Aggregation Problem

'***We used to think that if we knew one, we knew two, because one and one are two. We are finding that we must learn a great deal more about 'and'.***'
Arthur Eddington (1882–1944)
[English astronomer, in Mackay, 1977].

1 Aggregation Problem

Complex society rests on diverging principles of legitimacy and social primacy (Wacquant, 1997). Consequently, there is a plurality of legitimate perspectives involved in the valuation of collective concerns: they are all rational but not reducible to a single 'correct' view (Ravetz, 2003) and to the same common denominator. Kenneth Arrow (1951) had proved the logical impossibility of devising a democratic and maximally rational collective choice between multiple alternatives because many individual and collective choices are not logically consistent but contaminated with irrationality. In turn, in a complex social system we have an aggregation problem (Simon, Ando, 1961; Foster, Potts, 2007). It increasingly paralyses society from organising effective collective action such as in pursuing sustainable development between its incommensurable domains, economic, social, and environmental.

Coleman traces the roots of the aggregation problem in social sciences to the philosophies of Hobbes, Smith, Locke, Rousseau, and Mill from the seventeenth to the nineteenth centuries (in Morçöl, 2012). 'In all these philosophies the central question was how purposeful actions of individuals were connected to macrosocial phenomena. How do acts of individual actors lead to macro events? How dissatisfaction becomes revolution, individual fears turn into panic, individual aspirations into market demand or supply' (Ibid.)?

An explanation of the laws that link the individual to the collective remains one of the least developed aspects of the social sciences (Coleman, in Åberg, 2000). Leaving out a big-picture vision about large scale social matters and confining the analysis, either to an overly disaggregated view or to a too narrowly aggregated view, is detrimental to the possibility of their consistent comprehension. All sciences reveal a micro-macro divide, and even the most advanced among them have not reconciled the two levels theoretically (Turner, 2006). List and Polak (2010) furthermore observe that the problem, in general, remains unresolved, as the literature has mainly concentrated on how to avoid it. The aggregation problem is also generally absent from conventional textbooks (Elsner, 2007). The social sciences have largely 'ignored the problem and instead focused on individual actions or studied macro processes and structures only' (Morçöl, 2012).

One area of applicative social research that is especially hurt by the unresolved aggregation problem is the evaluation of multi-level and multi-domain impacts of policy interventions (Virtanen, Uusikylä, 2004) on the general public good. To evaluate (governmental) policy impacts is to collect detailed factual evidence of the performance of numerous independent policy measures assessed against numerous independent evaluation criteria. Standard approaches usually evaluate impacts only with intra-sectorial appraisals, such as when assessing the economic impacts of economic policies. These imply a commensurable situation, which allows for simple accumulation of assessed detailed impacts into an aggregate impact indicator.

A given policy achievement depends not only on its own efforts but also on the side-effects of many other policy interventions that take place simultaneously with completely different goals as well as with independent evaluation criteria. Separately implemented sectorial policies either hinder, reinforce or ignore one another's core aspirations. Their side-effects trigger non-linear

situations, in which the whole can be smaller or larger than the sum of objectively measured achievements of all intervening policies.

In complex conditions, the summation procedure is far from trivial (Veen, Otter, 2002). The impacts obtained through evaluation on different levels and in different domains do not add into an aggregate indicator of the overall impact, because impacts are not expressed in terms of one another (Gutiérrez et al., 2013). There is no uniform basis for comparison among diverse policy achievements. Yet without an aggregate measure of policy impacts, one cannot know how to address policy challenges with structural consistency, such as what should be prioritised and how to enhance synergies.

The importance of consistent aggregation has already been recognised in policy impact evaluation (Rotmans, 2002). The Impact Assessment Board (IAB, 2009) that advises the European Commission estimated that the majority of evaluation studies provided to the Commission supply the kind of information that fails to inform policy makers about the overall results of their policies. Policy impact evaluation has particularly failed to produce forms of knowledge that inform strategic policy considerations. Compatible conclusions follow from Hageboeck et al. (2013) in their meta-evaluation of 340 USAID evaluation reports, and in Huitema et al. (2011), who assessed 260 evaluation studies prepared for the needs of the EU climate policy. Standard evaluation approaches, in the EU in particular, struggle to appropriately summarise partially compatible policy impacts on diverse evaluation criteria, such as the Strategic Impact Assessment (2001/42/EC), Impact Assessment Guidelines (SEC(2005)791), and the Territorial Impact Assessment (TIA; ESPON - 3.2, 2006; next Case Study).

The lack of explicit justification of the aggregation procedure is the Achilles heel of evaluation efforts (Scriven, 1994). Unsurprisingly, there is widespread recognition of the failure of policy impact evaluation to live up to its promise of improving public governance and contributing to consistent efforts to enhance the general welfare of societies. This calls for a review of the foundations of aggregation methodology in policy impact evaluation (Scriven, 1994) as well as in the methodology of social research in general.

Methodologists in policy impact evaluation have proposed a wide range of substitute approaches to the aggregation of diverse assessments that lack a common denominator. They have developed various procedures for ‘qualitative synthesis’, such as meta-narrative synthesis, meta-study, meta-ethnography, thematic synthesis, textual narrative synthesis, and framework synthesis (Barnett et al., 2009; Gerrits, Verweij, 2015). These approaches aim to narrow the aggregation problem by adopting strategies of avoidance (Barnett et al., 2009). Some authors try to reinvent commensurability by constructing overarching meta-categories as substitute common denominators or developing synthetic constructs, frameworks of synthesis, and meta-summaries. On other occasions, the possibility of synthesis is ‘recovered’ by instituting ‘reciprocal translation algorithms’, by exploring contradictory claims in inputs to synthesis, by replacing quantitative synthesis with critical interpretative synthesis, or by replacing aggregating data with comparative understanding (Ibid.).

Yet the core aggregative challenge remains unresolved in substitute approaches as long as they do not respond to the fundamental question: how do we logically consistently aggregate analytically obtained quantitative results with dissimilar common denominators without wasting relevant information? More specifically, which is the most inclusive approach to synthesis in exclusive conditions of social complexity? The new aggregation methodology must specifically explain how to incorporate horizontal diversity into an aggregation algorithm as an essentially vertical procedure. These are the main challenges of the first Case Study.

It is comprised of a prospective (ex-ante) evaluation of the sustainable impact of a regional development program for the Slovenian North-Eastern region Pomurje for 2007 – 2013 (RP;

Radej, 2006). The study evaluates whether the submitted Program proposal is likely to contribute to a reversal of critical regional trends to meet sustainable development goals. Evaluation is accomplished by methodologically comparing three alternative methods of summation of detailed assessment findings into evaluation results: descriptive summation in a disaggregated or micro level approach; a highly aggregated or macro level approach; and an intermediate, only partially aggregated or meso level approach. The expectation is that different aggregation approaches will bring about dissimilar evaluative conclusions and lead to diverse and possibly even contradictory policy advice (Gutiérrez et al., 2013), especially at the strategic level of considerations.

Pomurje is the least developed Slovenian region, bordering with Croatia, Hungary and Austria. It covers 6.6% of the national territory, 5.5% of the national population and contributes 4.3% of the gross domestic product (GDP) while achieving less than 70% of the national average income per capita. The region has a strong cultural identity, with a very distinctive regional dialect and unique ecological amenities. More than a third of its territory is protected as landscape parks or nature reserves, including the melancholically poetic landscape along the Mura River.

The region's economic capital, which is mainly employed in agriculture and food processing, is fragile but has been improving since the 1990s. The region's social capital is very frail and depleting even further. Cold War borders surrounded the region for half a century – from the West (Austria) as well as from East (Hungary) – this is the only region of this kind in Central Europe. Due to the geostrategic realignments in Central and Eastern Europe at the beginning of the 1990s, the region suddenly found itself exposed to the main European transport corridor, which subjected it, unprepared, to intensive international flows of capital, goods, and people between Eastern and Western Europe. The accession of Slovenia to the EU in 2004 further imposed a more stringent border regime between Pomurje and the bordering region in the Republic of Croatia (at that time not yet a member of the EU), in this way restraining links between local communities on both sides of the border that had previously been traditionally close because of similar socio-economic and cultural conditions. Negative social shocks of the economic transition have further weakened the region's social and human capital, leading to continued depopulation, brain drain, long-term unemployment, as well as prolonged health and social risks, so that the majority of the regional population is officially classified as vulnerable (older population, long-term unemployed, women, ethnic minority, and poor).

Lags in structural development have accumulated in the region, especially in its social capital, despite the considerable inflow of investment from the national budget and EU funds in the past decades. Investments supported many new development projects in less advanced regions, but in the case of Pomurje, their priorities largely overlooked genuine local needs and local potentials. Decision-makers especially failed to address critical trends and trade-offs between domains of regional sustainability appropriately.

