



UNICAMP

Número: 137/2005

**UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE GEOCIÊNCIAS
PÓS-GRADUAÇÃO EM POLÍTICA CIENTÍFICA E
TECNOLÓGICA**

MAURO ZACKIEWICZ

Trajetórias e Desafios da Avaliação em Ciência, Tecnologia e Inovação

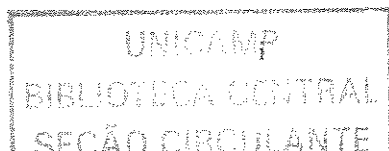
Tese apresentada ao Instituto de Geociências como
parte dos requisitos para obtenção do título de Doutor
em Política Científica e Tecnológica.

Orientador: Prof. Dr. Sergio Luiz Monteiro Salles Filho

CAMPINAS - SÃO PAULO

Fevereiro – 2005

Este exemplar corresponde a
redação final da tese defendida
por Mauro Zackiewicz
e aprovada pela Comissão Julgadora
em 28/02/2005



ORIENTADOR

**Catálogo na Publicação elaborada pela Biblioteca
do Instituto de Geociências/UNICAMP**

✱ Zackiewicz, Mauro
Z124 Trajetórias e Desafios da Avaliação em Ciência, Tecnologia e Inovação /
Mauro Zackiewicz – Campinas, SP: [s.n.], 2005.

Orientador: Sergio Luiz Monteiro Salles Filho
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Geociências.

1. Ciência e tecnologia. 2. Pesquisa e desenvolvimento – Avaliação. 3. Inovação
tecnológica. I. Salles Filho, Sergio. II. Universidade Estadual de Campinas,
Instituto de Geociências. III. Título.

UNIDADE	BC
Nº CHAMADA	T/UNICAMP
	Z124
V	EX
TOMBO BC/	63462
PROC.	16-P-20086-05
C	☐
D	☒
PREÇO	11,00
DATA	05/05/05
Nº CDD	

BIB ID- 349511



UNICAMP

**UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE GEOCIÊNCIAS
PÓS-GRADUAÇÃO EM
POLÍTICA CIENTÍFICA E TECNOLÓGICA**

AUTOR: MAURO ZACKIEWICZ

Trajetórias e Desafios da Avaliação em Ciência, Tecnologia e Inovação

ORIENTADOR: Prof. Dr. Sérgio Luiz Monteiro Salles Filho

Aprovada em 28/02/2005

EXAMINADORES:

Prof. Dr. Sérgio Luiz Monteiro Salles Filho

Prof. Dr. Rui Henrique P. Leite de Albuquerque

Prof. Dr. José Maria Ferreira Jardim da Silveira

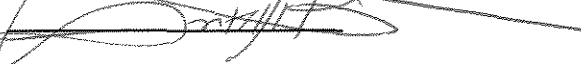
Prof. Dr. Abraham Sin Oih Yu

Prof. Dr. Antonio Márcio Buainain


_____- **Presidente**





Campinas, 28 de fevereiro de 2005

Agradecimentos

Sobre Moléculas e Música

A composer is a guy who goes around forcing his will on unsuspecting air molecules,
often with the assistance of unsuspecting musicians
Frank Zappa

Moléculas? A matéria é formada por moléculas. As idéias também surgem como moléculas. Daí bastam as condições favoráveis, talvez um ou mais catalisadores, interações imprevisíveis e pronto. Tem-se matéria prima para escrever uma tese. Tem-se a sequência de reações que permitem que se avance para substâncias que não se pensava serem possíveis.

A sorte de estar no lugar certo na hora certa. Que outro laboratório forneceria as condições acadêmicas que o DPCT tem sobrando? O ambiente multidisciplinar, as pessoas abertas a discussões profundas, a sensação de prazer em estudar e, não apenas em aprender, mas de apreender um objeto que não é simples. A crença na importância do conhecimento para o desenvolvimento da sociedade, da nação, da civilização.

E dentro do laboratório, o balão de ensaios chamado GEOPI. Um lugar pequeno e quente onde se concentram os reagentes. Ali não há como uma molécula passar impune, ela vai necessariamente se tornar outra, vai fazer parte de um complexo maior, vai liberar sua energia potencial. Entre a base teórica adquirida e a crítica ácida cultivada com humor, saís de múltiplas facetas vão se formando. E da interação com outros saís, catalisadores e toda a motivação que vem do inusitado, as reações não param de se tornar idéias.

