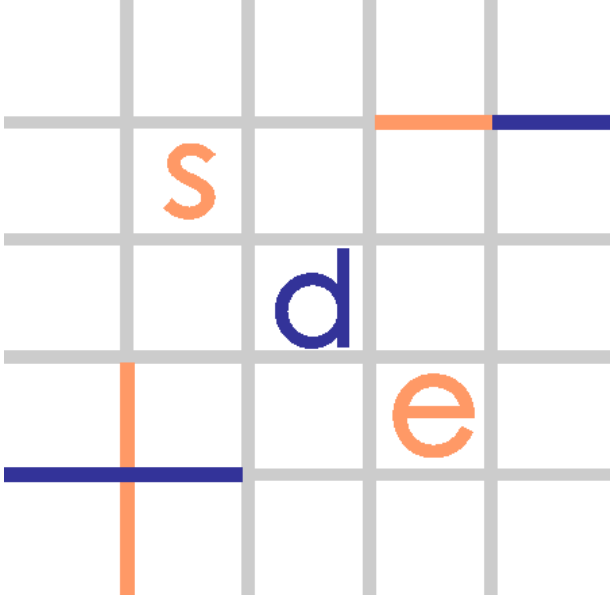
working papers



N o 3 v o l 2 0 1 0

Synthesis in Policy Impact Evaluation

Bojan RADEJ

Fourth revision (March 2011)

Ljubljana, March 2011
S l o v e n i a n E v a l u a t i o n S o c i e t y
Kardeljeva ploščad 17, Ljubljana
info@sdeval.si, http://www.sdeval.si

Slovenian Evaluation Society,
Working paper, vol. 1, no. 3
(March 2011, April 2010, April 2008)

Bojan Radej

Synthesis in Policy Impact Evaluation

Acknowledgements: I acknowledge valuable comments given to the previous version of the paper by anonimous referee and professors Mojca Golobic and Sreco Dragoš, both from University of Ljubljana.

Proposal for citation: Radej B. 2010. Synthesis in Policy Impact Assessement. Sage: Evaluation, 17/2(apr. 2011):133-150; http://evi.sagepub.com/content/17/2/

Ljubljana, March 2011

Synthesis in policy impact assessment

Bojan Radej

Slovenian Evaluation Society, Ljubljana, Slovenia, March 2011

Abstract

Impact evaluators of large-scale and multi-domain policy proposals have had difficulties in aggregating detailed assessments results into a summative evaluative conclusion. Assumptions about how to aggregate impacts from micro to macro level across multiple evaluation domains (economic, social, natural environment) differ and different approaches produce different end results. An aggregation problem arises because different policy impacts are not fully commensurable across scales (micro-, meso-, macro) and across domains of evaluation. A new approach to synthesis of results is proposed that takes incommensurability of policy impacts into account. Detailed impact assessments results are first partially aggregated into an input-output matrix of assessment domains (meso level) and then non-diagonally situated partial aggregates, defined as secondary impacts, are correlated. The aggregation problem is illustrated by the comparative assessment of the sustainability of the development program for the Pomurje region of Slovenia using three methods: micro (no aggregation of impact results), macro (full aggregation) and meso (partial aggregation) approaches. Only the meso approach is found consistent with the complexity of the challenge.

Keywords: aggregation, complexity, impact assessment

Introduction

To evaluate policy impacts is to collect evidence of the performance of a policy and make a value judgment about the policy's 'worth or merit' (Scriven, 1994). The importance of evidence-based policy making is one of the central postulates of the 'new public management'. In this model, policy impact assessment fulfils an essentially instrumental function in scientifically answering the question 'how effective are chosen policy means in achieving their specified public ends?' (Schwandt, 1997). There is, however, widespread recognition of the failure of evaluation research to live up to its promises of improving public management. It has been experiencing systemic failures in managing complex social issues and, in particular, in producing forms of knowledge that can strategically inform action to improve policy coordination. The Impact Assessment Board (IAB, 2009) estimates that some 60–80 percent of impact assessment studies currently provided to the European Commission supply the kind of information that does not inform policy makers as to whether their global objectives can be met.

A major challenge to the understanding of evaluation is posed by recent work on complexity theory. The concepts of chaos and complexity are well established in the natural sciences but have only recently come to the fore in the social sciences and in evaluative research. Complexity challenges basic assumptions that underpin positivist linear thinking. In the public sphere, complexity refers to multi-domain and multi-level problems, which are poorly understood with conventional causal models and thus not easily controllable. In the evaluation of complex issues, judgment comes from many independent sources, through many technical means, using multiple criteria embedded in different value systems. Evaluators thus necessarily face the challenge of how to consistently comprehend and integrate policy evidence.

A heterogeneous corpus of information produced in policy impact evaluation of complex policy interventions, is rendered sensible through the process of synthesis (Encyclopedia of Evaluation, 2004). However, when the observed evidence is not commensurable, it cannot be reduced to a common denominator (Funtowicz and Ravetz, 1994). As a synthesis cannot be obtained simply

by putting a jigsaw puzzle together to create a whole, there is an aggregation problem to be resolved.

A useful introductory example of the aggregation problem in social research is highlighted in relation to the theory of intersectionality in sociology (Kimberlé, 1994), which studies patterns of gender inequality. Intersectional theory examines how various socially and culturally constructed categories (domains) of discrimination which are incommensurable, such as gender, race and class interact in multiple areas, contributing to systematic inequality which is larger than the sum of three separately observed inequalities. Theory dismisses the additive claim that black women are twice as poor as white women due to both sexism and racism. According to Prins (2006), this theory emphasizes the fact that the complexity of the social inequality cannot be captured by such simple additive arithmetical frameworks.

Aggregation has been an extensively studied subject. There are many contexts in which we want to aggregate individuals' opinions or judgments such as in evaluative research and in research synthesis (Lipsey, 2009). List and Polak (2010) developed a theory of judgment aggregation to consider such situations. Nevertheless, they opine that the aggregation problem in general remains unresolved in social sciences because the literature has mainly concentrated on how to avoid it.

In a complex setting, there are essentially diverse or incommensurable viewpoints on social reality which value given social events in incompatible ways. When a controversial social phenomenon is evaluated, objectively measurable events may not be perceived in the same way by various stakeholders. There are two reasons for approaching policy impacts as incommensurable. First the evidence about them may apply to different scales of assessment so it must provide evidence of policy impacts at different levels: at micro, meso, or macro level; this induces incommensurability in scale such as when the given issue is perceived differently in local specifics and in its global entirety. This constitutes a 'vertical axis' of complexity and demands a multi-level approach to policy impact assessment. Second dimension of social complexity are different policy domains: as natural environmental, economic or social. These may put forward inconsistent policy demands which are equally legitimate and so equally important. This constitutes a 'horizontal axis' of complexity and requires parallel multi-criteria assessment of different evaluation domains resting on opposing primary values. Hence, incommensurability in scale and in domain of research are taken here as two coordination axes of 'organized complexity' (Easterling and Kok, 2002) or 'ordered social complexity' (Foster, 2004) which is situated between simplicity and chaos (un-organized complexity).

Social incommensurability (Munda, 2004) implies that no objective basis exists for rational choice between alternatives because different principles of legitimacy and social primacy must be reckoned with and reconciled (Wacquant, 1997) in evaluation of ordinary social dilemma. Incommensurability is also accompanied by knowledge structures which differ decisively in the propositions they take as their axioms. Kuhn (1970) argues that different sciences, say economics, sociology and ecology, integrate the same information in different ways. Different theories weigh the appearances of the same world differently and different types of models are suited for representing the judgment processes of a given social issue. This leads to disagreements about appropriate ways to sum up what we know about social reality. When confronted with multidisciplinary issues even competent, honest and disinterested scientists may arrive at different problem framings and conclusions because of systematic differences in the way they summarize available information (Mumpower and Stewart, 1996). When values are irreducibly plural, value conflicts are 'un-decidable' (Mouffe, 2000), i.e. there can be no rational

resolution of the conflict between them, and therefore also no clear cut judgment about 'merit and worth' of policy is possible.