2 Core Concepts

The purpose of synthesis is to render a heterogeneous corpus of information sensible (Encyclopaedia of Evaluation, 2004) from a more general viewpoint. Standard accounts of evaluative synthesis comprise a set of methods that aim to consolidate multiple valuations into a single valuation that represents the whole.[1]

[1] https://docs.oracle.com/cd/B19306_01/olap.102/b14349.pdf, III 2013.

Aggregation is the simplest method of direct micro to macro synthesis, aside from multiplication, standardisation, correlation, and integration. Aggregation demands only cumulative additions of similar elementary parts to a pile. The imperative of similarity requires strong symmetry and homogeneity between aggregated elements to secure the conservation of the qualities in transition between the lower levels of the elements to the higher level of the result – so that adding an apple to other apples cannot result in pile of oranges. 'Being symmetric' in general usage is synonymous with 'being harmonious', and in technical usage it is synonymous with 'commensurable', denoting that elements are of the same essence and so aggregatable by this essence.

When diverse social values translate into one ultimate value, to a single quantum of measurement, such as happiness, utility or money, then more means better, so our task is simple: aggregate and maximise (Scharffs, 2001)! When quantity is quality, there can be no conflict between quantity and quality. Trust in the neutrality of simple aggregation logic has led many social scientists to constitute this approach to synthesis as a guideline for consistent policy-making. The founder of modern utilitarianism, Jeremy Bentham, claimed (1789, in 'The Principles of Morals and Legislation') that the interest of the community could be neither more nor less than the sum of the interests of members who compose it. He proposed that governments pursue the maximisation strategy of 'the greatest happiness' principle as the fundamental measure of right and wrong in every society.

However, if all social values are reduced to a commensurable relation, values are no longer about diverse universals but simply a matter of calculation (Isaiah Berlin, in Scharffs, 2001). The assumption of commensurability between elementary parts as a precondition for forming wholes is generally true for physical entities in expressing their quantitative characteristics such as size, weight or speed. Yet it may not be true for social units (Scholes et al., 2013), which have a far less deterministic structure because they are heavily 'contaminated' with discordant value-based considerations which reflects in their incommensurability.

Furthermore, at its core the simple aggregation algorithm demands merely repetition of the elementary principle at a macro level. An aggregate is not qualitatively different and therefore not independent of its parts. A conventional approach to aggregation does not produce unity since unity is already involved in the parts, in their denominator. A new quality, such as 'collective novelty' (Perry, 1922) or a new wisdom about the whole cannot emerge from the aggregation of commensurable content (Allport, 1928). Simple aggregation is tautological (Ibid,), and relatively uninteresting (Wimsatt, 2002) for the study of multi-level and multi-scope social challenges.

The main obstacle to simple maximizing strategies is that it will usually be impossible to settle on which core value should be the one that is maximised (Scharffs, 2001) and which specific observed characteristic of reality should be taken as the commensurable representative for all others. In contemporary postmodern societies, there is no common good about which everyone agrees 'because ultimate values – our conceptions of what life is and what society should be – are beyond the range of mere logic' (Schumpeter, in Coleman, 2005) and calculation. Schumpeter accordingly refused Bentham's maximisation doctrine. The synthesis of social valuations needs to apply a different methodology in which detailed elements may not be commensurable and not fully aggregatable from a micro to a macro level.

The methodology of social research has developed a range of substitute approaches to overcoming the aggregation problem. Three main lines of approach can be distinguished for the narrower purposes of this inquiry: the non-aggregative, the fully aggregative and the intermediate one.

The first approach covers only methods, which do not permit cumulating detailed findings into aggregate conclusions whenever findings in one domain of research are not fully comparable in their core with findings in all other domains. The analytical results for each research domain must then stand on their own. They represent only themselves, leaving the researcher without clear conclusins about the overall result of inquiry.

A family of 'compensatory'[2] aggregative approaches is a representative of the second group of methodologies. In these, no specific value prevails in aggregation simply because not all values involved in social inquiry are treated as radically different but as equally relative. Compensatory approaches permit aggregation so that substantively diverse issues are first translated into a 'neutral form', for instance standardised into index numbers, and then aggregated into a composite index that has no common units of measurement (Nardo et al., in Zhou, Ang, 2009), as in the previously discussed methodology of calculating the UNDP's Human Development Index.

Another variation of the compensatory method relies on the application of social weights. For instance, higher weights are attached to environmental issues than to economic issues when applying the precautionary ecological principle. In other cases, weights are proportional: each preference in a collective choice has a weight equal to the proportion of the population supporting the values represented by each alternative option. Economists have developed spectre of techniques for smoothing out conflicts in collective choice at a lower level in order to achieve a maximal outcome at a higher level, such as by compensating losers for negative trade-offs (Pareto), or by weighting collective alternatives relative to a willingness to pay (Baumol, Oates, 1988).

Methodologies aimed at full aggregation of only partly commensurable analytical insights have offered valuable assistance especially in situations when underlying incompatibilities in measured phenomena remain small. Nevertheless, there are some critical concerns. Even though it is sometimes possible to recover, at least partially, some sort of commensurability between diverse qualities, a large number of additional parameters, such as trade-off coefficients and weights, must first be constructed, which may cause a loss of transparency and consistency in the aggregation model (Munda, 2004). In such cases, the substantive concerns about a researched matter diminish on account of increasing technical and statistical concerns about how to collect ideally fitting data and how to calculate the most appropriate version of statistical parameters. Furthermore, parameters may be very difficult to establish scientifically and are often selected ad hoc (Ibid.). In one way or the other, substitute strategies to aggregation not only contaminate with the same type of value judgment contradictions that restrained the implementation of simple aggregative methodology from the outset but also hide their bias behind statistical formalisms.

The third main group consists of intermediate approaches with multi-criteria methods. The multi-criteria analysis describes any structured approach to determining an overall picture from divergent contributions. The methodology considers separate aspects of inquiry in order to capture all domains of research indiscriminately (Söderbaum, 1998). However, due to the application of separate conflicting analytical criteria, multi-criteria methods do not always have a conclusive or unique solution. Even if they do produce clear-cut results, these comprise a limited set of interrelated aggregate propositions aimed at understanding only limited topics (Geels, 2007; Merton, 1968) but failing to address structural contradictions.

[2] Non-compensatory approaches of aggregation can aggregate the corpus of information by relying on only one set of assessment results, such as the best or worst achievements between multiple criteria (see Munda, 2012).

Analogous to diversity in a methodology of social research, Söderbaum (1998) distinguishes between three sets of multi-criteria policy impact evaluation methods according to the degree of aggregation of detailed impacts assessment results: disaggregated, aggregated and intermediate methods.

Examples of disaggregated methods are monitoring, environmental impact assessments, or the Leopold assessment matrix. Leopold et al. (1971) proposed a detailed impact assessment method at the micro level from which an aggregation of fragmented results remains absent as a matter of principle. They developed an approach that goes beyond intra-domain assessment about how a particular policy intervention, such as an economic one, directly affects primary targeted economic assessment criteria. Leopold et al. instead placed the horizontal assessment of side-effects or unwanted (secondary) effects in the centre of evaluation concerns, specifically economic effects on the environmental assessment criteria. [3] This approach surpasses, up to the present day, the scope of many prominent evaluation guidelines since they are mainly concerned only with an assessment of direct impacts. Unfortunately, however, Leopold et al. decisively refused to capitalise on their methodological breakthrough.

In the finest positivist tradition of analytical assessment, Leopold et al. assess policy performance through a pedantic description of its numerous side-effects. Their matrix presents 100 most polluting industrial sectors in rows and 88 evaluation criteria of environmental impact in columns, creating a matrix with 8,800 cells – each further divided into four sub-sections that describe every impact by its size (large/medium/small), direction (positive/negative/neutral), probability (high/low) and the assessment of risk involved in each individual impact (critical or not).

They assessed policy impacts in sufficient detail to enable maximally informed policy decisions. For them, the task of the evaluator is merely to inform and comment on specifics, not to generalise. Refusal of aggregation of detailed assessments is essential for neutral evaluation, argues Leopold, as it draws a demarcation line between evaluator and policy-maker to protect the former from value judgments and political interference (Kunseler, 2007). Since Leopold, evaluators largely resist calls to generalise their findings, or at best they take a passive descriptive approach to synthesis, such as the EU's Impact Assessment Guidelines (SEC(2005)791), one of the most influential reference documents in the EU.