Esse foi o ambiente. Um balão de ensaios que esquentou junto muita coisa. Mais de uma dezena de projetos de pesquisa realizados no tema avaliação. Alguns olhando o passado, outros o presente e vários deles o futuro. Muitas pessoas competentes trabalhando em grupo. E a imprudência de se fascinar com problemas difíceis.

Mas vamos aos agradecimentos. Primeiro, ao catalisador-mor, Sergio Salles Filho, o pedaço de metal sem o qual nada disso seria possível. Em segundo lugar, Adriana Bin, muito mais maleável que o orientador desse trabalho, por ser uma terra rara e de alta afinidade com Mz_2 , além de caracterizada por baixa energia de ativação quando em contato

direto, gera reações exo-idéias fantásticas. Ainda na turma dos companheiros de balão de ensaios, é inevitável mencionar os reagentes Bia, Rui Albuquerque, Ana Maria, Claudenício, Débora, Sonia Paulino, Solange, Sergio CP Paulino e Sonia Til, que estiveram todos comigo esses anos concentrados na mesma solução. Há também a turma dos sobrenadantes, Ana Lucia Assad, Celso Lage, Paula, Simony, Fernando Colugnati, Benami, Tacita, David e Felipe.

Além dessas pessoas queridas, agradeço também a todos os professores do Departamento de Política Científica e Tecnológica pela excelente acolhida durante minha estadia. Em especial Leda Gitahy, primeira responsável pelo meu interesse de entender melhor do que se tratava a C&T e fonte de muitas indicações de leitura. Menciono aqui também o professor André Furtado, interlocutor incansável e fundamental para a conformação das idéias desenvolvidas neste trabalho. E claro, a CAPES, que tornou viável financeiramente este estudo.

Para mim, foi como se este trabalho tivesse se iniciado de uma amostra retirada do balão de ensaios geopiano e circulado por diferentes bancadas do laboratório, até que se completasse o tempo da reação. Nasceu como uma matéria ainda bruta, que precisava ser trabalhada, para que se tornasse mais inteligível e agradável ao público. Eu tinha os instrumentos, tinha o tom, sabia já os conceitos que deveria transmitir. Faltava escrever a partitura, a música que iria comunicá-los.

Escrever uma tese é antes de tudo um trabalho de composição. Inicia-se com a página em branco, sobre a qual depois se delineiam melodias marcantes, que não podem ser mais evitadas, elas se tornam conceitos e com eles trabalhamos formando movimentos, seções de *tutti* e de *solo*, que, se forem bem sucedidas, chamam a atenção do público e valorizam o estilo do autor.

Mas, sinto se decepciono o leitor com essa afirmação, esse estilo nada tem a ver com inspiração. Ele foi deliberadamente copiado. Foi forjado, foi minuciosamente arquitetado. E nesse árduo percurso, pessoas fundamentais estiveram presentes. Desde sempre, meu pai, incentivador precoce da minha curiosidade e do meu senso crítico, minha mãe, por me ensinar o sentido da força de vontade e minha irmã, por me dar umas boas pancadas quando eu começava a achar que estava por cima. Da escola primária, professor Maurício, por me ensinar a não ter medo de textos difíceis. Do colegial, professores

Amadeu e Albanese, por escancararem a beleza e a flexibilidade da matemática. Do curso de engenharia na Unicamp, professores Ramon e Celso por me mostrarem o mundo "além" da técnica. Essas pessoas simbolizam as bases da minha trajetória intelectual, a todas elas sou muito grato e peço compreensão pelo fato de tê-las deslavadamente copiado.