The consequence of social complexity (incommensurability) is that what one observes and thinks is always predefined in domain and in scale of his/her judgment. This requires assessing different domains and scales as separate dimensions of assessment. A cumulative impact assessment of such complex policy settings is far from trivial (Van der Veen and Otter, 2002). One of the immediate consequences is the need to review the foundations of the methodology for aggregation (Scriven, 1994) in the assessment of complex policy proposals. Standard assessment methods are designed for the appraisal of assumingly homogeneous policy interventions with assumingly commensurable impacts (Elbers et al., 2007; Rotmans, 2002), while governments intervene in issues that are heterogeneous. Policy makers deal with numerous sectoral policies as well as with a great number of goals – all of which may hinder, support or reinforce each other. The idea that policies should be evaluated with differentiation between their incommensurable aspects (domains and scales) has long been recognized in policy evaluation (Rotmans, 2002; Weaver and Rotmans, 2006) but poorly implemented in policy impact assessment.

The lack of explicit justification of the aggregation procedure is the Achilles heel of assessment efforts (Scriven, 1994). Different algorithms for aggregating detailed assessment results across assessment domains and scales yield different results and lead to contradictory policy conclusions. The majority of standard assessment approaches are struggling with how to appropriately address incommensurable oppositions in policy impact assessment, such as the EU's strategic impact assessment (2001/42/EC), the Impact Assessment Guidelines (SEC(2005)791), the territorial impact assessment (TIA; ESPON - 3.2, 2006), and the ex-ante assessment of the contribution of the EU structural funds to regional sustainability (GHK et al., 2002). There is an apparent paradigm crisis in impact assessment (Hertin et al., 2007; Virtanen and Uusikylä, 2004). Foster and Potts (2007) argue that the crisis is in part a result of unresolved aggregation problem. There is a disagreement over assumptions about the aggregation of the assessed detailed impacts across multiple assessment domains and from micro to macro level. Older methodological approaches strictly speaking do not aggregate detailed assessment results. Luna Leopold et al. (1971) were the first to address the issue. They proposed a detailed assessment method at the micro level from which synthesis of results remains absent. Besides, Leopold's matrix is binary because it is only concerned with two domains – economic and natural environment and assesses the possible side or secondary effects of the former on the latter (see section 2).

Recently Ekins and Medhurst (2003, 2006) proposed a macro assessment approach from a complete multi-domain perspective. They claim that the majority of the assessed policy impacts conform to the normatively prescribed critical system thresholds, so they can be evaluated as 'normal' and therefore commensurable. They propose a vertically and horizontally aggregated version of the Leopold matrix, expanded from two to four assessment domains (social, human, economic, natural environmental; the obtained matrix is named here as the Leopold-Ekins-Medhurst matrix or LEM). They further allow for aggregation of assessed impacts for all assessment criteria within each of four assessment domains that are placed in the columns of the LEM. They aggregate policy impacts only within each impact domain, but not between them. In this way they obtain four composite indicators of a policy impact. Their work is an important step towards cumulative assessment methodology in the framework of social incommensurability. An analogous logic to LEM is now accepted in various procedures, such as the EU's strategic impact assessment (2001/42/EC), territorial impact assessment (ESPON - 3.2,

2006), and assessment of the contribution of EU structural funds to regional sustainability (GHK et al., 2002).

Nevertheless, LEM's aggregation approach is inappropriate in its second step, when it allows for the summation of all policy measures' impacts irrespective of their source of impact. The problem is that a given policy's impacts on different assessment domains are not qualitatively the same and not commensurable. This article argues that fragmented assessment results can be aggregated in LEM only by the given source *and* particular impact domains, i.e. *partially*. This reorganizes LEM into the square input-output Leontief (1970) matrix (see section 3).

The methodological divergence of various aggregation approaches is illustrated by a comparative ex-ante assessment of the development programme for the Slovenian region Pomurje for 2007–2013 (RDPP; Radej, 2006). Pomurje is the least advanced Slovenian region (at NUTS 2; with 6.6% of the national territory and 4.3% of national GDP) with a strong cultural and ecological identity – more than a third of its territory is protected nature reserves including unique landscape along the River Mura. Its economic capital is fragile but improving since the mid 1990s. Social capital is very frail and further depleting. For half a century, the region had been surrounded by cold war borders. Since the early 1990s, it has found itself on the main European transport corridor which exposed it unprepared to international flows of people and goods. The accession of Slovenia to the EU also imposed a more restrictive border regime between Pomurje and Croatia (an EU candidate country), which previously (in Yugoslavia) were traditionally close. These trends have further weakened social capital, leading to continued depopulation, brain-drain, long-term unemployment, prolonged health and social risks for vulnerable groups (the majority of the population is officially classified as vulnerable). Regional development lags have accumulated in regional social capital despite the increased inflow of resources earmarked for less-advanced regions from the national or EU budget in the past two decades, because not enough emphasis was placed on genuine local needs. This suggests that national and regional policy makers have failed to appropriately address critical regional trends and trade-offs between economic, social and natural environment domains.

A new approach to aggregation of assessed policy impacts is proposed. The distinction between commensurable and incommensurable social events is crucial in assessment, but not sufficient. What is needed is also distinction between weakly commensurable and weakly incommensurable policy impacts which is only possible when evaluative judgement is relocated from micro or macro level to meso level. The aim here is to contribute to what Sanderson (2000) calls 'a holistic approach' to assessment which identifies not only how effective a policy is in achieving its primary aims but also if it produces tensions or creates synergies between them.

Standard approach

First-generation methodologies for matrical impact assessment of complex policy interventions were proposed by geologist Luna Leopold and colleagues (1971). Their approach encompasses only two domains (economy, natural environment). It is focused at the micro level only assessing impact of individual economic measures' impacts on selected environmental criteria. The approach is nevertheless appealing because it goes beyond monitoring policy performance – i.e. how a particular economic policy measure impacts a particular economic assessment criterion. But Leopold evaluates side effects or secondary impacts of economic policy measures on non-economic, natural environmental assessment criteria.

In the finest positivist tradition of analytical science, Leopold et al. (1971) aim to assess the complex policy issue through the detailed description of its numerous impacts. They listed in their matrix the 100 most important economic policy measures horizontally in the rows of the matrix and 88 criteria of natural environmental impact vertically in the columns. This created a matrix with 8800 cells – each further divided into four sections that describe every impact by its size (large/medium/ small), direction (positive/negative/neutral), probability (high/low) and the amount of risk (critical or not). In this way impacts are assessed in sufficient detail to enable maximally informed decisions of elected politicians about the impacts of their policy proposals.

Appropriately recognizing the incommensurability between the economic and natural environmental domains of an assessment, Leopold explicitly rejected the summation of diverse impacts into a composed impact indicator. He claims that detailed assessment results should be presented disaggregated, leaving policy makers with full responsibility for the evaluation synthesis and for drawing its policy implications. Refusal of aggregation is essential for neutral evaluation, says Leopold, as it draws a demarcation line between evaluator and policy maker to protect the former from the value judgments and political interference (Kunseler, 2007). This argument has been accepted as an evaluation standard for decades and is maintained in the EU's Impact Assessment Guidelines (SEC(2005)791).

Rejection of summation in evaluation and shifting this task to policy makers is problematic. Refusing to summarize is 'letting the client down at exactly the moment they need you most' (Scriven, 1994). Assessment results make political decisions more informed but not necessarily easier (Diamond, 2005). It is exactly politicians' failure as social aggregators that call for policy evaluation in the first place – recall Arrow's impossibility theorem (1951). A problem is especially evident in the assessment of complex 'cross-cutting' social issues (Sanderson, 2000) such as sustainable development, gender equality and social cohesion. Sanderson notes that assessment that seeks to isolate policy instruments will produce results of limited usefulness due to their limited external validity. Assessment which simply produces non-overlapping information tends to underplay inherent system contradictions, legitimizing disregard of stakeholders' concerns in policy making (Stake, 2001). Without any explanation of how different parts of a public programme work together, assessment fails to satisfy information needs at the strategic level and produces banal answers to multi-dimensional problems (Virtanen and Uusikylä, 2004). Finally, when results are left unrelated in their interpretation, it is impossible to substantiate evaluation findings, which leaves them exposed to political manipulation.