Leopold's approach explicitly refuses to deal with value-based oppositions so it must be characterised as an example of analytical methodology that represents essentially a non-

---

[3] The methodology of the Leopold matrix was introduced at the end of 60's. It was applied in Slovenia already in 1970 by biologist Stane Peterlin (1937-). According to his own explanation (in an e-mail on 1.XI.2010 to his son, Marko Peterlin, who kindly forwarded it to the author), he found Leopold's paper '*Landscape Esthetics: How to Quantify the Scenics of a River Valley*', published in Natural History Journal in 1969 (vol. 78, pages 36–45). Mr. Peterlin applied Leopold's assessment approach at the Slovenian agency for heritage protection, together with Franc Vardjan, architect, and Milan Orožen Adamič, geographer, in the environmental impact assessment of seven alternative locations for the planned construction of a hydroelectric plant in Kobarid on the Soča river, one of few remaining pristine rivers in Europe with breathtaking visual attractiveness. Mr. Peterlin cannot confirm the chain of events that followed their negative assessment of the plans, but the fact is that the Soča river remains undisturbed to the present day in its natural beauty. The Leopold matrix was soon applied again in the environmental impact assessment of plans to build an oil refinery on the outskirts of Ljubljana, the capital of the country (Peterlin S. et al. 1972. *Ocena možnih lokacij za rafinerijo nafte v okolici Ljubljane s stališča varstva oklja, posebej varstva narave*. Zavod za spomeniško varstvo SR Slovenija, as stated in Marušič J. 1993. *Optimizacijski postopki kot sredstvo za vključevanje varovalnih presoj v celokupno in z okoljem skladno prostorsko načrtovanje*. Ljubljana: Biotechnical Faculty UL, Department of Landscape Architecture). The refinery project was also abandoned, as proposed by the authors of the evaluation study. I am grateful to Mojca Golobič for this information.

evaluative approach to policy impact evaluation. Non aggregated assessment produces non-overlapping information separately for each evaluation domain, which tends to underplay inherent system contradictions in disregard of legitimate stakeholder concerns in the public realm (Stake, 2001). Sanderson (2000) warns that cross-sectional evaluation, which seeks to isolate policy instruments and assess their impacts separately from one another, will only produce an information overload. Detailed assessment results undoubtedly make policy decisions more informed but not necessarily easier (Diamond, 2005). Non aggregated results remain of limited usefulness due to limited correspondence to the complex reality in which policy-makers operate. They especially fail to satisfy information needs at the strategic level of decision-making by producing banal answers to 'cross-cutting' and multi-dimensional questions (Virtanen, Uusikylä, 2004).

Rejection of the summation of findings and shifting this task to policy-makers, the media and politicians is dubious (Stiglitz et al., 2009). Leaving it up to the users of evaluation results to make the synthesis supposes that they are equipped to do this consistently and neutrally, which can be difficult to justify, taking into account their ideological, cultural, ethnic, social and other value based predispositions. It is, after all, precisely the failure of politicians and scientists as social aggregators, recall Arrow's theorem of impossibility that demanded the introduction of policy impact evaluation in collective choice in the first place. The absence of evaluative synthesis findings leaves the interpretation of results fully exposed to manipulation. Scriven (1994) appropriately concluded that rejection of summation in impact assessment means 'letting the client down at exactly the moment they need you most'.

At the other extreme are highly aggregated evaluation approaches, such as cost-benefit or cost-effectiveness analyses, or experimental methods that apply counterfactual evidence to identify core drivers of policy success and measure its effectiveness. These approaches are useful especially when a consensus about 'specific valuation rules' (Söderbaum, 1998) has been achieved.

One of the first declared macro-evaluation methods with multiple domains in the EU was the Strategic Environmental Assessment (SEA Directive, 2001/42/EC). It assesses all strategically important environmental aspects of economic projects at macro level but only piece by piece, giving no indication about how to cumulate diverse economic impacts on various environmental criteria into a compound indicator of overall environmental impact.

The missing chain in the aggregative methodology of evaluation has been contributed by Ekins and Medhurst (2003, 2006) in their multi-domain assessment of the EU Structural Funds programs' impacts on sustainable development in regions (sub-national level). Ekins previously proposed the Four-capitals model (economic, social, environmental, and human capital) to serve as a conceptual framework for a more structured comprehension of sustainable development (Ekins, 1992). The multi-part concept of sustainability can be traced back to the Brundtland report (WCED, 1987), and to the first global conference on sustainable development in Rio de Janeiro (UNCED, 1992), where Paul Ekins from UK's Policy Studies Institute and Professor of Sustainable Development at the University of Westminster, and Mohan Munasinghe, high ranking advisor at the World Bank from Sri Lanka, independently proposed an analogous idea.

Ekins and Medhurst worked out a partially aggregated version of the Leopold matrix, extended from two (economy, environment) to four domains to evaluate, in the accomplished Case Study, the RP's multifarious impacts on the four capitals of regional sustainability. The extended impact assessment matrix is named here the Leopold-Ekins-Medhurst matrix or LEM. It allows for aggregation of assessed impacts for all of the program's intervention measures within each of four evaluation domains that are placed in its columns, but not between them, appropriately accounting for their incommensurability. The EU's territorial impact assessment (ESPON – 3.2,

2006) has given grounds for developing an analogous approach to evaluative synthesis (see the subsequent Case Study).

However, in their aggregative approach Ekins and Medhurst overlooked the fact that various policy impacts on the same evaluation criterion may be also incommensurable. Policy impacts are aggregatable only by intervention measures with similar impacts so that, for instance, the impacts of economic intervention measures are aggregated separately from the impacts of social measures. This means that only partial aggregation of impacts is permitted between four evaluation domains of sustainability. Following this demand, LEM must be reorganised into an input-output matrix, composed of equal number of program intervention domains (in rows) and evaluation domains (in columns).

This is the meso matrix. It hierarchically exists above the micro level of the Leopold matrix because it aggregates from it. At the same time, as a set of only partial aggregates, it exists below the macro level of the LEM. The meso matrix, therefore, comprises an intermediate level of policy impact evaluation. It is a relevant tool since complex social issues, like sustainable development, are themselves built upon multiple domains as meso level units of evaluation (Dopfer et al., 2004; Easterling, Kok, 2002).

The meso matrix aggregates its contents from the inside, only by main sources and by main areas of impact, instead of bottom up (micro to macro) as in LEM. The matrix has very desirable properties for evaluation of impacts because it differentiates between the primary concerns (direct impacts on the negative diagonal of the meso matrix) and secondary concerns (indirect impacts, non-diagonally placed) so that one can simultaneously evaluate impacts in an intrasectorial (non-overlapping) as well as in an intersectorial (overlapping) perspective.

The Case Study illustrates the aggregation problem in policy impact evaluation, first only in the standard vertical manner, aggregating findings from micro to macro and then also in the newly proposed mesoscopic manner that integrates horizontal logic into evaluative synthesis.

3 Results: Micro vs Macro

An interdisciplinary group of experts estimated the possible positive, neutral or negative impacts of the 47 proposed intervention measures of the Regional Program (RP) on four sets of two evaluation criteria for social, economic, environmental, and human capital of regional sustainability (Table II.1). In an ideal case, assessment of detailed impacts would be acquired analytically from detailed sectoral research, from surveys among stakeholders and from statistical data (such as in Radej et al., 2015). When this is not feasible, usually because of lack of resources, or in ex-ante evaluations that assess future effects, as in our case, the task may be accomplished with ad hoc expert assessment of impacts. It is important to ensure at least that the expert assessments draw on the best available knowledge in concerned thematic areas and that their expertise be sufficiently diverse to cover all evaluation domains indiscriminately.

Expert opinions about the direction of the RP’s impacts were considerably divergent in some instances. Reassessment of discordant impact assessments resolved only about a third of the expert’s disagreements. The remaining disagreements were neutralised by averaging of assessments into a summary indicator of impact. The threat of cancelling-out of divergent assessments in itself encouraged a cooperative attitude in the team of experts. Contrary to this solution, some evaluation approaches, such as the Common Assessment Framework (EIPA, 2006; Macur, Radej, 2017) advise that experts reach a consensus in the assessment of impacts. However, forcing consensus for every assessment detail is risky because it may expose existing asymmetries within the assessment team, which can lead to a kind of closed, exclusive process (Connelly, Richardson, 2004) in which the dominant agent will prevail.

Disagreements between experts may be irresolvable but this does not mean they are unaggregatable. Detailed assessments are not only equally well founded and legitimate, but also very specific and incomplete claims, and therefore not eligible for the status of strict incommensurability, as wrongly assumed by Leopold in a methodologically analogous situation. Nevertheless, disagreements in expert impact assessment deserve additional attention, separately from an aggregation of results, if they are systematic across connected issues.

Table II.1 presents the Leopold matrix with detailed assessment of the impacts of the Pomurje Regional Program on selected regional sustainability evaluation criteria.