Tendo já essa bagagem preciosa a tiracolo, nesses anos que estudei no DPCT, convivi e interagi com muitas outras pessoas sonoras. E em meio às inúmeras variações sobre o tema que marcaram esses anos, entre concertos, *jam sessions* ou mesmo violão de barzinho, a música não seria a mesma sem as contribuições da turma dos seminários de Lógica Fuzzy na USP: Pedro Tonelli, Laécio, Neli, Julio Cesar Pereira e outros, de Luís Flávio Autran Gomes (IBMEC/RJ); de Geraldo Stachetti (Embrapa), de Hélio Barros (Programa Prospectar - MCT), de Marcio Miranda Santos, Dalci dos Santos e Gilda Massari (CGEE), do Isaias Macedo e do Gilberto Jannuzzi (NIPE–Unicamp), da Angela Kageyama, do José Maria da Silveira e do Antonio Marcio "Tuca" Buainain (IE/Unicamp) e de mais muita gente com quem tive o prazer de trabalhar nos últimos tempos. Espero que todos vocês se reconheçam neste trabalho.

Agradeço finalmente meus amigos do peito Rafael e Rubens, fundamentais para que eu pudesse terminar esta peça, e ao maestro de *bits e bytes* André "Capivara" Drummond, responsável pela materialização do programa para o método de avaliação aqui apresentado.

Espero que esta tese seja útil e lida. Imagino que quem trabalha com avaliação vai, no mínimo, se divertir ao descobrir que mais alguém compartilha com suas aflições. E se você não trabalha, não sabe o que está perdendo. É um campo que reúne prática e teoria, ciências sociais e exatas e que, a cada novo trabalho, um desafio nunca antes enfrentado aparece. É um campo no qual é preciso saber onde encontrar os músicos certos e como soprar uma melodia que faça sentido, que agrade a platéia, mas que, principalmente, faça parte de uma composição que convença a nós próprios.

Bem, sugiro então escolher um bom disco e desejo uma ótima leitura!

Sumário

Introdução.....	1
Parte I – Elementos e Desafios Metodológicos da Avaliação.....	11
Distinção e Vínculo Entre Avaliar e Decidir.....	13
I. Avaliações pressupõem objetos e sujeitos.....	16
II. Atributos e critérios.....	21
III. Decisão.....	25
IV. Processos decisórios ou de avaliação?	28
V. A noção de causalidade.....	34
VI. Avaliação como legitimação social.....	38
VII. O problema da escolha social.....	41
VIII. Avaliação e instituições.....	47
IX. Atributos para caracterizar avaliações.....	52
Avaliação e Seus Métodos.....	55
X. Avaliação objetivada.....	58
XI. Anomalias da avaliação objetivada.....	67
XII. Métodos objetivados de avaliação – aplicação do esquema analítico.....	75
XIII. A estrutura das preferências.....	80
XIV. Métodos multicritério de apoio à decisão	86
XV. Sobre o significado de cientificidade em avaliações.....	95
Parte II – Avaliação na C&T em Rede e na Coordenação da Inovação.....	101
Avaliação da Ciência e da Tecnologia.....	103
I. Três diferentes formas institucionais das atividades de C&T.....	104
II. Avaliação na Pequena C&T.....	115
III. Avaliação na Grande C&T.....	122
IV. Macro tendências na organização da produção de conhecimento.....	131
V. Avaliação na C&T em Rede.....	138
O Método de Avaliação em Múltiplas Dimensões (MDM).....	145
VI. Antecedentes e princípios gerais.....	146
VII. Aspectos formais do método de avaliação em múltiplas dimensões.....	153
VIII. As variáveis associadas a uma estrutura de avaliação.....	158
IX. A obtenção das medidas.....	165
X. A escolha dos valores representativos.....	168
XI. Indicadores de qualidade da estrutura de avaliação.....	172
XII. Aspectos práticos da avaliação de impactos com o método MDM.....	174
XIII. Diretrizes para a coleta de dados.....	183
XIV. Comentários finais sobre a avaliação em múltiplas dimensões	186
Conclusões.....	191
Referências Bibliográficas.....	201
Anexo 1 – Detalhamento do Método Multicritério Electre III.....	217
As definições necessárias	217
Algoritmo da relação de superação (S).....	221