Even if politicians and policy makers were able and willing to jump into evaluators' shoes and undertake evaluation synthesis on their own, it remains logically impossible, following Arrow, to derive strategic policy advice directly from fragmented inquiry. This is illustrated by the problems of evaluative synthesis in the case of impact assessment of the Regional Development Program for Pomurje 2007–2013 (Table 1).

Table 1 was obtained in a standard way: a group of experts convened a workshop and applied the Delphi method to assess the possible impacts of the 47 proposed RDPP measures (social, economic, natural environment) on a selected set of assessment criteria that also relate to social, economic, natural environment factors. In order to simplify the description of methods, experts' opinions are presented here only in terms of the direction of impacts: be they positive, neutral or negative. However, extensive qualitative and quantitative research was invested in studying the characteristics of these individual impacts. The fragmented assessment results presented in Table 1 can be summarized following three lines of conventional reasoning in evaluation:

(i) a prevalence of the programme's positive impacts suggests that a majority of the measures will positively contribute to regional development, which supports its endorsement;

(ii) negative impacts focus evaluator's attention on the weakest parts of the proposal which ought to be improved or abandoned in the programme;

(iii) neutral impacts (0) are not problematic.

These three lines of reasoning would suggest that the evaluator and policy maker should focus their attention on negative impacts. However, this evaluation approach is not appropriate in the context of social complexity where positive (and sometimes neutral) impacts may not be unproblematic; nor can a negative impact simply be treated as unacceptable.

The prevalence of positive impacts does not support the conclusion that the proposed policy is adequate as submitted, but only that the policy proposal has been prepared by a basically competent public authority. Proposals when prepared by a democratically elected government are carefully scrutinized as well as painfully negotiated among various group interests, usually long before they are submitted for impact assessment.

There are some additional reasons why a dominance of positive impacts shall not by itself lead evaluators to positively assess the policy proposal as a whole. Impacts are sometimes assessed against criteria selected by formally responsible implementation agencies themselves. Even when this is not the case, impacts are assessed in relation to individual criteria and thus in isolation from each other. Successful realization of separate policy measures cannot by itself guarantee positive society-wide impact of the policy as a whole if its goals (or assessment criteria) are in conflict – a very common situation in public sector policy making.

A prevalence of positive impacts in a Leopold matrix can, at its best, inform policy makers about their effectiveness observed from an atomistic perspective (micro view), while it does not enable a systematic conclusion about the proposal's integral impact on the overall welfare. Only when systematic evidence of positive impacts is obtained, can evaluators assess the appropriateness of the overall proposal. But 'systematic evidence' can be identified only at higher levels of evaluation, when detailed results of assessment are properly summarized.

The imperative of aggregation sparks two methodological concerns, both of which are linked to the conditions under which negative impacts may be tradable and thus aggregatable with particular positive impacts.

The *first* concern relates to situations in which different experts cannot reach consensus on the direction of an impact (positive or negative). Some approaches, such as CAF (Common Assessment Framework, 2006) – a self-evaluation tool suggest that revealed differences in the assessment of a given issue need to be discussed with the aim of reaching consensus about direction and even about intensity of impact. However, forcing consensus for every single assessment detail is risky because it could invoke asymmetries within the assessment team – such as their different negotiating skills which can lead to a kind of closed, exclusive process (Connelly and Richardson, 2004) where the dominant actor prevails. Sankey (1995) is instead more in favour of 'rational disagreement' where disagreeing experts verify arguments for their claims and discuss them – not to reach consensus but at least to confirm validity of arguments that stand behind each claim. If conflicting expert assessments are well founded, disagreement between them is justified and irresolvable. So another summative option, which is also applied in the presented case study, is cancelling out the opposing assessments. Different expert opinions are not only equally valid but are also only partial claims when seen from the wider perspective

of the overall programme evaluation. A threat that experts' polarized disagreements are netted out should induce cooperative efforts between different experts – to avoid the cancelling of their judgments through synthesis.

Table 1: Leopold's micro view

Impact Domains / Program Measures	Economic		Social		Environmental	
	GDP growth	Investm. intensity	Unemp-loyment	Migration	Abate-ment ex-penditure	Sewerage connectio.
1 Development lag	+	+	0	-	+	+
2 Competitiveness	+	+	-	-	-	+
3 Investment promotion	+	+	+	+	0	+
4 Endogenous advantages	+	+	+	+	+	+
5 Entrepreneurship	+	+	-	+	0	0
6 Regional tourist organizational model	0	0	0	0	0	0
7 Pomurje as a tourist destination	+	0	0	+	0	0
8 Destination management	0	0	0	0	+	+
9 Destination marketing	+	0	0	+	0	0
10 Human resources in tourism	0	+	+	+	0	0
11 Quality management	0	+	0	0	+	+
12 Tourist infrastructure investment	+	+	0	0	+	+
13 R&D in tourism	+	+	0	0	0	0
14 Health inequality (criteria)	0	0	0	0	0	0
15 Health promotion network	+	+	+	+	0	0
16 Health inequality – regional	+	0	+	+	0	0
17 Health inequality – vulnerable groups	+	0	+	+	0	0
18 Quality, access to health services	+	0	0	0	0	0
19 Healthy environment	0	+	0	0	+	+
20 Mental health	0	0	+	+	0	0
21 Agriculture modernisation	+	+	-	-	+	+
22 Environmental agriculture	+	+	+	+	+	+
23 Entrepreneurship in agriculture	+	0	+	0	0	0
24 Human development in agriculture	0	0	+	+	0	0
25 Value added growth	+	+	-	-	+	+
26 Products, services – farms	+	+	+	+	+	+
27 Products, services - agro industry	+	+	-	0	+	+
28 Marketing agro-products	+	+	0	0	0	0
29 Rural development –products & services	+	+	+	+	0	0
30 Countryside development	+	+	+	+	+	+
31 Rural entrepreneurship	+	+	+	0	0	0
32 Rural stakeholders' co-operation	+	0	+	+	0	0
33 Water supply	+	+	+	0	+	+
34 Transport infrastructure	+	+	+	+	+	0
35 Alternative, local energy sources	+	+	+	+	+	0
36 Energy distribution network	+	+	0	0	0	0
37 Access to IT services	+	+	+	+	0	0
38 Waste waters, collection & treatment	+	+	+	0	+	+
39 Solid waste management	+	+	+	0	+	+
40 Communally equipped zones	+	+	+	+	+	+
41 Water quality	+	+	+	0	+	+
42 Revitalisation of hot-spots	-	-	0	0	+	+
43 Illegal land-filling, monitoring	+	+	+	0	+	+
44 Nature and culture conservation	-	+	-	0	+	0
45 Energy policy	+	-	+	0	+	0
46 Spatial planning	+	+	+	+	0	0
47 Communication strategies	+	+	0	0	0	0

Source: Radej, 2006.

The second difficulty is related to aggregating positive and negative impacts of a given policy measure on various assessment criteria (or similarly, positive and negative impacts of *different* policy measures assessed against the same evaluation criteria). For instance, in Table 1, may negative impacts of entrepreneurship promotion on employment in Pomurje be outweighed by positive impacts of entrepreneurship promotion on migration? Or another example: is it permitted in evaluation to cancel out additional tons of greenhouse pollution (negative impact) with additional purchase of tradable pollution permits (positive impact, because this finances natural environmental investment at the permit seller's plant)? Greenhouse emissions cause irreversible changes in the atmospheric conditions and so the economic and climate domain of policy making need to be treated as incommensurable.