Table II.1: RP's impacts – presented at micro level (Leopold's matrix)

Evaluation Domains and Criteria / Regional Program Measures	Economic		Social		Environmental		Human	
	Income growth	Invest. Intensity	Unemp-loyment	Outward Migration	Abatement Expenditure	Sewerage Connections	Students per capita	Regional Ageing
1 Development lags	+	+	0	-	+	+	+	0
2 Competitiveness	+	+	-	-	-	+	+	0
3 Investment promotion	+	+	+	+	0	+	0	0
4 Endogenous advantages	+	+	+	+	+	+	+	+
5 Entrepreneurship	+	+	-	+	0	0	+	0
6 Regional tourist organisation model	0	0	0	0	0	0	0	0
7 Pomurje as a tourist destination	+	0	0	+	0	0	+	0
8 Destination management	0	0	0	0	+	+	+	0
9 Destination marketing	+	0	0	+	0	0	0	0
10 Human resources in tourism	0	+	+	+	0	0	+	0
11 Quality management	0	+	0	0	+	+	+	0
12 Tourist infrastructure investment	+	+	0	0	+	+	0	0
13 R&D in tourism	+	+	0	0	0	0	+	0
14 Health inequality (criteria)	0	0	0	0	0	0	0	0
15 Health promotion network	+	+	+	+	0	0	+	0
16 Health inequality – regional	+	0	+	+	0	0	+	+
17 Health inequality–vulnerable groups	+	0	+	+	0	0	+	+
18 Quality, access to health	+	0	0	0	0	0	+	+
19 Healthy environment	0	+	0	0	+	+	0	0
20 Mental health	0	0	+	+	0	0	0	+
21 Agriculture modernisation	+	+	-	-	+	+	+	0
22 Environmental agriculture	+	+	+	+	+	+	+	0
23 Entrepreneurship in agriculture	+	0	+	0	0	0	+	0
24 Human development in agriculture	0	0	+	+	0	0	+	0
25 Value added growth	+	+	-	-	+	+	0	0
26 Products, services – farms	+	+	+	+	+	+	+	+
27 Products, services-agro industry	+	+	-	0	+	+	+	0
28 Marketing agro-products	+	+	0	0	0	0	0	0
29 Rural developm., products, services	+	+	+	+	0	0	+	+
30 Countryside development	+	+	+	+	+	+	+	+
31 Rural entrepreneurship	+	+	+	0	0	0	+	0
32 Rural stakeholders' co-operation	+	0	+	+	0	0	+	0
33 Water supply	+	+	+	0	+	+	0	0
34 Transport infrastructure	+	+	+	+	+	0	0	+
35 Alternative, local energy	+	+	+	+	+	0	+	0
36 Energy distribution network	+	+	0	0	0	0	0	0
37 Access to IT services	+	+	+	+	0	0	+	+
38 Waste waters treatment	+	+	+	0	+	+	0	0
39 Solid waste management	+	+	+	0	+	+	0	0
40 Communally equipped zones	+	+	+	+	+	+	+	0
41 Water quality	+	+	+	0	+	+	0	+
42 Revitalisation of hot-spots	-	-	0	0	+	+	0	0
43 Illegal land-filling, monitoring	+	+	+	0	+	+	0	0

44 Nature and culture conservation	-	+	-	0	+	0	+	0
45 Energy policy	+	-	+	0	+	0	+	0
46 Spatial planning	+	+	+	+	0	0	+	0
47 Communication strategies	+	+	0	0	0	0	+	0

Source: Radej, 2006.

Legend: Scores: '+' positive impact, '0' absent or neutral impacts, '-' negative impact.

Leopold would summarise the findings descriptively. A prevalence of the program's positive impacts would suggest that a majority of the RP's interventions favourably influence the program area. Negative impacts would focus the attention of policy-makers on the weakest parts of the proposed program in all four evaluated domains that should be improved with future amendments, either by abandoning problematic interventions with the most damaging cross-sector impacts or by providing corrective actions and compensation in order to offset voluntary victims. Absent or neutral impacts (0) in evaluation are usually not seen as problematic.

A simple descriptive synthesis of evaluation findings into recommendations for improving the Program is innocently naïve, but essentially wrong in complex situations. By concentrating attention only on incidences of negative impacts, the evaluator loses sight of the program as a whole, and consequently loses sight of regional sustainability. Furthermore, 'neutral' (0) impacts are not irrelevant to evaluation since they indicate an absence of cooperation, synergy, and overlap, which are defining features of sustainability of regional development.

Neither is the prevailing positive impact authoritative evidence of an excellent program design. Suspicion is linked to several reasons. Positive results should not be taken as decisive evidence of policy success since the Regional Program is prepared by a competent institution, specialised for governing regional development in close cooperation with stakeholders and their specific needs. Furthermore, the program was carefully scrutinised before being submitted to evaluation, first by sectoral experts and then in negotiations among various group interests. This largely ensured that the institution responsible for preparation of the Program could submit a document that guaranteed a wide array of positive contributions to the resolution of major regional challenges. Policy impact evaluation is not necessary to discover such obvious things.

Positive effects are often only assessed against criteria selected by a formally responsible implementation authority, as outputs, not by end beneficiaries, as results, or by the general public (as social-wide impacts). Moreover, positive effects cannot guarantee society-wide positive impacts if policy actions or assessment criteria or both are in conflict – also a very common situation. The Regional Program as a whole gives appropriate contribution to regional sustainability only when positive impacts are systematic across all domains of evaluation and between them. However, a systematic insight arises in evaluation only when detailed assessment results, positive and negative, are properly synthesised with an algorithm that is in line with the synergetic intervention logic of sustainable development.

There are important obstacles to aggregating the positive and negative impacts of a given intervention measure on various assessment criteria or, analogously, the impacts of different intervention measures assessed against the same evaluation criterion. In aggregation, the evaluator has to take a position on the fundamental issue of compensability (Munda, 2012), i.e. the possibility of offsetting a negative impact of a given policy measure on one criterion with a positive impact on another. For instance, in Table II.1, might the positive impacts of abolishing the main development lags (intervention measure 1) on income growth outweigh in synthesis all the negative impacts of abolishing development lags on migration in Pomurje? Analogously, with more eminent example, is it permitted in principle to cancel out an additional ton of greenhouse gas emissions (a negative impact) with additional purchases of tradable pollution permits (a positive impact, because their proceedings pay for environmental protection measures at the

permit seller's plant)? Greenhouse emissions trigger irreversible deterioration in climate conditions; therefore, they cannot be interchangeable but incommensurable with environmental protection goals. Thus, a trade-off between additional greenhouse gases and money is neither adequate as a general principle in policy-making nor in an evaluation of its impacts.

However, one needs to observe the climate issue as a complex matter that operates at multiple levels with different evaluation principles relevant on each level. Trade-offs between different aspects of the public good may not be incommensurable in every single case. On a micro level, conflicts between principal valuation domains arise only in extreme situations, when breaching some normative thresholds, but not within threshold limits where, in fact, a large majority of trade-offs take place. Normative thresholds, for instance economic, ecological, and social (see Muradian, 2001), define how much of something valuable can be sacrificed with trade-off for some other valuable thing without incurring long term damage to the system as a whole. Thresholds then mark a tipping point, beyond (or below) which a small quantitative change in one part of a system might have a disproportionately large effect on the entire social system, involving high risk to its overall integrity.

The negative planetary environmental impact of each additional gram of greenhouse emissions is, of course, ignorant of the threshold limits in Slovenia or in any other country, but its wide social impacts need not be. Additional emissions and environmental trade-offs can be to some considerable extent offset, for instance, with additional forestation to absorb additional emissions, or by faster accumulation of human capital that will substitute depletable resources intergenerationally – as is assumed in the World Bank's Index of Genuine Savings (Hamilton, Clemens, 1999).

As long as threshold constraints are not violated, trade-offs on the micro level are free from macroscopic consequences. In such cases, concrete exchange rate depends only on the considerations of protagonists who are directly involved in trade-offs. Within threshold limits, an agent does not sense the qualitative difference (Luce, in Munda, 2006) involved in the trade-offs or is willing to adopt fair compensation for damage when affected as a voluntary victim. Beyond (or in some case below) normative limits, any further trade-offs are ruled out on account of systemic reasons, even if the victims' consent could be still preserved.

Regional Programs are prepared by institutions of the system so that their impacts should generally be in line with the thresholds they themselves prescribe. In turn, impacts of their interventions must be at least to considerable extent exchangeable. Normative thresholds then draw a network of internal barriers to full and uniform micro to macro aggregation. Within this network, conflicting impacts must be partially aggregatable.

It is obvious that Leopold's assessment approach with strictly demanding strong commensurability between detailed impacts as a prerequisite for synthesis is far too restrictive rule to instigate fruitful synthesis. Policy impacts that fully comply to all domain-specific threshold constraints are in fact only 'weakly commensurable', which means positive and negative impacts between two given evaluation domains are partially aggregatable by source and area of policy impacts. Accounting for the difference between strong and weak commensurability importantly improves aggregative potential in impact evaluation of complex social concerns.

A better understanding of the nature of value incommensurability in evaluation enabled Ekins and Medhurst in constructing their LEM. They realised they can summarise detailed impacts from Leopold's matrix (Table II.1) in two constrained ways: individual impacts for all evaluation criteria only within each evaluation domain, and all impacts of intervention measures within individual sectorial policies. Results of aggregation are presented in Table II.2.