Índice de Figuras e Quadros

Figura 2.1: Esquema geral de hierarquização de objetivos para decisão multicritério.....	88
Figura 4.1: Na realidade completa os impactos são complexos.....	151
Figura 4.2: As dimensões de impactos analisadas são independentes e delimitadas.....	152
Figura 4.3: A síntese das medidas de impactos as tornam reticulados homogêneos....	152
Figura 4.4: Esquema geral de uma estrutura para avaliar impactos.....	156
Figura 4.5: Cadeia causal da geração de impactos.....	161
Figura 4.6: A múltipla causalidade dos impactos.....	161
Figura 4.7: Representação gráfica da Equação 4.3.....	163
Figura 4.8: Modos Direto e Indireto para obtenção de medidas de impacto.....	166
Figura 4.9: Parte da estrutura de avaliação para impactos econômicos de tecnologias na agricultura.....	179
Figura 4.10: Interface para edição de questões para os componentes básicos da estrutura de avaliação do método MDM.....	185
Figura 4.11: Interface para definição dos atores participantes e suas categorias.....	186
Figura 4.12: Resultados de campo para a avaliação dos impactos econômicos do PROCANA.....	187
Figura 4.12: Resultados de campo para a avaliação dos impactos econômicos do PROCANA, desagregado para os produtores da região de Ribeirão Preto.....	188
Quadro 1.1 – Portadores de instituições nos diferentes pilares institucionais	50
Quadro 2. 1 – Características que opõem os Paradigmas Qualitativo e Quantitativo de avaliação (Bogdan e Taylor, 1975).....	72
Quadro 4.1 -- Exemplo de indeterminação aparente entre diferentes escalas de medida e impacto.....	167



UNICAMP

**UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE GEOCIÊNCIAS
PÓS-GRADUAÇÃO EM POLÍTICA CIENTÍFICA E
TECNOLÓGICA**

Trajetórias e Desafios da Avaliação em Ciência, Tecnologia e Inovação

RESUMO

Tese de Doutorado

Mauro Zackiewicz

A discussão sobre desafios e trajetórias da avaliação em ciência, tecnologia e inovação (C,T&I) está diretamente ligada aos desafios de compreender como participação e aprendizado se relacionam com avaliação. A partir de um esquema analítico geral, são discutidos os fundamentos, a variação de formas possíveis e os principais desafios colocados aos processos de avaliação. Alguns problemas importantes em avaliações merecem destaque, como os limites lógicos à escolha social, a delimitação das relações de causalidade e atribuição, a legitimação via objetivação das avaliações e as inevitáveis influências subjetivas. A Parte I deste trabalho aborda o tema avaliação de modo relativamente abstrato, com ênfase em aspectos conceituais e lógicos dos processos individuais e sociais que geram as avaliações e seus métodos. Na Parte II, a discussão se volta aos diferentes contextos para os quais métodos de avaliação da ciência e da tecnologia foram desenvolvidos. Os principais métodos correntemente empregados para avaliar a C&T são apresentados e um método especificamente desenhado para a avaliação no contexto da C&T em Rede é descrito detalhadamente. Essa forma institucional da ciência e tecnologia é caracterizada pela co-existência de múltiplos arranjos entre múltiplos atores da sociedade, orientados à produção de conhecimento e inovação. Nela, características como participação, aprendizado e promoção de efeitos de coordenação visando inovação são distintivas. Essas premissas são ponto de partida para o Método de Avaliação em Múltiplas Dimensões (MDM) apresentado e discutido ao final do trabalho.



UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE GEOCIÊNCIAS
PÓS-GRADUAÇÃO EM POLÍTICA CIENTÍFICA E
TECNOLÓGICA

Evaluation of Science, Technology and Innovation: Trajectories and Challenges

ABSTRACT

Mauro Zackiewicz

The challenges and trajectories facing evaluation practices in science, technology and innovation are directly related to the comprehension of the links between evaluation, learning and actor's participation. A general analytical framework is proposed to discuss the evaluation practices fundamentals, their methodological variability and their main threats. Some important problems are highlighted, as the logical limits to social choice, the causal attribution between phenomena, the legitimacy of objective structures *vis a vis* the inexorable subjective influences. Part I presents in a formal way the logical and conceptual aspects of individual and collective methods and processes of evaluation. Part II presents three different institutional contexts of science and technology activities and the different methodological approaches that each one requires and legitimates. The main methods currently used to evaluate science and technology are discussed under this framework and a specific method designed to work in a network context is detailed. The institutional context of the Network Science and Technology is characterized by the coexistence of heterogeneous actors and organizations arrangements oriented to produce knowledge and innovation. At this context, elements like participation, learning and coordination effects are distinctive. These are premises of the Multi Dimensional Method of Evaluation (MDM) presented and discussed at the end of this work.