Thus a trade-off between greenhouse gases and money is not adequate as a general principle. However, from a multi-scale perspective, some judgments are localized and so different from the global ones. Trade-offs between income and greenhouse emissions are not incommensurable in every single case, or at least people are not willing to treat them as such.

To incorporate this peculiarity in policy evaluation, system thresholds – such as ecological and social safety standards – have emerged (for a survey of literature see Muradian, 2001). Threshold marks a 'tipping point', beyond which a small additional change can have a disproportionately large effect on the entire society and could endanger its basic integrity. A policy proposal is by definition not allowed, through any of its impacts, to disregard system thresholds, such as legal norms and safety standards, because the high probability of negative policy impacts involving critically high system risk would make the policy proposal illegitimate.

In social and evaluation research, quantitative thresholds reflect the fact that there are discontinuities in the measurement of value-based social phenomena and value addition (Mason, 2006) that are consistent with discontinuity in individual and social values. The concept of system thresholds is closely linked to incommensurability of social phenomena. As Wiggins (1997) explains, two values are incommensurable if there is no general way in which A and B trade-off across the range of situations of choice and comparison in which they are present. But such examples are rare in social context. Social phenomena are usually incommensurable 'only' beyond (or below or both) their threshold values but not necessarily within these limits. Within the safety limits an agent either does not sense the qualitative difference between two distinct social conditions or refuses to declare a preference for one or the other (Luce, 1956 in Munda, 2006), as in the case of minor natural environmental damage that stays within 'safe' ecological standards. System thresholds define objective criteria against which policy impacts are conditionally commensurable. They delineate a space of social normality within which concerned parties in a micro context freely interact and therefore also freely trade-off positive with negative impacts, depending entirely on specific, subjective, and value-based considerations of those directly affected by these trade-offs. Recognition of system thresholds allows progress in the cumulative impact evaluation. One of the first macro-evaluation methods following such reasoning was strategic environmental assessment (SEA Directive, 2001/42/EC; Sadler and Verheem, 1996), but it gives no indication about how to cumulate natural environmental impacts and parallel them in an aggregate way to economic ones.

The missing link has been contributed by Ekins and Medhurst (2003, 2006) in their novel approach to assessment of the EU structural funds' impact on regional sustainable development. They proposed the 'four capitals model' which evaluates in parallel economic, social, natural environmental and human impacts of a given policy proposal. This particular logic of multiple welfare domains can be traced back to the Brundtland report (WCED, 1987), and to the

conference on sustainable development in Rio de Janeiro (Ekins, 1992; Munasinghe, 1992; UNCED, 1992).

Ekins and Medhurst have proposed their matrix (LEM) as a highly compacted form of the Leopold matrix. The columns are reduced from 88 fields of possible natural environmental impacts to four domains of sustainability, each covered by a smaller number of assessment criteria (only two in the illustrative example, presented in Table 1; in LEM these are already summed up horizontally within each domain; see Table 2). For simplicity, the 'four-capital model' of Ekins and Medhurst is reduced in this article to a 'three capitals model' because this is sufficient to explain the aggregation problem for a three-domain issue. The number of rows in LEM matrix may be as large as in the case of Leopold matrix, but it is also reduced, as suggested in Ekins and Medhurst, from 47 programme measures (Table 1) to six main regional policies (see Table 2).

Table 2: LEM's macro view

Policy domains / Sectoral policies	Economic	Social	Environment
Value added growth (rows 1-5)*	+++	-	+
Tourism (rows 6-13)	+	0	0
Health (rows 14-20)	0	+	0
Rural development (rows 21-32)	+++	+	++
Infrastructure (rows 33-40)	+++	++	++
Natural environment (rows 41-47)	+	+	+
Summary impact of RDPP (rows 1-47)	++	+	+

Source: Radej, 2006. *Note*: * Summary of rows 1-5 in Table 1, etc.

LEM presents impacts on a wider range of scores compared to Table 1: impacts are expressed on the range from the most robust positive impact with the highest score (+++) to the most negative impact with the lowest score (---), with all other seven intermediate possibilities included. When uncertain about how to round-off aggregate impact from Table 1 to Table 2, a decision was taken based on the comparison of the financial weight of the related measures. So in the final row of Table 2 a vertically aggregated impacts of the entire RDPP is presented for each of three assessment domains.

Table 2 presents the infrastructure development will be the most welfare-enhancing policy, followed by rural development policy. The most problematic is the negative impact of value-added growth on social welfare. The impact of health policy and tourism are disappointing – the explanation is that their measures mostly relate to the preparation of plans and regional organization structures. But the overall impacts of the programme, presented in aggregate at the bottom row of Table 2, do not appear problematic. The programme will improve regional sustainability in all three domains, though more in the economic domain (++) than in social and natural environmental ones (+). The summary impacts of the RDPP are therefore unevenly positive but differences are small. Thus the assessment concludes that the programme will have a rather weak overall impact on regional sustainability.

However, the main task of assessment of complex programme is not to examine how sectoral policies achieve their narrowly defined goals. LEM cannot say much about RDPP's impact on the integrity of regional policy making in terms of its internal oppositions from the perspective of conflicts and synergies between policy domains. Although LEM constructs a macro level view

on RDPP's overall impacts, it does so via bottom-up aggregation of the impacts observed at the micro level, thus retaining the micro-based assessment logic put forward by Leopold.

Meso approach

If detailed assessment results are not cumulated, as in Leopold's method, the assessment produces findings that are microscopic – too fragmented for the evaluation of complex issues. In contrast, full aggregation, such as in macroscopic LEM, results in findings that are too aggregated to enable a judgment about the policy's inherent structural tensions among assessment domains. In this section we first discuss the arguments about inconsistent aggregation in LEM and then negotiate a compromise between micro and macro approaches which goes beyond conventional dichotomies.

Ekins and Medhurst did not acknowledge in LEM that the assessed impacts are vertically not fully aggregatable despite stringent conformity to their associated system thresholds. Column aggregation assumes the homogeneity of the impacts of all different policy measures on a given assessment criteria. However, many studies demonstrate that a given policy does not influence different areas of impact in the same way (Schnellenbach, 2005). Policy impacts are direct (primary) or indirect (secondary). They are direct when they affect the target impact area, but indirect or unintended when they impact on areas which are not primarily targeted with a given policy measure. Such impacts often fall under the jurisdiction of other policy domain (Rotmans, 2006). Hence, each policy measure impacts each evaluation criteria differently and subsequently impacts are not vertically fully aggregatable.

Non-neutrality in impact is confirmed even for those sectoral policies that had previously been taken as the most homogenous in impact such as monetary (Lucas, 1972) and tax policy (Leith and Thadden, 2006). For instance, impact of increased interest rates in monetary policy is directly more relevant for the economic than for non-economic actors. In principle, sectoraly specialised policy interventions should *always* be analysed in relation to the general interest they are supposed to serve (Donzelot, in Burchell et al., 1991) which comprises economic as well as non-economic issues and direct as well as indirect policy impacts. Yet, indirect or secondary impacts routinely fade out in policy evaluation because it is assumed that they are 'too complex' and thus impossible to track.

Sectoral specialization, and therefore the non-holistic nature of policies, implies not only that a distinction between incommensurable impact domains must be preserved in evaluation (Ostmann, 2006) but also that secondary impacts should be taken into account as equally important. It is exactly the distinction between primary and secondary impacts, that enables partial vertical summation in LEM. Aggregation is partial because the cumulation of assessment evidence is constrained by different *numeraires* resulting from the differentiated value frameworks in diverse policy domains. For example, economic and social policy's impacts on the natural environment are not commensurable so they are aggregated separately (economic impacts on environment separately from social impacts on natural environment).