Table II.2 exhibits the LEM on a selected illustrative example. For simplicity of explanation, we provisionally reduce the original Four-capitals model to only three evaluative domains, economic, social and environmental. A narrower setting is entirely sufficient to discuss the aggregation problem in the evaluation of sustainable development. We will remove this simplification later.

The number of rows in LEM reduces with aggregation from 47 program measures to only six main regional sectorial policies. LEM presents impacts on a wider scale of assessment scores, as compared to Table II.1, from the most robust positive impact with the highest score (three pluses, '+++') to the most negative impact with the lowest score (three minuses, '– – –'), with all five intermediate possibilities, including neutral or absent impact (0). When uncertain about how to round-off partially aggregated detailed impacts from Table II.1 to Table II.2, the evaluation decided in favour of the impact contributed by the financially heftier RP intervention measure (Radej, 2006).

LEM presents multifarious assessment results in an aggregated form on the level of six policy sectors and three main evaluation domains of sustainability. What are its main messages? Infrastructure development, as one of the six sectorial policies in the RP, will produce more positive effects on three sustainability domains than any other sector policy and therefore appears as the most welfare-enhancing set of intervention measures, followed by rural development policy. The most problematic is the negative impact of the manufacturing sector's intervention measures on regional social capital, mainly because they aim to increase cost efficiency in companies by lowering input costs, which results in more lay-offs, higher unemployment of the regional work force and further increased outward migration to other Slovenian regions and to other countries. The sustainable impacts of the health sector and tourism policy are largely absent – the reason being that the proposed intervention measures in the programming period generally do not relate to investments or to the provision of new services, but merely to the preparation of plans and regional organisational (infra)structures.

The direction of the RP's overall impacts on three domains of regional sustainability, as presented in aggregate in the bottom row of Table II.2 as the highest level of aggregation in LEM, do not appear too problematic. The program will improve regional sustainability in all three capitals, though a more positive impact is expected in the economic (++, two out of three possible plusses) than in social and environmental domains of sustainability (+). The summary impacts of the RP are feeble and unevenly positive, yet the differences between domains are relatively small and do not seem to defy regional ambitions for balanced sustainable development. LEM's synthesis row brings the evaluator to the macroscopic conclusion that the program will have rather fragile but systematically positive effects on regional sustainability. The RP should be eligible for adoption under the condition that observable improvements are made in the final version of the Program to assure a more positive net impact of the RP on social capital that is regionally most vulnerable. Taken as a whole, the RP appears not to be sufficiently ambitious, yet not an unacceptable development instrument for the Pomurje region.

Table II.2: RP's impacts – presented at macro level of sustainability (LEM)

Evaluation Domains / Policy Sectors	Economic (E)	Social (S)	Environment (N)
1 Manufacturing (rows 1-5)*	+++	–	+
2 Tourism (rows 6-13)	+	0	0
3 Health (rows 14-20)	0	+	0
4 Rural development (rows 21-32)	+++	+	++
5 Infrastructure (rows 33-40)	+++	++	++
6 Natural Environment (rows 41-47)	+	+	+

Summary Impact	++	+	+

Source of data: Table II.1.

Legend:

* Summary of rows 1-5 in Table II.1.

Impact Scores: '+++' Very positive, '++' Moderately positive, '+' Weakly positive, '0' Absent or Neutral, '-' Weakly negative, '--' Moderately negative, '---' Very negative impact.

What can one learn from comparing the two immoderate readings of the assessment results in Table II.1 and II.2? If detailed assessment results remain disaggregated, as in Leopold's microscopic approach, then the evaluation produces findings that are too fragmented to be able to inform regional policies about strategic concerns regarding its overall direction of development. A high level of aggregation of the assessment results solves this problem. However, a macroscopic reading of the results cannot identify structural synergies between three domains of sustainability because it cannot show how they impact one another and overlap horizontally. According to the idea of sustainable development, achieving these overlaps and appropriately synthesising their diverse messages is imperative for an overall positive evaluation of the Program.

Standard micro or macro methodologies of aggregation produce either overly fragmented or overly aggregated results due to either a too restrictive or a too expansive distinction between commensurable and incommensurable principles in the summation of detailed findings. Even though the evaluation of impacts requires a large effort to collect detailed information, the informative value of the evaluative findings for regional decision-makers remains in both cases shallow.

The aggregative challenge of social complexity demands a more subtle approach, able to negotiate a logically consistent middle ground between emphasising radical difference and need for synthesis.

4 Meso Level Results

The origin of the troubles in the aggregation procedure applied by Ekins and Medhurst is a methodologically inconsistent application of incommensurable relations. They acknowledged it in LEM only between evaluation domains but not between policy domains. As a result they failed to recognise that the impacts originating from different sectors are also incommensurable and thus not fully aggregatable by columns in LEM.

The column aggregation in Table II.2 methodologically assumes homogeneity of policy impacts – that different sectors affect given evaluation criteria in the same way and thus neutrally – which is a highly dubious conjecture. Many studies indeed demonstrate that a given policy does not influence all areas of impact in the same way (Schnellenbach, 2005). Discriminatory sectorial impact is confirmed even for those policies that had previously been thought to be the most sector-neutral, such as for monetary policy (Lucas, 1972) and for tax policy (Leith, Thadden, 2006).

Every sectorial policy affects the goals of its own sector differently than it affects the goals of other policy sectors with entirely different responsibilities. By their nature, policy impacts are either direct, when pursuing their primary goals vertically (impacts of economic interventions on economic criteria), or indirect, that is, side-effects on areas that fall under the jurisdiction of some other policy sector (Rotmans, 2006) with divergent goals and dissimilar evaluation criteria, such as impacts of economic interventions on social and environmental goals.

Policy interventions are sector-based and specialised, so one should assess their intended effects separately from their unintended side-effects. In consequence, the evaluator should take side-effects as equally important in evaluation with the direct impacts. Yet side-effects as secondary in importance often fade out in the evaluation because it is assumed that 'they are too complex' (Morçöl, 2011) and impossible to track. A meta-evaluation of USAID evaluation studies found that only 15% of the studies reported unplanned effects, and only 10% of the studies discussed secondary causes that might contribute to the results (Hageboeck et al., 2013). Huitema arrived at a similarly upsetting conclusion, discovering that some 60-80% of evaluation studies prepared for the needs of EU climate policy either avoid or attempt to diminish the complexity of the evaluated objects (Huitema et al., 2011), by focusing only on directly observable policy impacts while side-lining indirect impacts. Systematic disregard for extensive secondary impacts in policy-making might explain why well designed and implemented policies, founded on strong values and even on common sense, often lead to disappointing overall results (Chapman, 2004) for end beneficiaries and for society at large.

In response to these critical observations, a new generation of evaluation approaches (Guba, Lincoln, 1989) or a new wave of evaluation studies (Vedung, 2010) has arisen, which see evaluation as an essentially complex challenge. New approaches explicitly outline the importance of secondary impacts in policy evaluation by highlighting synergies in inter-sectorial overlaps and 'horizontal themes', such as gender equality or sustainable development, that span equally across all sector-based policy concerns.

Accounting for indirect impacts in evaluation calls for a specific aggregation procedure in two regards, following the partial aggregation rule. Firstly, summarizing indirect impacts separately from direct ones – in this way abandoning LEM's summary row in Table II.2 as inconsistent. And secondly, aggregating different types of indirect impacts independently. For example, the economic and social interventions' impacts on environmental criteria are not commensurable and therefore must be assessed and aggregated separately – economic impacts on the natural environment separately from social impacts on the natural environment.

The partial aggregation rule initially seems at odds with the incommensurability thesis. Yet value incommensurability of social facts does not imply full incomparability, it only denies comparability with the universal common denominator. Some authors argue against a generalizing concept of strong incommensurability (Nola, Sankey, 2000). They propose instead making a distinction between relations of strong and weak incommensurability. Martinez-Alier et al. (1998) specifically 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.

Two intermediate or hybrid categories exist in the middle between strong commensurability and strong incommensurability of impacts: weak commensurability and weak incommensurability. Policy impact is 'weakly commensurable' when limitations are imposed in aggregation so that it can be accomplished only partially. In the Case Study, impacts are aggregatable only within specific policy and evaluation domain within a domain-specific network of thresholds. Furthermore, impacts that are weakly commensurable in two (or more) incommensurable domains simultaneously are 'weakly incommensurable'. For instance, hybrid socio-economic impacts are weakly commensurable in the social and in economic domains of sustainability.

Recognising the existence of weak incommensurable impacts is essential because they are assessed against two mismatching sets of evaluation criteria. As such, weakly incommensurable impacts become drivers of evaluative synthesis. Munda (2016) goes so far – probably too far – as to claim that weak incommensurability needs to be the main methodological foundation for the evaluation of well-being. Mesoscopic methodology implies instead that social facts are partly aggregatable (weakly commensurable) at the first, micro-meso phase of aggregation and weakly

incommensurable only in the second, meso-macro phase of aggregation. On each level, different sorts of in/commensurability operate and they are equally important in building aggregate comprehension of complex social concerns.