Introdução

Avaliação é um tema fundamental para a Política Científica e Tecnológica. O estudo do desenvolvimento científico e tecnológico, desde a discussão dos incentivos possíveis até a decisão sobre intervenções necessárias recorre sempre a algum tipo de avaliação. Seja de modo formal ou informal, avaliar é ponto de partida para se tomar decisões e, portanto, é etapa inescapável da formulação de qualquer política ou planejamento.

Avaliação é também um tema vasto e com ampla interpenetração em relação a muitos outros temas. Pode-se começar o exame das avaliações pelos diferentes métodos de avaliação, que são muitos. Pode-se também listar diferentes objetivos para avaliações, diferentes pressupostos teóricos e diferentes influências institucionais e culturais. Por cada um desses caminhos se revela parte do universo que constitui o tema avaliação. Porém é inevitável que essas rotas de investigação se cruzem e se sobreponham a todo o momento, tornando difícil apreender em poucas linhas todo o escopo de práticas, inter relações e significados associado ao tema avaliação. Não obstante, este trabalho é resultado de um esforço arquitetado para organizar a miríade de concepções, métodos, usos e motivações que envolvem as avaliações no campo da Política Científica e Tecnológica.

Todavia, sendo o campo coberto pela Política Científica e Tecnológica também vasto e também interpenetrado por diversas disciplinas, não é prudente tratar nele o tema avaliação sem relacioná-lo com a avaliação realizada a partir de referências mais gerais, dadas pelos procedimentos de avaliação de programas sociais e por outros métodos de avaliação empregados em diversas áreas das ciências sociais. Por outro lado, é importante que não se perca de vista as especificidades das atividades científicas e tecnológicas, elas próprias diferenciadas por adotarem ritos particulares de avaliação, que as distinguem de outras atividades sociais.

Mas há ainda mais uma motivação – talvez a mais importante – para o interesse do tema avaliação. O atual destaque dado por teóricos, governos e empresas aos processos de inovação, associado às dificuldades práticas de materializar essas mesmas inovações, paulatinamente traz à tona a necessidade de se repensar o significado e de se propor novos métodos de avaliação. Nesse contexto de promoção da inovação, são necessárias novas e

criativas soluções de planejamento e política para fortalecer a capacidade social de produção de conhecimentos e de criação de novos mercados baseados nesses conhecimentos. A essência do fortalecimento dos sistemas de inovação é o fortalecimento da capacidade social de produzir, circular e aplicar conhecimentos. Evidentemente, isso não é trivial e requer abordagens capazes de promover efeitos de sinergia e de coordenação coletiva que não se logram a partir de iniciativas isoladas ou de políticas fragmentadas.

Há, portanto, espaço para um novo tipo de planejamento e de política. Um planejamento que envolva mais que dívida, que compartilhe mais que segregue, que articule redes mais que defina hierarquias. Novas palavras chaves distinguem um novo padrão para o desenvolvimento científico e tecnológico, assim como também ocorre na organização das atividades produtivas e em outras esferas da sociedade. A avaliação nesse contexto também adquire novas palavras chaves: *participação* e *aprendizado*. Ambas refletem diretamente os novos objetivos do planejamento e da política de inovação. Sem participação ou sem aprendizado não há inovação.

Desse modo, a discussão sobre desafios e trajetórias da avaliação em ciência, tecnologia e inovação (C,T&I) está diretamente ligada aos desafios de compreender como participação e aprendizado se relacionam com avaliação. Como eles podem ser potencializados nas avaliações? Quais são as implicações metodológicas e conceituais? O que afinal significa participação? E de que tipo de aprendizado está se tratando quando se fala de inovação? Estas e outras questões relacionadas são os motores que motivam esta investigação e que ao mesmo tempo colocam importantes desafios para o desenvolvimento metodológico.

À primeira vista, participação e aprendizado parecem premissas salutaras para avaliações – e ninguém duvida disso. Entretanto, introduzir mecanismos de participação em avaliações não é algo simples nem necessariamente virtuoso. Também o aprendizado, seja individual seja institucional, não é necessariamente benéfico para as avaliações. De fato, ambos provocam problemas metodológicos profundos, que serão investigados em detalhe neste trabalho.