At first sight partial aggregation seems at odds with the strong version of the incommensurability thesis. Martinez-Alier et al. (1998) point out that in situations when there is an irreducible value conflict in public affairs, we can only search for weak comparability as a facilitator of collective discourse. Some authors have explicitly argued against the strong incommensurability thesis (Morgan, 2007; Nola and Sankey, 2000). They proposed making a distinction between relations of strong and weak in/commensurability. Impact is said to be weakly commensurable when

specific limitations are imposed in summation procedure such as with the partial aggregation rule. Impacts that are weakly commensurable in two or more incommensurable domains of the assessment, like 'hybrid' socio-economic impacts, are said to be weakly incommensurable. The difference between strong and weak incommensurability can be illustrated with an example. It is entirely possible to mix apples and oranges in fruit juice because their juicy meat is weakly incommensurable, they can be tastefully mixed under certain conditions. Despite such compatibility it will never be possible to grow an apple tree from the seed of orange, because seeds in their primary essence are strongly incommensurable. Only in the strong case can one not aggregate apples with oranges. Weakly incommensurable impacts are secondary to both, economic and social domain when observed separately. Secondary policy impacts are hybrids, so they can be evaluated against two otherwise incompatible set of judgments (domains) so they imply a delicate possibility of translation between principal incompatibilities. Recognition of weak in/commensurability of impacts is therefore a precondition for consistent evaluative synthesis in the assessment of complex policy issues.

Taking into account weak commensurability of impacts, evaluators needs to regroup all rows, in Table 1 in the same way as impact areas are grouped in columns – by their three incommensurable domains (see Table 3). This divides the Leopold matrix into nine subsections.

Table 3: The meso view

Impact Domains / Policy Domains	E	S	N
E	(E∩E) +++	(E∩S) -	(E∩N) +
S	(S∩E) +++	(S∩S) ++	(S∩N) ++
N	(N∩E) +	(N∩S) +	(N∩N) +

Source: Radej, 2006.

When assessed policy impacts are partially aggregated within appropriate sections of Leopold matrix, partially aggregated impact is obtained for each sub-section. These sub-aggregates can be presented together in a square 'input – output' or a Leontief's matrix (its 'central quadrant'). Leontief initially developed his matrix to facilitate inter-sectoral studies, because it is suitable for an explicit presentation of the tension between sectors, direct and indirect. For instance, if we have agriculture, industry and services as three sectors in an economy, the rows and columns in a matrix can illustrate how the sectors are linked. Agriculture requires some capital goods and chemicals from industry, and concurrently it supplies its output to the industry and services as their. intermediate inputs. Additional agricultural needs for industrial goods will also indirectly induce demands for services needed to enable increased industrial production, etc.

The square matrix exists hierarchically above the micro level (Leopold matrix) because it is aggregated from it. At the same time, as a partial aggregate, the matrix exists at a lower level than the macro (LEM). The input-output matrix thus presents an intermediate or meso view of the assessed impacts. This approach to evaluation is therefore 'meso-matrical'. The concept of a mesomatrix is important because the complex social system is built upon meso (Dopfer et al., 2004). This also conforms to Easterling and Kok (2002) who argued that meso is the perspective from where the modelling of social complexity is the most tractable 'a priori'.

The meso approach is holistic because it assesses policy impacts across all domains (demanding horizontal synthesis) and at different levels at which a policy is evaluated (requiring vertical synthesis; Rogers, 2008). Being rooted simultaneously both in micro and in macro uni-modal

logics, meso exhibits bi-modality (Dopfer, 2006). This enables the evaluator to adopt a dual perspective (Dopfer et al., 2004), which also permits the application of what Malthus (in Cremaschi and Dascal, 1996) described as the 'doctrine of the middle'. A meso perspective treads the path between 'difference and sameness' (Allmendinger, 2002) which enables mid-level articulation of oppositions that accompany public choice and in this way facilitates the understanding of the basis of deep system conflicts (Mertens, 1999). Evaluators can, for instance, intervene in conflicts from the meso perspective to help actors understand where their disagreements have epistemological and ethical roots and help expose the incommensurable meaning systems by which these facts are being interpreted (Bovens et al., 2008). As a 'plural-relativistic' view (Geertz, 2000) of social complexity, meso covers many parallel views of one closed reality containing many (pre) existing substantial contexts. With its bimodal and intersectional characteristics meso is situated in 'the un-excluded middle' (Wallerstein, 2004) of the social complexity.

A meso-matrix refocuses evaluator's attention from the performance assessment of sectoral policies' effectiveness to the evaluation *of the relations (trade-offs) among assessment domains* – their oppositions and synergies. It reworks the presentation of policy proposal in the perspective of overlaps between its policy's domains (inputs, in rows) and its assessment domains (outputs, in columns). Intersection between input and output domains is denoted with the intersection sign '∩' from the set theory. For example, economic policy impact on the social domain is denoted as E∩S (E intersects S). A meso-matrix of RDPP's partially aggregated impacts is presented in Table 3.1

Table 3 is mainly not interesting for the policy maker who is narrowly concerned only with the assessment of performance of his/her own actions and is not able or willing to question their systemwide impacts. Table 3 is mainly relevant for higher level policy makers who are concerned with overall policy consistency taking into account primary and secondary impacts alike. Table 3 conveys a qualitatively different view of policy achievements compared to Tables 1 and 2. Table 2 (LEM) does not produce a qualitatively different view compared to Table 1 since they are both focused on the assessment of successful realization of their primary (sectoral) goals. In Table 3 the evaluative conclusion is 'emergent'. This means that evaluative conclusions cannot be reduced to the fragmented pieces of information about individual impacts but emerges as a new qualitative judgement. To understand how judgement emerges in the meso view one needs first to correlate the elements of Table 3. A correlation is applied when one studies connectedness between pairs of variables in causal models. In our case 'to correlate' means two directional assessment of relationship between two domains, such as E∩N, with its diagonally symmetric opposite in Table 3, N∩E; this evaluates the reciprocal relationship (NE) between natural environment (N) and economy (E), which explains trade-offs between two assessment domains and how they work together in the assessed policy.

Standard programme theory explains logic of policy intervention in linear fashion as one directional relationship between policy as a cause and its intended effects ('economy' impacts 'environment', or the opposite). Meso evaluation is non-linear as it relates to causally cyclical issues ('economy' impacts 'natural environment' and simultaneously 'natural environment' impacts 'economy'; NE). Meso perspective requires 'corelational evaluation theory'. Emergence of qualitatively new judgement then takes place just because the evaluation logic changes from linear micro assessment of policy effectiveness to corelational assessment of the main policy synergies at meso level.

The square matrix in Table 3 consists of two distinctive classes of relationship: (i) primary impacts (E∩E, S∩S, N∩N) are located on the diagonal, top left to bottom right; (ii) overlaps or secondary impacts (E∩N, etc.) are located below or above this diagonal. The diagonal elements present partially aggregated results of intended or primary policy impacts. As strongly incommensurable they cannot be aggregated any further and are interpreted as they are – nothing qualitatively new can emerge from their interpretation. Diagonal elements of Table 3 indicate that: (i) regional economic policy would be very successful in achieving its own primary goals (three pluses); (ii) moderately successful in social policy (two pluses); and (iii) only weakly effective in environmental protection (one plus). Therefore social, environmental and economic domains of sustainability are not treated indiscriminately in RDPP. This observation does not match with the previously obtained conclusion from summary row of Table 2 that suggested a broadly balanced impact of the programme on the three assessment domains. Besides, RDPP's cumulative impact on economic and social domains is assessed higher in the input-output matrix than in LEM. Insight into RDPP's internal consistency of trade-offs between policy domains is obtained with the correlation of non-diagonally located and weakly incommensurable partial aggregates. It produces three correlates in Table 4: SE, for the 'socio-economic' intersection, denoting bi-directional overlap between S∩E and E∩S; NE (see above) and analogously SN, as the socio-environmental correlate.

