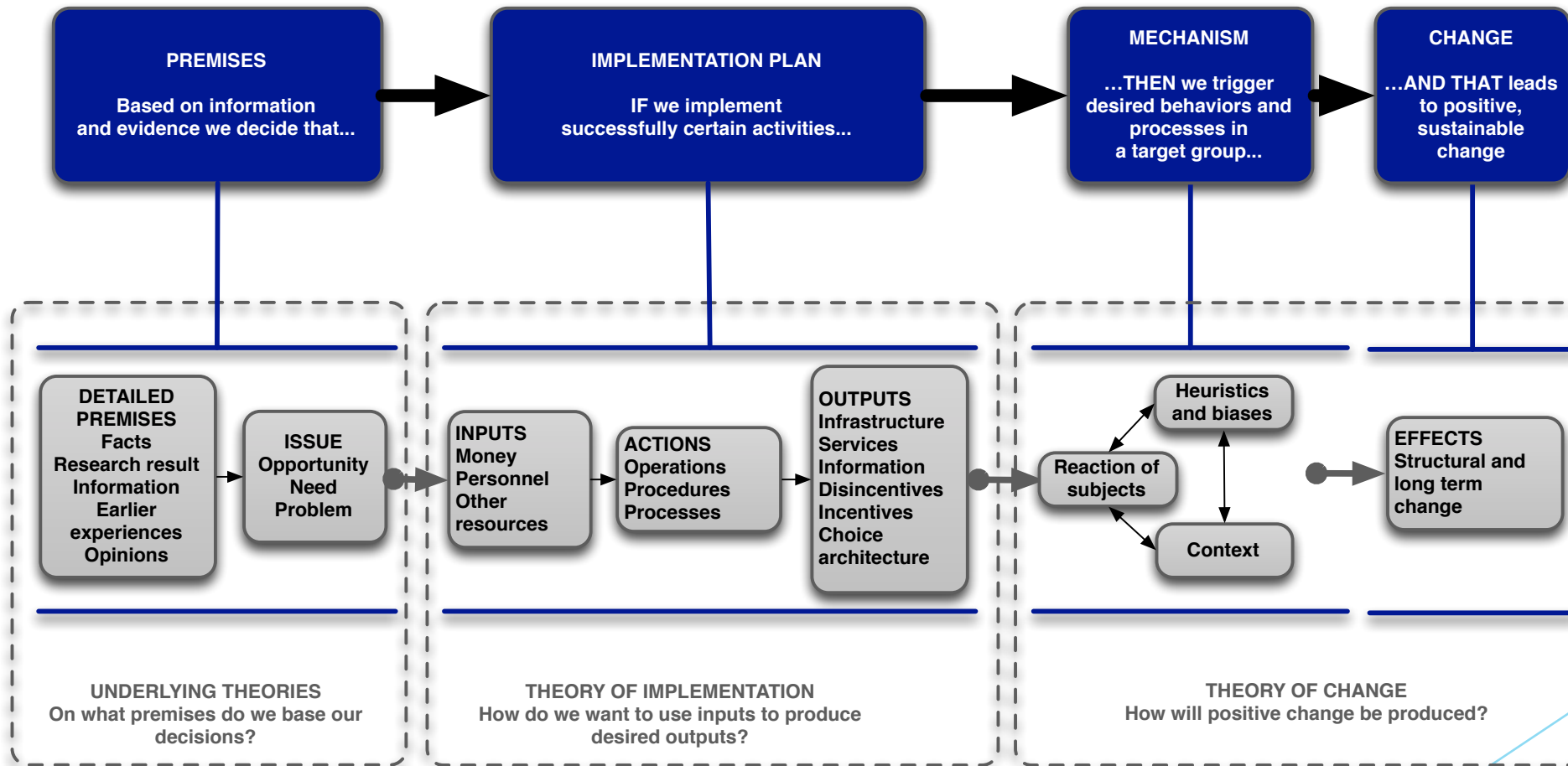
We Overstate the Ease of Flow of Evidence
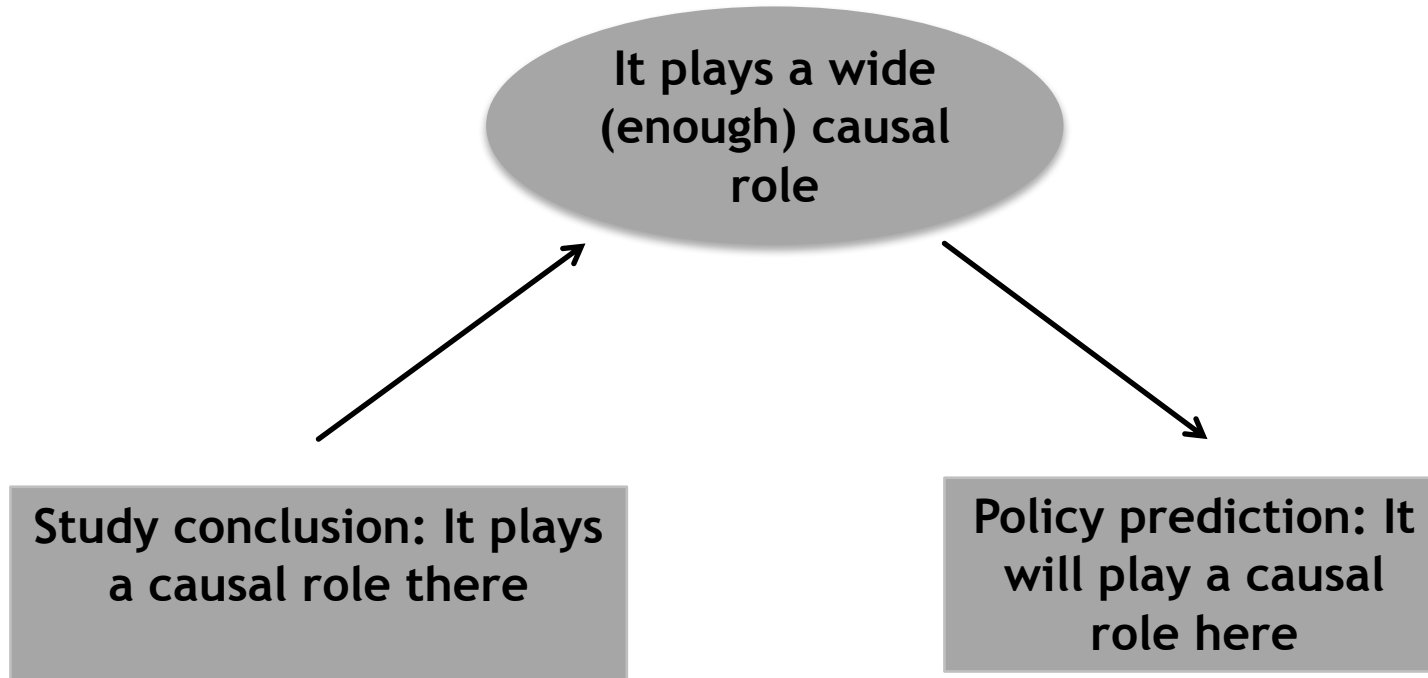
It plays a wide (enough) causal role