The difference between strong and weak incommensurability is important so it may be helpful to illustrate it by a juicy example that shows how natural is their co-existence. It is usually not appropriate to add apples and oranges because they are two different things. It is nevertheless entirely possible to mix the juices of apples and oranges, as their secondary content. Juices are only weakly incommensurable since they can be tastefully combined in a cocktail under certain threshold conditions provided by a bartender's recipe. Despite this, however, it will still never be possible to grow an apple tree from the seed of an orange, because the seeds of apples and oranges are strongly incommensurable, incompatible 'in their core', in primary substantive essence. In the same manner, social valuations are strongly incommensurable only between their core principal assertions, but weakly incommensurable on the periphery, in non-principal matters, where they overlap, this is where a major part of social interactions takes place and overlap.

The distinction between strong and weak incommensurability essentially improves the possibilities for aggregation of assessed impacts in complex evaluation. An evaluator can resolve the methodological inconsistency in Ekins and Medhurst by regrouping all the intervention measures in Tables II.1 and II.2 (rows) in the same way that impact areas organise – in columns – by three incommensurable domains. This divides the Leopold matrix as well as LEM into three sections vertically and only three sections horizontally, resulting in the construction of a square input-output matrix or meso matrix of impacts with nine sub-sections as separate fields of the matrix. A partial aggregate in each sub-section arises from summarising detailed impacts by a source (relevant rows) and area of impact (relevant columns; Table II.3).

A meso-matrical evaluation enhances two separate but indispensable views on how the Regional Program affects sustainability domains. The first view relates to direct impacts that are located on the diagonal of the matrix; it comprises assessment of the RP's effectiveness – how does it achieve its primary goals in each of its vertical domains or pillars of sustainability (previously referred to as the firstness of thirdness). The indirect or secondary impacts are located on the non-diagonal fields of the meso matrix. They describe trade-offs between the horizontally overlapping sustainability domains (the secondness of thirdness).

Table II.3 presents the impacts of three RP policy domains on three domains of sustainability in the Pomurje region. The sign '∩' denotes a horizontal overlap between domains. The sign is obtained from set theory, which is relevant since the Venn diagram, which represents a complex situation at the meso level, is one of its logical tools (see Figure I.2.5). The impact of economic policy on the social domain of evaluation is denoted as $E \cap S$ (*E* overlaps *S*), while the impact of social policy on the economic domain is $S \cap E$.

Note that in policy impact evaluation, $E \cap S$ is not identical to $S \cap E$. In set theory, an overlap consists of shared characteristics between the elements of two sets, equally belonging to both sets. Evaluation, however, works with incommensurable 'variables' so elements of overlap must be presented separately, with $E \cap S$ and $S \cap E$ as different elements of the same overlap in the Venn diagram.

Table II.3: RP's impacts – presented at the meso level (meso matrix)

Evaluation 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 of data: Table II.2.

Legends:

- Policy domains: * Economic. Aggregated from rows 1 and 2 from Table II.2; ** Social. Rows 3 and 4 from Table II.2; *** Environmental. Rows 5 and 6 from Table II.2.
- Impact scores: The same as in Table II.2.

The diagonal elements in Table II.3 are incommensurable so they cannot aggregate any further. They are ready for a descriptive explanation as they are. The evaluation suggests that implementation of the RP would be very successful in promoting economic development criteria (maximum, three pluses, instead of two in Table II.2), moderately successful in enhancing social development criteria (two pluses, instead of one in Table II.2) and only weakly effective in achieving primary environmental criteria (one plus in both Tables).

Table II.3 suggests that the RP handles the three sustainability domains in an unbalanced way. This observation does not agree with the one previously obtained from LEM. Note, however, that the summary row aggregates in Table II.2 and the diagonal elements in Table II.3 do not represent the same content. The former is a compound indicator of the total impacts (direct and indirect) of all intervention measures on a given evaluation domain, in this way ignoring incommensurability of some impacts (LEM's sin), while the latter more methodologically consistently expresses only the aggregate of direct, intra-domain impacts.

The secondary impacts (in overlaps between *E*∩*N*, *N*∩*E*, *S*∩E etc.), are located below and above the diagonal in Table II.3. They disclose the RP's cross-cutting impacts between domains of sustainability. This content is absent from the assessment results presented in Tables II.1 and II.2. Each cross-sectional relation in the meso matrix, such as *E*∩*N*, has its inversely symmetric counterpart, *N*∩*E*, and when they are correlated they describe the key characteristics of a relationship between two domains (*E*∩*N* and *N*∩*E*; simplified as *NE*, between environmental (natural) and economic domain). The meso matrix produces another two two-sided bilateral overlaps in Table II.4: *SE*, for the 'socio-economic' overlap and *NS*, for the overlap between (natural) environmental and social sustainability.

S is the only domain in Table II.3 that induces high positive indirect impacts – very positively on *E* and with moderately positive side-effects on *N*. *S* is most supportive for other domains of sustainability, but least supported by their favourable side-effects. Just the opposite is the case for *E*, it gives lesser support to other domains despite receiving their compelling support. The side impacts of *N* are, in this regard, the most balanced: *N* does not eagerly support the other sustainability domains, nor is it importantly supported by them, in this way producing separateness of environmental development from other regional concerns in the Program. The character of the relations between sustainability domains in terms of their indirect impacts indicates number of systematic imbalances in the RP. In particular, the RP's poor indirect impact on *S* will not contribute sufficiently to a reversal of unfavourable trends in regional social capital, relative to *E* and *N*.

Table II.4 presents evaluation of synergies between all sustainability domains.

Table II.4: The RP's correlation matrix of synergies between sustainability domains

	E	S	N
E	EE, Economic Sustainability (+++): Very effective	SE, Socio-Economic Sustainability (+++ , -): Strong overlap but involving negative trade-offs in favour of E	NE, Environmental-Economic Sustainability (+, +): Small overlap, balanced between N and E

S	-	SS, Social Sustainability (++): Moderately effective	NS, Environmental-Social Sustainability (++, +): Moderate overlap, unbalanced in favour of N
N	-	-	NN, Environmental Sustainability (+): Poor effectiveness

Source of data and Legend for Impact scores: The same as in Table II.3.

The two-sided bilateral overlap in *SE* (*S*∩*E* and *E*∩*S*) is strong but damaging for the social domain. The moderately strong two-sided bilateral overlap in *NS* (*N*∩*S* and *S*∩*N*) could be enhanced only if environmental measures better recognised the needs of social capital such as the strengthening local communities. Here we recall the previous finding (Table II.3), that the RP's intervention measures aimed at environment protection and nature conservation will not be very effective even in pursuing their primary goals. Thus, at least in relative terms, the RP will indirectly impose an additional lagging behind for social capital for very meagre environmental improvements. Finally, *NE* (*E*∩*N* and *N*∩*E*) is balanced, but only faintly positive.

The overall conclusion from a meso level evaluation of impacts between the three domains of sustainability is that the RP inconsistently contributes to regional sustainability, so it does not secure a reversing of critical development trends in Pomurje. This is again a less positive evaluation of the RP compared to the conclusions obtained from Tables II.1 and II.2.

Diverging evaluation conclusions derived from Tables II.1, II.2, and II.4 are by no means the consequence of different detailed impact assessments at the micro level. They arise solely from the summative endeavour with dissimilar assumptions about commensurability relations between evaluated contents on micro, macro, and on meso levels of evaluation.

4.1 Horizontal Extension

By explaining the mesoscopic aggregation procedure for the three-part version of Ekins's model, an intermediate goal of the Case Study is achieved. This invites us to return to the original four-part model of sustainability in order to develop deeper insight into the horizontal aspect of mesoscopic synthesis. Figure II.1 compares three- and four-part models in the form of a three- and a four-part Venn diagrams with the meso level of evaluation decomposed into three or more sublevels.

A brief theoretical introduction may be helpful for grasping the hierarchical structure of meso levels. Dopfer, Potts and Foster (DFP; 2004) developed a theoretical basis for the hierarchical organisation of meso logic. They do not understand meso as a single level but at three sub-levels representing a three-phase process of a system's trajectory from old to new as Meso 1, 2 and 3 in accordance with three steps of evolutionary process. The initial triadic micro–meso–macro hierarchy reshapes, in concordance with DFP, into a meso–meso–meso hierarchy. Meso is no longer merely an intermediating phase in the transition between micro and macro, but arises as a core perspective for inquiring complex social matters.

DFP describe Meso 1 as the micro to meso process of spontaneous emergence of novelty (rule, product, technology, group...) by early adoption of novelty between initial followers due to their affinities or similarity in relation to a novelty. In the Case Study, Meso 1 comprises the partial aggregation of detailed and weakly commensurable impacts into cells as 'niches' of the meso matrix. This is the lowest, relational level of mesoscopic evaluation.