Table 4: Correlation matrix of synergy between policy domains

Output / Input	E	S	N
E	Economic sustainability (E∩E) = (+++) → Robust positive impact	Socio-economic sustainability (S∩E, E∩S) = (+++, -) → Robust link between E and S but very unbalanced in favour of E	Eco-commercial sustainability (N∩E, E∩N) = (+, +) → Low correlation, balanced impacts between E and N
S		Social Sustainability (S∩S) = (++) → Moderately positive impact	Socio-natural sustainability (S∩N, N∩S) = (++, +) → Moderately strong correlation, unbalanced in favour of N
N			Natural sustainability (N∩N) = (+) → Low positive impact

The correlation result in SE is strong but damaging for S which indicates a socially constraining economic policy (Table 4). Low correlation in SN results from a socially limited environment protection policy. Here we recall the previous observation that environment protection policy will not be very effective in pursuing its primary goals – so the proposed Programme will at least in relative terms further impose a social burden for only slight anticipated improvement in environmental sustainability. A socially constraining economic and environment protection policy both further weaken already fragile social capital. Besides, the economic domain very poorly integrates with S and with N; similar can be said about the feeble secondary impacts of N on E and S. Table 4 thus reveals non-social character of RDPP, which is highly problematic taking into account the baseline conditions with S already deeply depressed in comparison with E and N. This observation is entirely absent from the assessment results presented in Table 1 and 2. The final conclusion is that the RDPP is not consistently contributing to regional sustainability. This is the opposite conclusion of the one obtained from Table 2.

Discrepancy in conclusions derived from Tables 1, 2 and 4 are, of course, not due to different detailed expert assessment at the micro level, which remains the same in all three instances. It emerges solely from the summative endeavour of evaluation. Policy failure to implement more cohesive policies is therefore not necessarily a result of intentional bias, but can be a consequence of inadequate strategic evaluation of policy initiatives in the presence of social complexity. Different aggregation assumptions lead to different evaluative conclusions, and inappropriate assumptions in results synthesis lead to misleading policy advice. This validates the premise given in the introduction that many difficulties arising from inconsistent impact evaluation are caused by an inadequate handling with social complexity, with flawed assumptions about the nature of (in)comparability of social issues. The conclusion is that, in the context of social incommensurability, while a synthesis is vital to evaluation methodology (see Lipsey, 2009), the opposite is also true: a theory of synthesis is the starting point from which the more appropriate evaluation approach can be developed.

Conclusions

This article has explored how standard approaches in matrical impact assessment take into account social incommensurability of policy impacts in formation of evaluative conclusions. The meso synthesis is proposed as an alternative to both micro and macro-based approaches. Leopold et al. (1971) accurately perceived natural environment and economy as two incommensurable domains of valuation; but they failed to see that their approach to the assessment is actually concerned only with secondary impacts, which are in fact weakly commensurable, so they can be partially aggregated from micro to meso level. On the other side, Ekins and Medhurst (2006) accurately observe that aggregation of impacts within a given assessment domain is formally correct. They provide a procedure in cumulative assessment for providing synthesized evaluative conclusions. They appropriately allow only for partial aggregation of impacts on the output side of assessment, but they fail to apply the same principle consistently on the input side of policy domains. So at the end their macro assessment approach basically remains 'micro-foundationalist'.

The controversy has been illustrated by the case of ex-ante impact assessment but there is no reason why the same logic could not also be applied in ex-post evaluation context. The proposed seems appropriate also for process evaluation because they both refer to intermediate issues.

The article demonstrates that social incommensurability is not an irresolvable obstacle to more holistic reasoning in public policy. Assumption of incommensurability constitutes only a safety mechanism which reminds us that social issues are complex and cannot be explained in their entirety from any specific point of view. However, strong differentiations are important for only a small number of social issues. Even though contemporary societies are built on incommensurable oppositions which cause strong social fractures and invoke strong dis/agreement, the majority of issues important for the reproduction of everyday social life are linked by weak ties (Granovetter, 1983). This article employs this observation by suggesting how to utilize weakly commensurable and weakly incommensurable evidence in policy evaluation in order to improve understanding of cohesive processes in society while at the same time preserving its high diversity and plurality.

Specialized sectoral policies produce biased impacts but functionally they are equally important. In such a situation, when there is no straightforward mechanism to install an optimal public policy, a policy proposal that is the most secondary effective ought to be chosen (cf. Demsetz, 1969). As societies grow more complex, policy makers should be increasingly aware not only of

their own agency's primary objectives narrowly defined, but also of wider implications and unwanted effects of their (in)activity on wider society and its welfare. Disregard for secondary impacts might explain why good individual policies, based on strong values and even common sense, often lead to disappointing overall results (Chapman, 2004). The same idea of importance of secondary issues is relevant to the thought of both Hayek and Popper, who take the view that the unintended consequences of action are the principal concern of social science and that the existence of unintended consequences is a precondition for the very possibility of the scientific understanding of a complex society (Vernon, 1976).

Today majority of evaluation studies try to earn their neutrality in the first evaluation step of scientifically objective assessment of policy impacts. This article sheds light on the cases when neutrality should be earned also in the synthesis of detailed assessment results. Our experiment not only confirms the need for the cumulative evaluation of complex policy proposals, but places synthesis concerns, via neutrality considerations, into the centre of the efforts aimed at improvement of policy impact evaluation.

So, meso approach to evaluation is not simply a synthesizing logic of drawing general conclusions from elementary issues. Synthesis requires a careful framing of the research problem which calls for a specific formalism that theoretically remains faithful to the framework of study (Barroso et al., 2003). In this regard the meso approach offers a distinctive theory-based method of judgment and distinctive programme theory because it develops a distinctive basis for explanation as to what it is about a policy which makes it work.

References

Allmendinger P. Towards a Post-Positivist Typology of Planning Theory. Sage, Planning Theory, 1/1(2002):77–99.

Arrow K. 1951. Social Choice and Individual Values, (2nd ed., 1963), New Haven: Yale University Press. 124 pp.

Barroso J., C. J. Gollop, M. Sandelowski, J. Meynell, P.F.Pearce, L.J. Collonis. The challenges of searching for and retrieving qualitative studies. Western Journal of Nursing Research, 25(2003):153-178, in Bonds, 2007.

Bar-Yam Y. A Mathematical Theory of Strong Emergence Using Multiscale Variety. Wiley, Complexity, 9/6 (2004):15-24.

Bondas T., E.O.C. Hall. Challenges in Approaching Metasynthesis Research. Sage, Qualitative Health Research, 17/1(2007):113-21.

Bovens M., P. 't Hart, S. Kuipers. 2008. The politics of policy evaluation. The Oxford Handbook of Public Policy, Chapter 15, pp. 317-33.

Burchell G., C. Gordon, P. Miller (eds). 1991. The Foucault Effect: Studies in governmentality. Hertfordshire: Harvester Wheatsheaf, 303 pp.

Chapman J. 2004. System failure - Why governments must learn to think differently. Second edition, London: Demos, http://www.demos.co.uk, [IV/09].

Common Assessment Framework. 2006. Brussels: EIPA – European Institute of Public Administration, http://www.eipa.eu/en/campaigns/go/&tid=20, 51 pp., [VI/09].

Connelly S. T. Richardson. 2004. Exclusion: The necessary difference between ideal and practical consensus. Journal of Environmental Planning and Management 47/1(January 2004):3-17.

Cremaschi S., M. Dascal. Malthus and Ricardo on Economic Methodology. History of Political Economy, 28/3(1996):475-511.

Demsetz H. Information and Efficiency: Another Viewpoint. Journal of Law and Economics 12(1969):1-22, in Schnellenbach, 2005.

Diamond J. Establishing a Performance Management Framework for Government. Instituto de Estudios Fiscales, Presupuesto y Gasto Público, Working Paper 40(2005):159-83.

Donzelot J. 1991. The Mobilisation of Society, in Burchell et al., p. 169-79.

Dopfer K. 2006. The Origins of Meso Economics Schumpeter's Legacy. Jena: Max Planck Institute of Economics, Evolutionary Economics Group, The Papers on Economics and Evolution No. 610, 44 pp https://papers.econ.mpg.de/evo/discussionpapers/2006-10.pdf [IV/07].