Study conclusion: It plays a causal role there

Policy prediction: It will play a causal role here

Source: Cartwright, N. (2013). Knowing what we are talking about: why evidence doesn't always travel. *Evidence & Policy: A Journal of Research, Debate and Practice*, 9(1), 97-112.
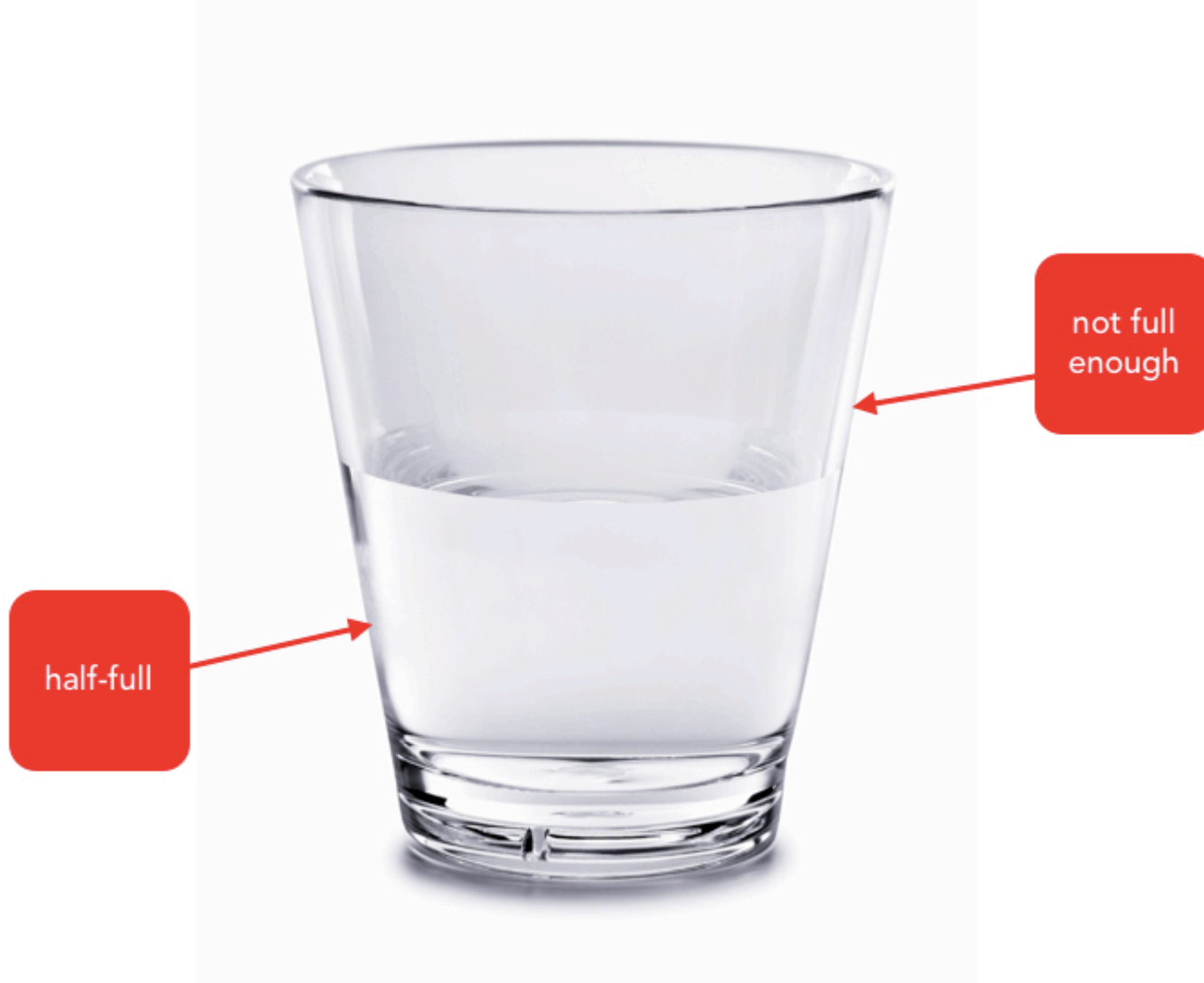
What is needed for a well-supported effectiveness (impact) prediction?