Meso 2 consists of a meso to macro correlative process (at 'regime' level) in which novelty is integrated into a broader context through mixing and blending. The Case Study applies Meso 2 by correlating weakly incommensurable impacts in three overlaps between evaluation domains of the meso matrix, *SN*, *NE*, and *ES* (in Tables II.4 above and II.5b, below). Meso 2 is the evaluative level, on which the overlaps between constructs of Meso 1 arise (DFP; 2004).

Figure II.1: Venn diagram of sustainable development

Figure II.1a: Venn diagram of meso levels of sustainable development

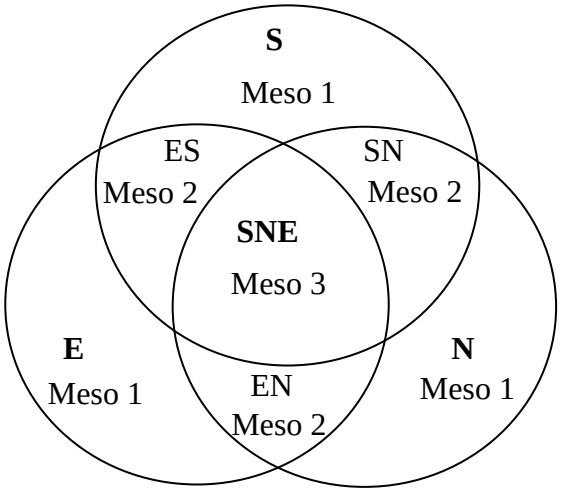
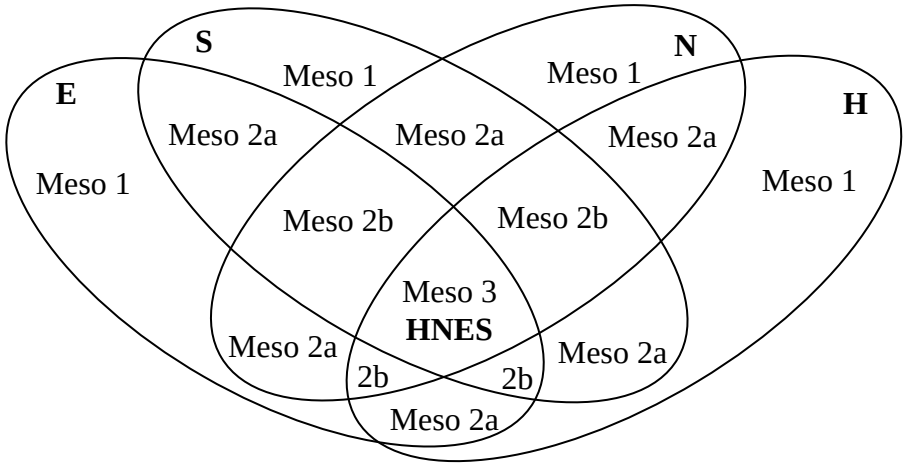


Figure II.1b: Venn diagram of the Four-capital model of sustainability



Source: Author.

The evolutionary novelty becomes a new order at a 'landscape' level when it reaches Meso 3 – the overlap between overlaps in the Venn diagram. With the concept of Meso 3, DFP account for the meta-correlation between the emergent products of Meso 2. The outcome of ordering on the Meso 2 level is typically circular at Meso 3 due to characteristic non-transitivity (see Arrow's theorem) of triangulated binary overlaps. In consequence, aggregation results at Meso 3 are not singular but multiple and partly inconsistent, so the overall message from evaluation of conflicts and synergies is not evident from obtained results at lower levels. Meta-overlap at the highest level of aggregation can be comprehended only interpretatively.

Instructed by DFP, one can extend the hierarchy of meso levels one step further. DFP produced a hierarchical theory of the meso level in the vertical direction of complexity only, while the horizontal differences are not included in their model (even though they acknowledge their relevance in side comments). As evolutionists of the Schumpeterian tradition, DFP were not sufficiently equipped to deal with the horizontal aspect of vertical emergence. This curtailed the prospects of developing their evolutionary model further towards social complexity.

Involving the horizontal axis into the aggregation algorithm is mandatory when synthesising incommensurable contents. By adding the fourth domain (H, for Human capital) of regional

sustainability to the evaluation model (Figure II.1b), horizontal axis of aggregation can be explained more in depth.

By adding *H* as the fourth evaluation domain, the synthesis procedure lengthens by one full cycle of horizontal correlations on Meso 2 sublevel. As a result, Meso 2 must be further decomposed into sub-levels Meso 2a and Meso 2b. The RP's cross-sectional impacts (Meso 2a) are now presented by six dual overlaps (*ES*, *EN*, *SN*, *SH*, *HE*, *HN*), where previously they had been presented by only three (*ES*, *EN*, *SN*; Table II.4). In addition, four triple overlaps are obtained (*HSE*, *EHN*, *SNE*, *NSH*; Meso 2b), where previously there was only one (*SNE*, as DFP's Meso 3 in Figure II.1a). In the Four-capital model, Meso 3 only shifts upwards from triple to a quadruple overlap (*HNES*, Figure II.1b). By adding the fourth horizontal domain, mesoscopic reasoning does not become increasingly relative as it would in standard approaches but extends 'into its own middle' (Prigogine, Stenger, 1979), into ever deeper and more heterogeneous overlaps.

The mesoscopic synthesis of four evaluation domains formally starts with the definition of the quadruple overlap (*HNES*, the Meso 3) between four domains (the Meso 1, Peirce's firstness in thirdness):

$$HNES = H \cap N \cap E \cap S.$$

HNES as the quadruple overlap can rewrite into the overlap between four triple overlaps of Meso 2b as it appears in the Venn diagram, Figure II.1b:

$$HNES = HSN \cap HSE \cap EHN \cap SNE. \quad (1)$$

Reformulation is practical because it translates a four-part correlation matrix into four sub-matrices of the third order (in line with Peirce). Triadic overlaps resolve the same as in Table II.4 (via non-diagonal fields of the correlation matrix) at the Meso 2a level, with a set of dual overlaps (secondness in thirdness):

$$HSN = H \cap S \cap N = (H \cap S) \cap (H \cap N) \cap (S \cap N); \text{ analogously for } HSE, EHN, SNE. \quad (2)$$

If further simplified by replacing $H \cap S = HS$, etc. we obtain presentation at the Meso 2 sublevel as schematised in Figure II.1a.

The formula of quadruple overlap (1) converts through (2):

$$HNES = (HS \cap HN \cap SN) \cap (HS \cap HE \cap SE) \cap (HN \cap HE \cap NE) \cap (SN \cap SE \cap NE). \quad (3)$$

Due to the idempotency of overlapping sets, it can be simplified, $HS \cap HS = HS$, etc., so reducing the equation (3) to the expression:

$$HNES = (HS \cap NE) \cap (HE \cap SN) \cap (SE \cap HN). \quad (4)$$

Therefore, Meso 3 materialises with (4) in three overlaps between dual overlaps evaluated at the Meso 2a sub-level:

$$(HS \cap NE) \rightarrow \text{View } (A),$$

$$(HE \cap SN) \rightarrow \text{View } (B),$$

$$(SE \cap HN) \rightarrow \text{View } (C).$$

Evaluating aggregate results from overlaps at the Meso 2a sub-level is most efficient since it reduces result to only three integrated components instead of four separate components if explained from the Meso 1 level (*H*, *N*, *E*, *S*) or from the Meso 2b sub-level (*HSN*, *HSE*, *EHN*, *SNE*). Three main perspectives of meta-overlap constitute three compound indicators, which present in cycle the three panoramic views of the RP's overall effects on regional sustainability.

Standard science does not like circular reasoning because it is irrational. However, it should not be incorrect to apply circular explanations to social issues, which are cyclical by nature. Circular methodology builds understanding by an 'alternating cycle of destruction and creation'.[4] Circular reading first deconstructs insights obtained from the previous view (*A*) as incomplete; this calls for a different aggregate reading (*B*), which successfully complements *A*, but is incomplete in some other respect that is involved only in (*C*), which again conceals a certain deficit that leads to *A*, and so forth...

The circular reading is valuable because it never asserts the pursuit of a final answer, so it never marks the end of the inquiry. It secures that the results of synthesis emerge only provisionally, and so can never assert themselves as 'structures of dominance'.[5] The circular reading perpetuates instead an ongoing questioning that keeps one's mind open to the idea that there may be more complete and even alternative understandings (Derrida).[6] Openness for radically diverse perspectives seems a more appropriate strategy for attaining neutrality in the evaluation of complex matters than Leopold's principal abstinence from comparison of incommensurable contents. The holistic understanding of poorly compatible contributions is also for Kuhn (1970) the precondition for a neutral comprehension of complex matters.