Dopfer K., J. Foster, J. Potts. Micro–meso–macro. Springer, Journal of Evolutionary Economy 14/3(2004):263–79, http://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID721599_code114131.pdf?abstractid=721599&mirid=1, [VI/07].

Easterling W.E., K. Kok. Emergent Properties of Scale in Global Environmental Modelling – Are There Any? Integrated Assessment 3/2–3(2002):233–46, http://www.ingentaconnect.com/content/routledg/nias/2002/00000003/F0020002/art00012;jsessionid=18gjekrjrn3oh.alice, [XII/06].

Ekins P. 1992. A Four-Capital Model of Wealth Creation; in Real-Life Economics: Understanding Wealth Creation (in Ekins P., M. Max-Neef, eds.). London, New York: Routledge, p. 147-55.

Ekins P., Medhurst J. The European Structural Funds and Sustainable Development: A Methodology and Indicator Framework for Evaluation. Sage: Evaluation, 12/4(October 2006):474-95.

Ekins P., Medhurst J. 2003. Evaluating the Contribution of the European Structural Funds to Sustainable Development. Presented at the 5th European Conference on Evaluation of Structural Funds, Budapest, June 26-27, 48 pp, http://europa.eu.int/comm/regional_policy/sources/docgener/evaluation/rado_en.htm, [XI/04]

Elbers C., J.W. Gunning, K. de Hoop. Assessing Budget Support with Statistical Impact Evaluation: a Methodological Proposal. Amsterdam: Tinbergen Institute, Discussion Paper 2007-075/2, 29 pp., http://ssrn.com/abstract=1016945, [VII/08].

ESPON - 3.2, Vol. 5. 2006. Spatial Scenarios and Orientations in Relation to the ESDP and Cohesion Policy - Territorial Impact Assessment, Final Report, p. 2-97, http://www.espon.eu/mmp/online/website/content/projects/260/716/file_2786/fr-3.2_April2007-full.pdf, [IX/07].

European Commission. 2001. SEA Directive 2001/42/EC on the assessment of the effects of certain plans and programmes on the environment, http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32001L0042:EN:NOT, [X/06].

European Commission. 2006. Impact Assessment Guidelines. SEC(2005)791, http://ec.europa.eu/governance/impact/docs/key_docs/sec_2005_0791_en.pdf, [X/06].

Foster J. 2004. Why is Economics not a Complex Systems Science? The University of Queensland, School of Economics: Discussion Paper No. 336, 38 pp, http://www.econ.iastate.edu/tesfatsi/macrocas.foster.pdf [II/10].

Foster J., Potts J. 2007. A micro-meso-macro perspective on the methodology of evolutionary economics: integrating history, simulation and econometrics. The University of Queensland, School of Economics, Discussion Paper No. 343, 24 pp.

Funtowicz S.O., J.R. Ravetz. The worth of a songbird: ecological economics as a post-normal science. Ecological Economics 10(1994):197–207, in Martinez-Alier et al., 1998

Geertz C. 2000. Available Light. Princeton University Press, 271 pp., in Renselle, 2007.

GHK, PSI, IEEP, CE, National Evaluators. 2002. The Contribution of the Structural Funds to Sustainable Development: A Synthesis Report to DG Regio, EC. 2002. Volume 1-2. London, Brussels, http://ec.europa.eu/regional_policy/sources/docgener/evaluation/doc/sustainable_annexes.pdf [III/05]

Granovetter M. The Strength of Weak Ties: A Network Theory Revisited. Sociological Theory 1(1983):201-33, http://rfrost.people.si.umich.edu//courses/SI110/readings/In_Out_and_Beyond/Granovetter.pdf, [IX/09].

Hertin J., A. Jordan, M. Nilsson, B. Nykvist, D. Russel, J. Turnpenny. 2007. The practice of policy assessment in Europe: An institutional and political analysis. EU/FP6 Project MATISSE Working Paper 6, 52 pp., http://www.matisse-project.net/, [VII/08].

IAB. 2009. Impact Assessment Board report for 2008. Third strategic review of Better Regulation in the European Union. SEC(2009)55, 17 pp., http://ec.europa.eu/governance/impact/key_en.htm, [II/09].

Kimberlé C. W. 1994. Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color, in Fineman, M.A., R. Mykitiuk (eds.). The Public Nature of Private Violence. New York: Routledge, pp. 93-118.

Kuhn T.S. 1970. The Structure of Scientific Revolutions. Chicago: The University of Chicago Press, 210 pp., http://sciencepolicy.colorado.edu/about_us/meet_us/carl_mitcham/courses_taught/5110/classic_sts/structure_of_scientific_revolutions.pdf, [VII/08].

Leith C., L. von Thadden. 2006. Monetary and fiscal policy interactions in a new Keynesian model with capital accumulation and non-Ricardian consumers. Working Paper Series No 649, 42 pp, http://ssrn.com/abstract_id=908620, [VII/07].

Leontief W. Environmental Repercussion and the Economic Structure: An Input-Output Approach. Cambridge Mass.: The MIT Press, Review of Economics and Statistics, 52/3(1970):262-71.

Leopold L.B., F.E. Clarke, B.B. Hanshaw, J.R. Balsley. 1971. A procedure for evaluating environmental impact. Washington: Geological Survey Circular 645, 13 pp., http://eps.berkeley.edu/people/lunaleopold/(118)%20A%20Procedure%20for%20Evaluating%20Environmental%20Impact.pdf, [VII/08].

Lipsey M.W. Identifying interesting variables and analysis opportunities. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds). The Handbook of Research Synthesis and Meta-Analysis (2nd edition). NY: Russell Sage Foundation, 2009. Chapter 8, pp. 147-58.

List C., B. Polak. 2010. Introduction to judgment aggregation. Yale University, Cowles Foundation Discussion Paper No. 1753, http://ssrn.com/abstract=1548413, [VII/08].

Lucas R.E. Jr. Expectations and the neutrality of money. Blackwell, Journal of Economic Theory, 4/2(1972):103-24.

Martinez-Alier J., G. Munda, J. O'Neill. Weak comparability of values as a foundation for ecological economics. Elsevier, Ecological Economics 26(1998):277-86, Stagl, 2007.

Mason E. 2006. Value Pluralism; entry in Stanford Encyclopedia of Philosophy (in Zalta N.E, ed.), Stanford: Centre for the Study of Language and Information, Stanford University, http://plato.stanford.edu/entries/value-pluralism/, [II/07].

Mertens D.M. Inclusive Evaluation: Implications of Transformative Theory for Evaluation. Sage, American Journal of Evaluation, 20/1(1999):1-14.

Morgan D.L. Paradigms Lost and Pragmatism Regained: Methodological Implications of Combining Qualitative and Quantitative Methods. Sage, Journal of Mixed Methods Research, 1/1(2007):48-76.

Mouffe C. 2000. The Democratic Paradox. London & New York: Verso, in Trainor, 2008.

Mumpower J.L., T.R. Stewart. Expert Judgement and Expert Disagreement. Taylor and Francis, Thinking and Reasoning, 2/2-3(1996):191-211.

Munasinghe M. 1992. Environmental Economics and Sustainable Development. Paper presented at the UN Earth Summit, Rio de Janeiro, Washington: World Bank, Environment Paper No. 3.

Munda G. 2006. A NAIADE based approach for sustainability benchmarking. Inderscience, International Journal of Environmental Technology and Management, 6/1-2(2006):65–78, http://www.inderscience.com/storage/f113210112864597.pdf, [VI/07].

Munda G. Social multi-criteria evaluation: Methodological foundations and operational consequences. European Journal of Operational Research, 158/3(2004):662-77.

Muradian R. Ecological thresholds: a survey. Elsevier, Ecological Economics, 38/1(2001):7–24.