It will play a positive causal role here

The support factors for it are w, y, z	We have w, y, z here	It can play a positive causal role here

?	?	?	?	?	?	?	It can play the same role as there	It plays a positive causal role there (and there)
							?	STUDY	RCT	RCT

Source: Cartwright, N. (2013). Knowing what we are talking about: why evidence doesn't always travel. *Evidence & Policy: A Journal of Research, Debate and Practice*, 9(1), 97-112.

We Underestimate the Role of Volition Among Impactees and their Own Heuristics

Olejniczak, K. & Newcomer, K. (2014). "Moving towards accountability for learning";
in: Olejniczak, K. & Mazur, S. (eds.) Organizational Learning. A Framework for Public Administration, p.81-99. Warsaw: Scholar Publishing House.

We Overstate The Current Evaluation Capacity in Government

not full enough

half-full

Evaluation Capacity =
Both Demand and Supply

- Who is asking for the evidence?

- How do perceptions of the potential for political use of the evidence affect program managers?

- How clear is the understanding between providers and requestors on what sort of data (evidence) is needed?

- Are there sufficient resources within agencies to respond to demand?

- What about the lack of interaction and synergies among the different potential providers of evidence - such as at the U.S. federal level GPRA/GPRAMA reporting staff, internal evaluation staff, external evaluation contractors, SBST, data.gov teams, etc.!

Currently Multiple Groups Undertake Monitoring and Evaluation in and for U.S. Government Agencies

Monitoring
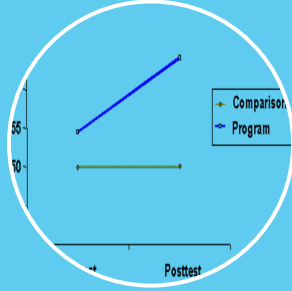
Impact Evaluation

Behavioral Economics

They tend to operate in separate and even uncommunicative units with competing priorities!

There are Signs of Progress in Improving the Quality of Evidence in Government

- Reputable Drivers are putting resources into efforts, e.g.:
  - The Pew MacArthur Results First Initiative (their state rating will be released soon)
  - The National Academy of Sciences, Engineering, and Medicine May 2016 report: *Advancing the Power of Economic Evidence to Inform Investments in Children, Youth, and Families*
  - The Arnold and William Grant Foundations
- Recent National Science Foundation support of initiatives to help policy researchers translate their findings for government users
- Professional Associations are supporting translational efforts, e.g., APPAM, AEA
- Communities of Practice abound, especially in public health
- The Commission on Evidence–Based Policy seems to be inclusive in terms of considering what constitutes evidence

25

Promising Practices from the Obama Administration

Promising Practice	Affects Supply or Demand?	Needed Support Factors
Knowledge Brokers, e.g., the Chief Evaluation Officers	Both	Brokers have technical expertise, interpersonal skills, and contextual wisdom
Learning Agendas	Demand	Strong leadership backing and encouragement to be innovative
Quarterly Reviews	Supply	Credible data, stress on learning, no punitive actions
Strategic Reviews	Both	Encouragement to be innovative, stress on learning not accountability

The Immediate Political Context for Evidence–Based Policy Making in the U.S.

- Program managers and other decision makers are caught between two masters- The President and Congress - and these entities are likely to have different priorities and values, as are the two major political parties regarding the use of evidence

- The implementation of virtually all federal programs and policies is undertaken through states, local governments, nonprofits and even private agents

- Federalism affects the flow of money to implement federal policies and programs –for example, formula grants given to states are hard to change into evidence-based grants

- The President relies on his or her Office of Management and Budget to "drive management reforms" and it is hard to not have the rest of government view these directives as compliance exercises

How and When do Decision-makers learn from Evidence?