The mesoscopic procedure of horizontal synthesis is illustrated in the Case Study. Table II.5 presents the impacts of the RP from Table II.1 aggregated into four evaluation domains: first in a meso matrix of impacts (5a, weak commensurability) and then in a correlative matrix of overlaps between sustainability domains (5b, weak incommensurability, non-diagonally).

Table II.5: RP's impacts on sustainability of the Pomurje region – four-part presentation

Table II.5a: Four-part input-output matrix of RP's impacts

	E	H	S	N
E	E∩E (+++)	E∩H (+)	E∩S (-)	E∩N (+)
H	H∩E (0)	H∩H (+)	H∩S (+)	H∩N (0)
S	S∩E (+++)	S∩H (+)	S∩S (++)	S∩N (++)
N	N∩E (+)	N∩H (+)	N∩S (+)	N∩N (+)

Table II.5b: Four-part correlation matrix of the RP's impacts

	E	H	S	N
E	EE (+++): Very effective	EH (0, +): Very small overlap, imbalanced in favour of H	ES (+++ , -): Strong overlap, imbalanced in favour of E	EN (+, +): Small overlap, balanced
H	-	HH (+): Poor effectiveness	HS (+, +): Small overlap, balanced	HN (0, +): Very small overlap, balanced
S	-	-	SS (++): Moderately effective	SN (++, +): Moderate overlap, imbalanced in favour of N
N	-	-	-	NN (+): Poor effectiveness

Source of data: Table II.1.

Legend for Impact scores: The same as in Table II.3.

[4] Boyd J.R. 1976. Destruction and creation. https://en.wikisource.org/wiki/Page:Destruction_%26_Creation.pdf/7, II 2017.

[5] Ibid.

[6] Jacques Derrida: Deconstruction, by Catherine Turner, http://criticallegalthinking.com, III 2017.

The first view, (*A*), evaluates the RP's impacts between the material (*E, N*) and non-material (*H* and *S*) aspects of regional sustainability. The correlation matrix (Table II.5b) evaluates it as: *EN* – Small overlap; and *HS* – Small overlap. Therefore, the RP contributes to this aspect of regional sustainability in a balanced way even though only at the lowest level of positive synergies. It is not only that economic impacts poorly support non-material areas of sustainability, the opposite is also the case. Absence of material contribution by the RP is especially characteristic for regional H. The view (*A*) is best representing regional pillarisation strategy of sustainable development.

The view (*B*) evaluates the interaction between the progressive factors of sustainability (*H, E,* that are the most responsive to policy stimulus and most mobile), and conservation factors (*S, N*; that are sedentary and so they normally unfold 'conservatively', such as with preservation of cultural traditions and protection of nature). Achieved overlap in HE is very small, while overlap in SN is moderately strong. The meta overlap ($EH \cap SN$) is small and emphasises relatively strong connectedness only between conservation factors of regional sustainability. The view (*B*) is in line with the identified passive attitude of regional policy-makers to progressive initiatives in the field of regional development. The RP operates as an instrument of inertia, instead of instigating a progressive change.

The third view, (*C*), evaluates the nature of the interactions between the factors of regional sustainability that can largely be produced within the programming period (*E, S*) and non-produced, or more precisely, self-produced factors (*H* and *N*) beyond the programming period for which a Program is prepared. Evaluations shows that the Program induces a medium-strong overlap between *ES* and *HN*. The RP both positively and negatively affects the produced factors of sustainability significantly more than self-produced ones. This suggests, in our understanding, that the RP forwards a specific theory of change. It is more concerned with process-oriented (input-output) factors of regional development and therefore aspires for incremental, medium-term change rather than foundational and autonomous changes in self-produced capitals. The Program responds to problems that can be fixed relatively quickly, while disregarding issues that are more persistent so it is not likely to trigger reversal of negative trends and breakthroughs in the sustainability of regional development.

The triangulation of three views that compose Meso 3 level results confirms the previous observations indicating that the evaluated RP has an inappropriate intervention logic. It is in favour of materialistic and produced domains of sustainability but only conservatively, while also lacking transformative moment. The RP forwards a consistently restrained model of regional development that is deeply divided between its driving forces, it emphasises conservation values and allows only cautious progression without interfering with deep-seated structural problems of regional development. As a Program with the mission to improve regional sustainability, RP is too weakly synergetic between sustainability domains.

Comparison between results for three- and for four-set aggregation approaches shows how mesoscopic reasoning develops its argumentation when new qualitative information is added to existing blocks of knowledge – by extending into itself. That is not by becoming increasingly detailed, relativist or compromised (multi-criteria, middle-range) but by producing more heterogeneous perspective of the whole.

5 Evaluative Synthesis in the Middle

The Case Study supports claims that assumption of value incommensurability of social facts at the meso level has a superior aggregation potential compared to the standard assumption of elementary commensurability at the micro level. The main reason is that mesoscopic procedure, by adding a horizontal aspect to vertical aggregation, no longer acts as an enemy of qualitative

difference. Quantitative and qualitative aspects are both indispensable in the aggregation of complex matters. A vertical synthesis that ignores radical differentiation is trivial. Already in the fifth century CE, the Macedonian Joannes Stobaeus said that 'things that were alike and of the same kind had no need of harmony, but those that were unlike and not of the same kind and of unequal order'.[7]

The mesoscopic methodology of aggregation is an artefact of a forthcoming antipostmodern 'Age of Synthesis' (Belloc, in Wilhelmsen, Bret, 1970), which searches for less exclusive approaches to connecting increasingly diverse things. Ritzer and Smart (2003) outlined that the algorithm of synthesis is one of the central issues in the methodology of social research because it is about standards of consistent understanding of a large picture and integral aspects of social reality. Synthesis is more than simply putting things back together after one has taken them apart (Bartlett, 2001). Geels (2002, 2007) emphasised that aggregation is not simply a routine process of drawing general lessons from local projects; one must also take into account how power, ignorance, and framing play a role. The act of synthesis requires deep thinking and creativity in finding the most appropriate way to connect things, relative to the way things work together (Bartlett, 2001), which means in concordance with the internal logic of the challenge at hand.

Every epoch of social history forwards a different logic of synthesis, and they are an integral part of the ruling political arithmetic (Ritzer, Smart, 2003). The logic of synthesis always rests on culturally specific formulae of intellectual coherence (Prigogine, Stengers, 1979). 'The type of mathematics found in any major Culture is a clue, or key, to the distinctive character of the Culture taken as a whole'.[8]

For a mesoscopic methodology of social complexity, the deductive and divisive character of the standard approach poses a serious problem and raises suspicion about its appropriate rationale. The methodology of social research needs to develop new approach to synthesis that is less ambitious in terms of universality, but more connective and explanatorily rich since it decreases exclusion of radical difference in formation of social wholes.

Antipostmodern evaluative synthesis seeks the middle ground between commensurable and incommensurable approaches to constructing wholes. Evaluative synthesis must provide integral explanation of contents that are of course constitutive of the complex social issue at hand, but only in a deeply dividing way, as well as its secondary contents that enable synthesis but only in matters that are not of primary importance to anybody (Radej et al., 2012). New approach to synthesis must equally consider a large number of local synergies and a small number of deep and irresolvable oppositions. For Foucault, neither difference nor unity is primary. They need to be in balance (in Olssen, 2002), which can materialise only on the meso plane (Althusser)[9] of inter-paradigmatic standards (Kordig, 1973) with an evaluative efforts.

Non-transitivity ceases to present a logical problem to synthesis when irrational contents are manipulated evaluatively – by first locating a void in incommensurable claims and then by iterative, deconstructively constructive synthesis in the overlapping areas of their peripheral inconsistencies. The mesoscopic algorithm rests equally on learning and forgetting, by iterating acts of deconstruction in the center and transformative creativity on the margin.

---

[7] Wikipedia, #Stobaeus, I 2011.

[8] Wilder R.L. The Cultural Basis Of Mathematics. http://www-history.mcs.st-andrews.ac.uk/Extras/Cultural_Basis_I.html, II 2011.

[9] In Levačić M. 2009. Materialna eksistenca ideologije. Air Beletrina, http://www.airbeletrina.si/index.php?option=com_content&task=view&id=1713&Itemid=82, II 2011.

The Case Study suggests that Arrowian micro to macro impossibility does not arise entirely from the complex nature of collective choice. His impossibility results also from the oversimplified micro-macro theory of synthesis, which is incapable of taking advantage of the aggregative potential of irrationality. Irrationality is inseparable from collective choice so it is not wise to ignore it or treat it as an obstacle to aggregation.

As noted in introduction, aggregation is a rather primitive approach to synthesis. It can accordingly offer only a small contribution to integrative understanding of social complexes. Society is more than a pile; it is an entity with contradictory mechanisms of collective action. Another difficulty is prevalence of uniform drivers of social integration, which fail to empower a majority of those included,.

How to provide for social integration in complex conditions is the main concern of the next Case Study.