Nola R., H. Sankey, eds. 2000. After Popper, Kuhn & Feyerabend: Issues in Theories of Scientific Method. Australasian Studies in History and Philosophy of Science, Kluwer. http://philosophy.wisc.edu/forster, [II/08].

Ostmann A. 2006. The aggregate and the representation of its parts. Bonn: Max Planck Institute for Research on Collective Goods, Preprints of the Max Planck Institute for Research on Collective Goods 2007/11b, 38 pp., http://ssrn.com/abstract=1024681, [IX/07].

Prins B. Narrative Accounts of Origins: A Blind Spot in the Intersectional Approach? European Journal of Women's Studies 13/3(2006):277-90.

Radej B. 2006. Assessment of structural funds effectiveness on the sustainable development: Pomurje regional case study. Final report, 2006, 6th Framework program - Framework project SDRTOOLS: Regional sustainability Appraisal: Developing evaluation methods and sustainable policies (Contract SCS8-CT-2004-502485), 80 pp.

Renselle D. 2007. Review of Clifford Geertz' Available Light. Carmel: Quantonics, http://www.quantonics.com/Review_of_Clifford_Geertz_Available_Light.html, [III/07]

Rogers P. Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions. Sage, Evaluation, 14/1(2008):29–48

Rotmans J. Scaling in Integrated Assessment: Problem or Challenge? Swets & Zeitlinger, Integrated Assessment, 3/2–3(2002):266–79.

Rotmans J. Tools for Integrated Sustainability Assessment: A two-track approach. Vancouver: University of British Columbia, The Integrated Assessment Journal, 6/4(2006):35–57.

Sadler B., R. Verheem. 1996. Strategic environmental assessment: Status, challenges and future directions. Hague: Ministry of Housing, Spatial Planning, the Environment. 188 pp.

Sanderson I. Evaluation in complex policy systems. Sage, Evaluation, 6/4(2000):433–454.

Sankey H. The Problem of Rational Theory-Choice. Epistemologia, 18(1995):299-312, http://www.philosophy.unimelb.edu.au/staff/Sankey/howard/howardpaper4.PDF, [III/09].

Schnellenbach J. The Dahrendorf hypothesis and its implications for (the theory of) economic policy-making. Cambridge Journal of Economics, 29/6(2005):997-1009.

Schwandt T. A. Evaluation as Practical Hermeneutics. Evaluation 3/1(1997):69–83.

Scriven M. The Final Synthesis. Sage, American Journal of Evaluation, 15/3(1994):367-82.

Stagl S. 2007. Emerging Methods for Sustainability Valuation and Appraisal - SDRN Rapid Research and Evidence Review. A report to the Sustainable Development Research Network. Final Report (January), 66 pp, http://admin.sd-research.org.uk/wp-content/uploads/2007/04/sdrnemsvareviewfinal.pdf, [VI/07].

Stake R.E. A Problematic Heading. American Journal of Evaluation; 22/3(2001):349–354.

Encyclopedia of Evaluation (#Synthesis). 2004. SAGE Publications. http://www.sage-ereference.com/evaluation/Article_n536.html [IV/10].

Trainor B.T. Politics as the quest for unity. Perspectivism, incommensurable values and agonistic politics. Sage, Philosophy & Social Criticism, 34/8(2008): 905–24.

UNCED. 1992. Agenda 21. United Nation Conference on Environment and Development, http://www.un.org/esa/sustdev/documents/agenda21/index.htm, [IX/06].

Veen A. Van Der, H. S. Otter. Scales in Space. Swets&Zeitlinger, Integrated Assessment, 3/2–3(2002):160–6.

Vernon R. The »Great Society« and the »Open Society«: Liberalism in Hayek and Popper. Canadian Journal of Political Science/Revue canadienne de science politique, 9/2(1976):261-76, http://www.jstor.org/stable/3230923, [III/09]

Virtanen P., P. Uusikylä. Exploring the Missing Links between Cause and Effect A Conceptual Framework for Understanding Micro–Macro Conversions in Programme. Sage, Evaluation, 10/1(2004):77–91.

Wacquant L. 1997. Reading Bourdieu's "Capital". Foreword to the English-language translation of The State Nobility. Cambridge: Polity Press, http://www.homme-moderne.org/societe/socio/wacquant/capital.html, [I/07].

Wallerstein I. 2004. The Uncertainties of Knowledge. Philadelphia: Temple University Press, 224 pp.

WCED. 1987. Our Common Future. Oxford and New York: Oxford University Press, World Commission on Environment and Development – WCED, 400 pp, http://www.un-documents.net/wced-ocf.htm, [IX/06].

Weaver P.M., J. Rotmans. 2006. Integrated Sustainability Assessment: What? Why? How? EU/FP6 Project MATISSE, Working Paper 1 (October), 22 pp., http://www.matisse-project.net/projectcomm/uploads/tx_article/Working_Paper_1_03.pdf, [VII/08].

Wiggins D. 1997. Incommensurability: Four Proposals, (in Chang R. (ed.), Incommensurability, Incomparability, and Practical Reason), Cambridge, Mass.: Harvard University Press, http://plato.stanford.edu/entries/value-incommensurable, [VII/08].

COLOPHON

About the Slovenian evaluation society

Society has been established in 2008 with the vision to affirm independent evaluation of complexity social phenomena. Society operates as a platform of the civil society. Its work is divided between three permanent commissions: for ethical codex and evaluation standards; for meta-evaluation; and for evaluation studies.

About the SES's Working papers series

Working paper series is freely accessible on the internet. It publishes scientific and technical papers from the aspect of evaluation of public policies and from related disciplines. Papers are reviewed and bibliographic catalogued (CIP) ensuring that information is available in all catalogues in Slovenia and elsewhere. Already published in SES/WP:

Volume 1:

1/2008, Excercising Aggregation (In Slovenian; B. Radej, 23 pp).
2/2008, Synthesis of Territorial Impact Assessment for Slovene Energy Programme (In Slo.; B. Radej, 43 pp)
3/2008, Synthesis in Policy Impact Evaluation (previous title: Meso-Matrical Synthesis of the Incommensurable) (B. Radej, 20 pp, revised April 2010).

Volume 2:

1/2009, Anti-systemic movement in unity and diversity (B. Radej, 12. pp)
2/2009, Meso-matrical Impact Assessment - peer to peer discussion of the working paper 3/2008; report for the period 8/08-2/09 (B. Radej, ed., 30. pp)
3/2009, Turistic regionalisation in Slovenia (in Slovenian, J. Kos Grabar, 29 pp).
4/2009, Impact assessment of government proposals (In Slovenian; B. Radej, 18 pp.)
5/2009, Performance Based Budgeting (In Slovenian; B. Radej, 33 pp.)

Volume 3:

1/2010, Beyond »New Public Management« doctrine in policy impact evaluation (B. Radej, M. Golobic, M. Cernic Istenic, 25 pp)
2/2010, Basics of impact evaluation for occassional and random users (In Slovenian; B. Radej, 43 pp.)
3/2010; Multi-level aspect of social complexity and its relevance for methodology of social research (In Slovenian; B. Radej, 59 pp.)

About the author

Bojan Radej is a methodologist in social research from Ljubljana. Master degree in macroeconomics, University of Ljubljana - Faculty of Economics (1993). Professional Experience Record: Governmental Institute of Macroeconomic Analysis and Development (1987-04; under-secretary to the government), areas of work: sustainable development (1998-04), chief manager of the modelling department (1993-5); initiator and the first editor of Slovenian Economic Mirror (1995-8); Co-Editor of IB journal (2001-04). Member of Slovenian Evaluation Society.

Publisher:	Slovensko društvo evalvatorjev.
Adress:	info@sdeval.si, Tabor 7, 1000 Ljubljana, Slovenija (c/o Radej)
Desgign of logo:	Naja Marot, UIRS, naja.marot@uirs.si
Supporters:	Inštitut za ekonomska raziskovanja, Ljubljana http://www.ier.si/