Information Processing

-Automatic Operations ("Fast Thinking")

-Controlled Operations (Slow)
- Worldview & Epistemology
- Expertise

-Judgmental Heuristics

-Emotional State

-Presentism
- Pure rate of time preference

Transmission Process

-Brokering/delivering the information

-Priming (timing matters!)

-Timeliness of Access

-Presentation of Data
- Logic visualization
- Data visualization

Nature of Information

-Source of Evidence
- Presumed credibility
- Reputed credibility

-Trustworthiness of Evidence
- Weight of evidence
- Strength of evidence
- Reliability of data

-Match Between Evidence and Receiver's Epistemological Preferences

-Signaling about Priorities in Research Designs Rigor from Respected Sources

Organization and Social Context

-Organizational Culture
- Leadership modeling of use of evidence
- Priority given data in decision making
- Focus on learning
- Support for risk taking
- Treatment of "errors"

-Social Supports
- Similarity in worldviews within group
- "Like-minded" peers
- Priority given to diversity of views

How Can Evaluators Contribute to Helping Decision-makers learn from Evidence?

Information Processing

-Automatic Operations ("Fast Thinking")

-Controlled Operations (Slow)
- Worldview & Epistemology
- Expertise

-Judgmental Heuristics

-Emotional State

-Presentism
- Pure rate of time preference

Transmission Process

-Brokering/delivering the information

-Priming (timing matters!)

-Timeliness of Access

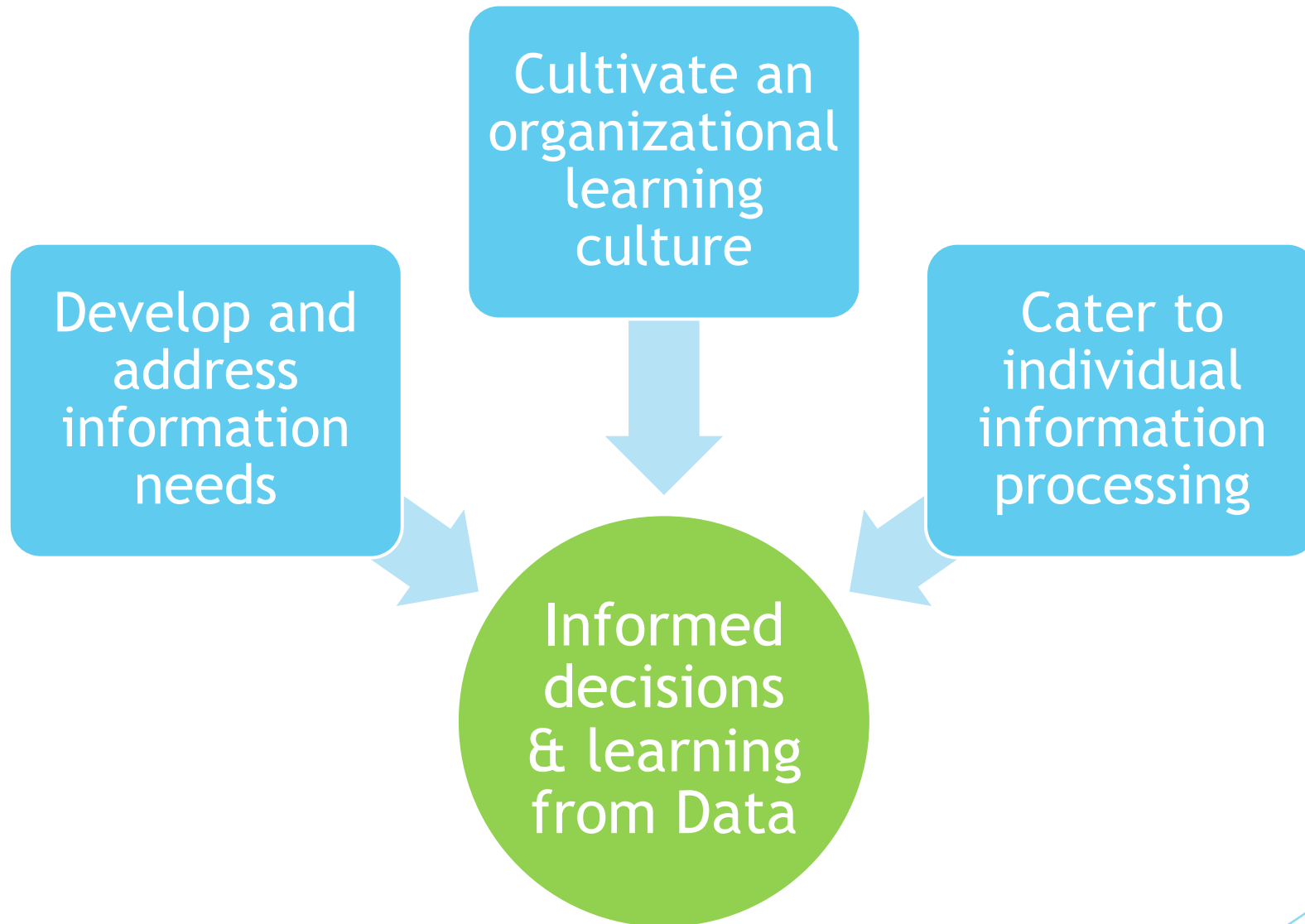-Presentation of Data
- Logic visualization
- Data visualization

Nature of Information

-Source of Evidence
- Presumed credibility
- Reputed credibility

-Trustworthiness of Evidence
- Weight of evidence
- Strength of evidence
- Reliability of data

-Match Between Evidence and Receiver's Epistemological Preferences

-Signaling about Priorities in Research Designs Rigor from Respected Sources

Organization and Social Context

-Organizational Culture
- Leadership modeling of use of evidence
- Priority given data in decision making
- Focus on learning
- Support for risk taking
- Treatment of "errors"

-Social Supports
- Similarity in worldviews within group
- "Like-minded" peers
- Priority given to diversity of views

Transmission Process

- Just as a there are many producers, there are many potential users of the evidence provided, e.g., different policy designer and implementers in complex service delivery and regulatory networks units

- **Knowledge brokering** is critical

- Understanding and strengthening the linkage between the producers of evaluative data and the many potential users of that information requires time and resources

- For Example: the network of 57 evaluation brokering units in Poland overseeing 900 evaluations of EU cohesion policy investments

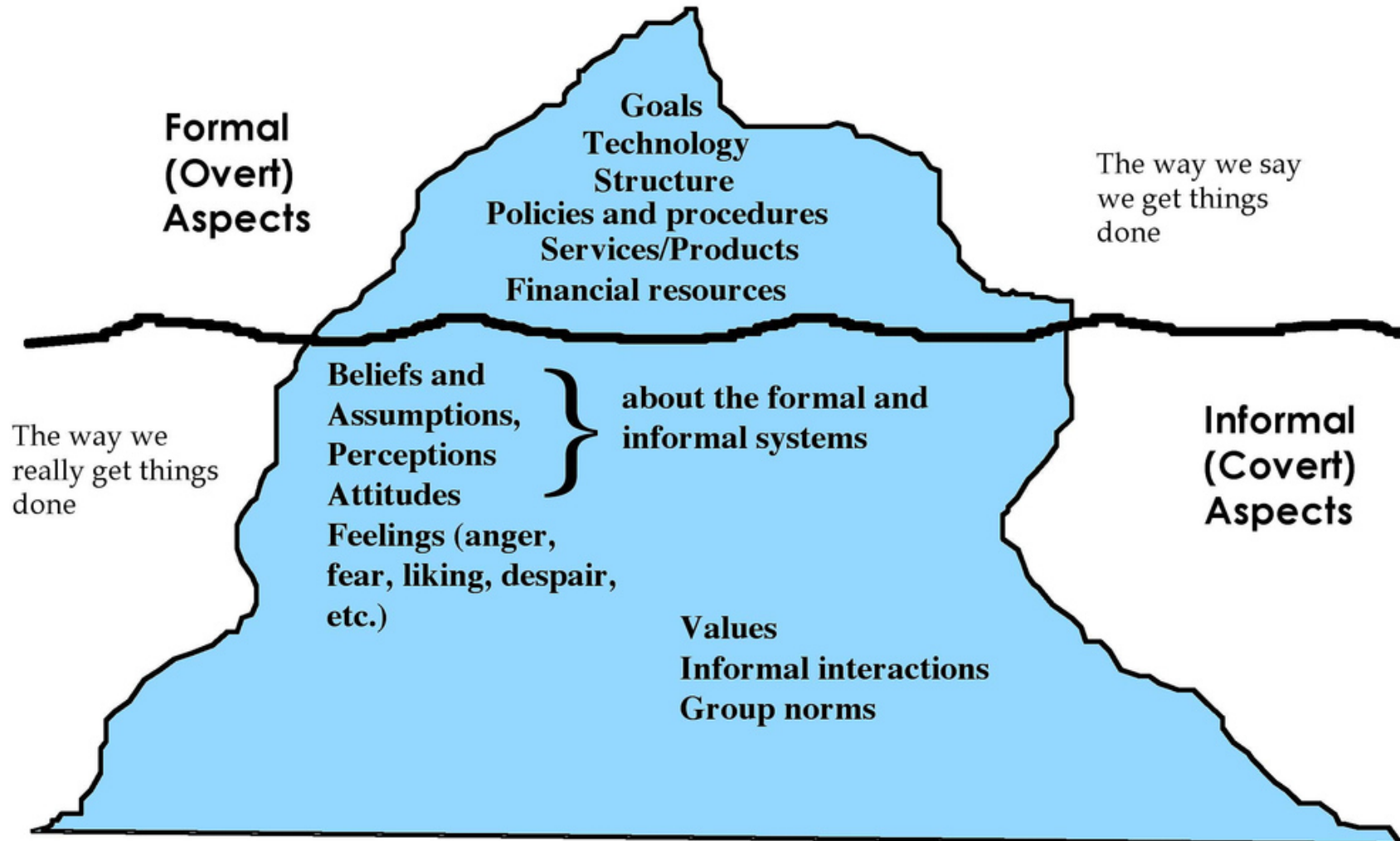# A simple framework….

Cultivate an organizational learning culture

Develop and address information needs

Cater to individual information processing

Informed decisions & learning from Data

Remember Evaluation Capacity = Both Demand and Supply

▶ Consider who is asking for the data/evidence and **who** might use the information provided and **how and when** they may use it

▶ Probe the extent to which there is a clear understanding between providers and requestors for what sorts of evidence is needed, e.g., brokering

▶ Assess whether or not sufficient resources are available to meet demand

▶ Address the lack of interaction and facilitate synergies among the different potential providers of evidence - such as monitoring and reporting staff, internal evaluation staff, external evaluation contractors, etc.

32

Organizational Culture is Difficult to Change

Formal (Overt) Aspects

The way we say we get things done

Goals
Technology
Structure
Policies and procedures
Services/Products
Financial resources

The way we really get things done

Informal (Covert) Aspects

Beliefs and Assumptions, Perceptions Attitudes Feelings (anger, fear, liking, despair, etc.)

} about the formal and informal systems

Values
Informal interactions
Group norms

Created by Stanley N. Herman. TRW Systems Group, 1970

33

What are Evaluation-Receptive Organizational Cultures?

▶ Engage in **self-reflection & self-examination**

   ▶ Deliberately **seek evidence** on what it's doing

   ▶ **Use results information** to challenge or support what it's doing

   ▶ Promote **candor**, challenge and **genuine dialogue**

▶ Engage in **evidence-based learning**

   ▶ Make **time** to learn

   ▶ **Learn from mistakes** and failures

   ▶ Encourage **knowledge sharing**

▶ Encourage **experimentation and change**

   ▶ Support deliberate **risk-taking**

   ▶ **Seek out new ways** of doing business
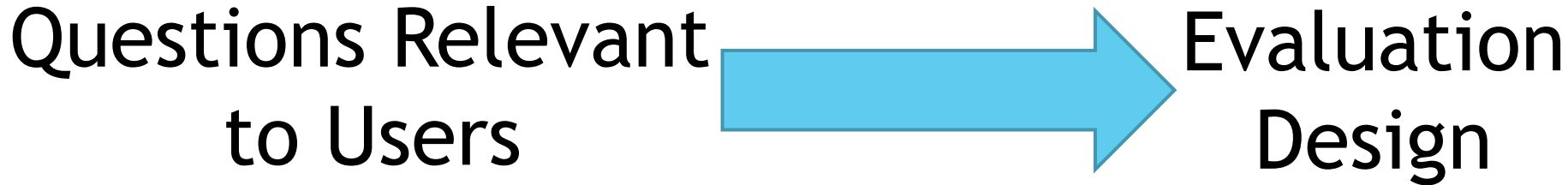
(See John Mayne, 2010)

34

Be Strategic and Intentional about Cultivating Evaluation-Receptive Cultures

- ▶ Assess and address the factors perpetuating a compliance mentality among potential users, especially clients
- ▶ Reward learning from monitoring and evaluation, e.g., Learning Audits in the Netherlands
- ▶ Cultivate capacity to support both the demand and supply of information, e.g., the Canadian approach
- ▶ Match evaluation approaches to questions appropriately and transparently
- ▶ Reward mixed methods approaches that integrate data collected via differing methods

Move To Strategic and Synergistic Use of Evaluation!

Monitoring

Behavioral Economics

Evaluation

Impact Evaluation

Implementation Evaluation
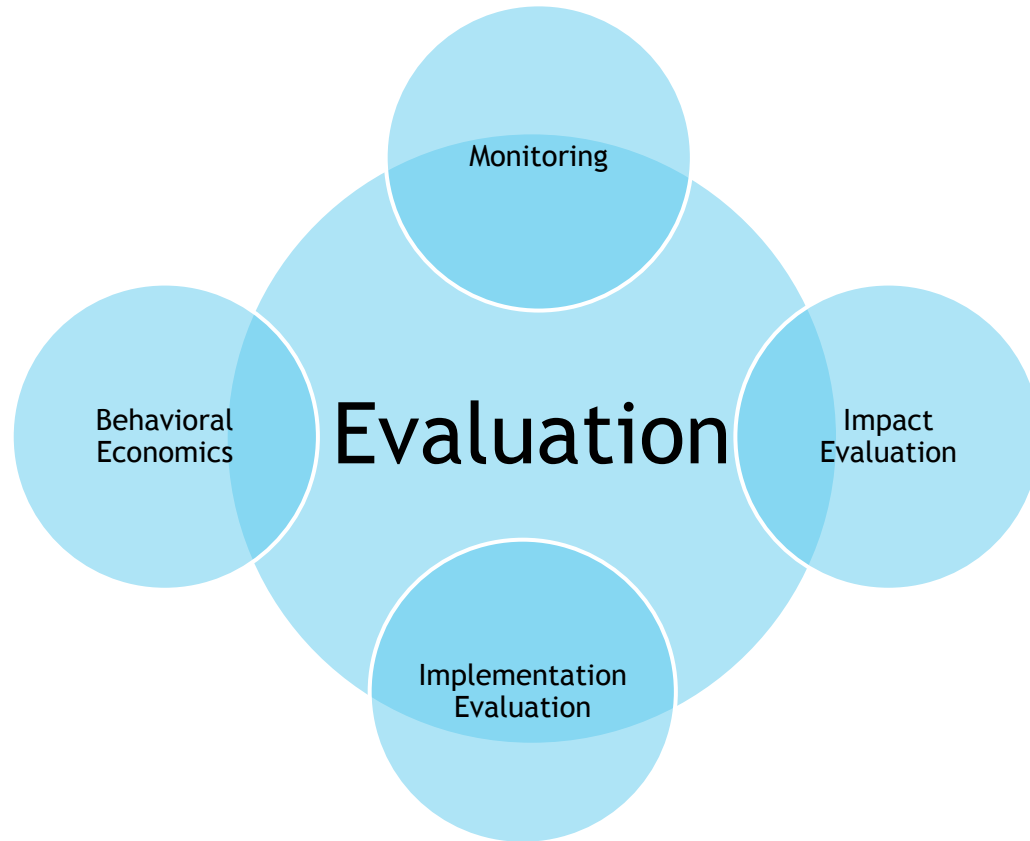
Help Information Users Frame Pertinent Questions and then Match the Questions with the Appropriate Evaluation Approach

Questions Relevant to Users ⟶ Evaluation Design

Match Evaluation Approach to Questions

Objective	Illustrative Questions	Possible Design
#1: Describe program activities	• How extensive and costly are the program activities? • How do implementation efforts vary across sites, beneficiaries, regions? • Has the program been implemented sufficiently to be evaluated?	• Monitoring • Exploratory Evaluations • Evaluability Assessments • Multiple Case Studies
#2: Probe targeting & implementation	• How closely are the protocols implemented with fidelity to the original design? • What key contextual factors are likely to affect achievement of intended outcomes? • How do contextual constraints affect the implementation of a intervention? • How does a new intervention interact with other potential solutions to recognized problems?	• Multiple Case Studies • Implementation or Process evaluations • Performance Audits • Compliance Audits • Problem-Driven Iterative Adaptation
#3: Measure the impact of policies & programs	• What are the average effects across different implementations of the intervention? • Has implementation of the program or policy produced results consistent with its design (espoused purpose)? • Is the implementation strategy more (or less) effective in relation to its costs?	• Experimental Designs/RCTs • Non-experimental Designs: Difference-in-difference, Propensity score matching, etc. • Cost-effectiveness & Benefit Cost Analysis • Systematic Reviews & Meta-Analyses
#4 : Explain how/ why programs & policies produce (un)intended effects	• How/why did the program have the intended effects? • To what extent has implementation of the program had important unanticipated negative spillover effects? • How likely is it that the program will have similar effects in other communities or in the future?	• Impact Pathways and Process tracing • System dynamics • Configurational analysis,

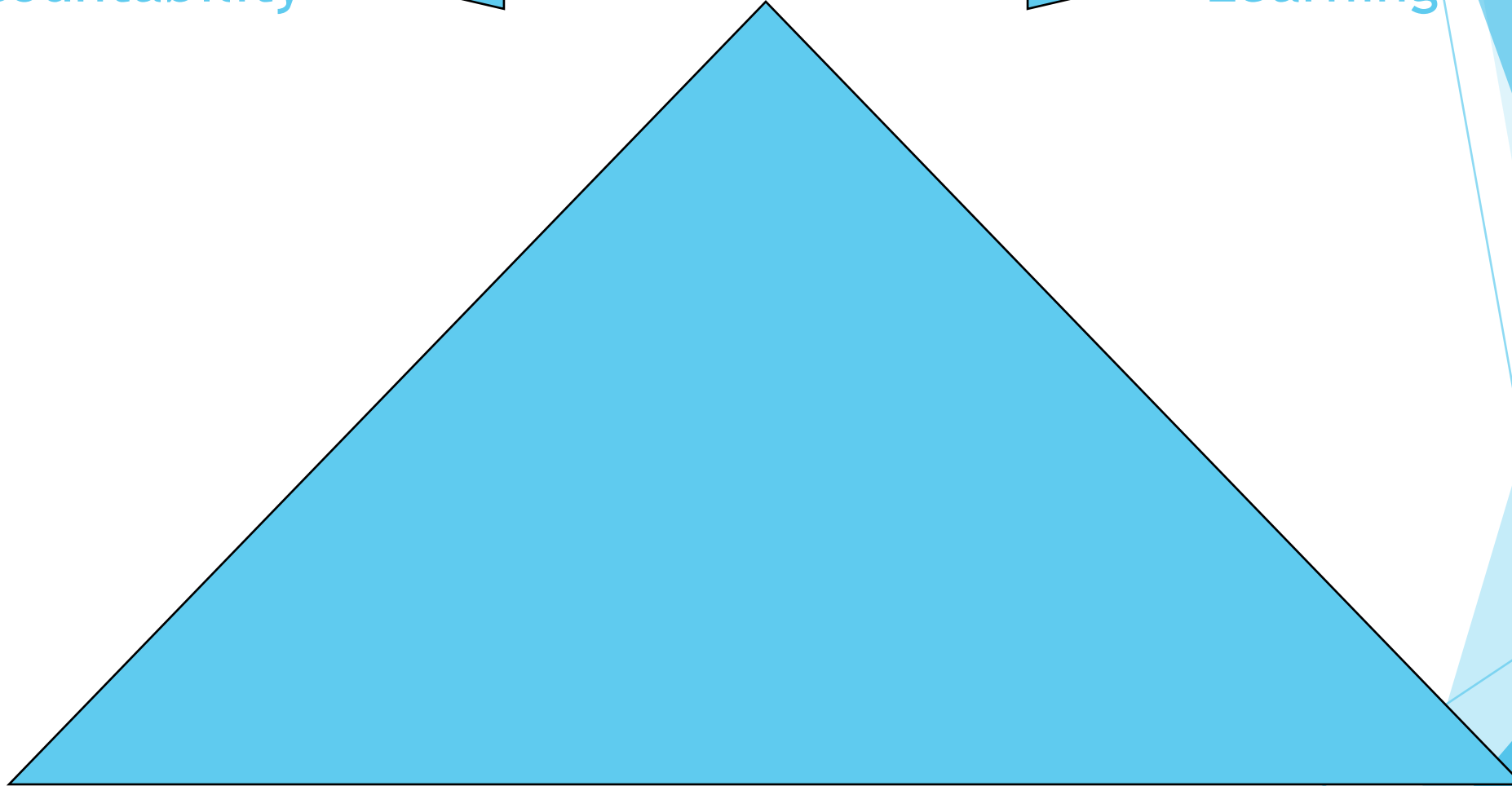# How do we balance accountability with learning from evaluation?

- ▶ Very, very carefully!!
- ▶ Signaling matters!
- ▶ Funders' reporting requirements matter!

A Delicate Balancing Act

Accountability ←——————————→ Learning

There is an ongoing tension between producing evidence to demonstrate accountability versus to promote learning.

Knowledge Generation and Use

Strategic Knowledge

Accounting for
Achieving Impact

Learning What
Works and Why

Accountability

Learning

Accounting for
Financial Compliance

Learning how to
Operate Efficiently

41

Operational Knowledge

Knowledge Generation and Use

Strategic Knowledge

Accounting for
Achieving Impact

Learning What
Works and Why

Accountability ⟵————————⟶ Learning

Priority Number 1!

Accounting for
Financial Compliance

Learning how to
Operate Efficiently

Operational Knowledge

42

Knowledge Generation and Use

Strategic Knowledge

Accounting for
Achieving Impact

Learning What
Works and Why

Accountability ← → Learning

Accounting for
Financial Compliance

Learning how to
Operate Efficiently

Operational Knowledge

43

Relevant References

▶ Dahler-Larsen, Peter. 2012. *The Evaluation Society*. Stanford University Press.

▶ Donaldson, S., C. Christie, and M. Mark (editors) 2015. *Credible and Actionable Evidence*, 2nd Edition. Sage.

▶ Head, B. 2015. "Toward More "Evidence-Informed" Policy Making?" *Public Administration Review*. Vol.76, Issue 3, pp. 472-484.

▶ Kahneman, D. 2011. *Thinking, Fast and Slow*. Farrar, Straus and Giroux Publishers.

▶ Mayne, J. 2010. "Building an evaluative culture: The key to effective evaluation and results management." *Canadian Journal of Program Evaluation*, 24(2), 1-30.

▶ Newcomer, K. and C. Brass. 2016. "Forging a Strategic and Comprehensive Approach to Evaluation within Public and Nonprofit Organizations: Integrating Measurement and Analytics within Evaluation." *American Journal of Evaluation*, Vol. 37 (1), 80-99.

▶ Olejniczak, K., E. Raimondo, and T. Kupiec. 2016. "Evaluation units as knowledge brokers: Testing and calibrating an innovative framework." *Evaluation*, Volume 22 (2)., 168-189.

▶ Sunstein. C. and R. Hastie. 2015. *Wiser: Getting Beyond Groupthink to Make Groups Smarter*. Harvard Business Review Press.

▶ World Bank Group. *Mind, Society and Behavior*. 2015.

Thank You!
Questions?

I can be reached at newcomer@gwu.edu

Trachtenberg School
of Public Policy
& Public Administration

THE GEORGE WASHINGTON UNIVERSITY