Essa linha de discussão vai revelar que importantes alterações conceituais e, especialmente, metodológicas são necessárias para que as avaliações cumpram o papel de promotoras de participação e aprendizado, sem se descaracterizarem, todavia, como

avaliações. Este trabalho defende a concepção de métodos de avaliação capazes de generalizar outros métodos e dotar o processo de avaliação das premissas de participação e aprendizado orientadas à inovação.

Em tempo, certamente é útil ao leitor saber um pouco mais sobre alguns antecedentes deste trabalho, e dos passos que levaram a seu formato final. A percepção de que participação e aprendizado eram pontos centrais para o debate sobre avaliação não surgiu senão após extensa vivência com a prática de execução de avaliações e com a experiência adquirida com a pesquisa e o desenvolvimento de metodologias de avaliação e prospecção tecnológica.

Assim, desde 1998 tive a oportunidade de me envolver com diversos casos de avaliação, sempre em equipe, a maioria com pesquisadores do Grupo de Estudos sobre Organização da Pesquisa e da Inovação (GEOPI). Entre eles se destacam um estudo de priorização de atividades no Centro Nacional de Pesquisa em Suínos e Aves (CNPISA) da EMBRAPA¹ e outro realizado para a Universidade Federal de São Paulo (Unifesp); a avaliação da parceria entre o WWF-Brasil e a missão da USAID no Brasil seguida pela avaliação do sistema de monitoramento das atividades dos parceiros da USAID no Brasil; o desenvolvimento metodológico para a prospecção tecnológica em saúde realizado para o CNPq; a pesquisa sobre o esforço tecnológico das Organizações Não Governamentais para compor o Diretório da Pesquisa Privada da FINEP; a avaliação do primeiro Edital do CT-Infra, também para a FINEP; os exercícios de prospecção tecnológica em energia e em recursos hídricos para o Centro de Gestão e Estudos Estratégicos (CGEE) e o desenvolvimento de uma metodologia para identificação e monitoramento de capacidade tecnológica e competências, primeiro aplicada para o campo da biotecnologia, por encomenda do Ministério da Ciência e Tecnologia, e, em seguida, para identificar o perfil tecnológico da Região Metropolitana de Campinas, visando a implantação do Parque Tecnológico em Campinas.²

Cada um desses casos forneceu elementos para a reflexão que este trabalho se propõe. Cada um, com suas próprias particularidades, chamou a atenção para certos aspectos fundamentais, tais como: a importância (e a dificuldade) da definição do objeto e

¹ A discussão crítica desse estudo foi objeto de minha pesquisa de mestrado (ZACKIEWICZ, 2000).

² Alguns desses relatórios podem ser encontrados em meio eletrônico em www.ige.unicamp.br/geopi, www.finep.gov.br e em www.cgее.org.br.

dos objetivos da avaliação; a importância do envolvimento das pessoas; a dificuldade de superar a aversão por avaliações presente em muitas pessoas ou organizações; e a dificuldade de se conseguir informação qualificada ou, por outro lado, a dificuldade de se lidar com a enormidade de informações muitas vezes disponível. Há também clara dificuldade de se trabalhar com técnicas amostrais e com modelos de causa e efeito quando o objeto da avaliação envolve tecnologias e inovação.

Porém, uma constatação derivada dessas experiências foi talvez a mais importante. As avaliações são ocasiões privilegiadas para a promoção do aprendizado pessoal e organizacional, para motivar a mudança e a resolução de conflitos. Uma avaliação gera muitas vezes apreensão em quem é avaliado ou em quem pode ser afetado por seus resultados e essas pessoas ou organizações podem quase sempre comprometer os resultados de uma avaliação, ao perceberem que seus interesses são contrariados. Desse modo, o grande desafio é transformar essa ameaça em apoio, comprometer os atores envolvidos e gerar um clima de cooperação tal que o processo e os resultados de uma avaliação façam sentido no contexto e para o contexto em que ela acontece, para que haja pró atividade e para que mesmo resultados desagradáveis possam ser interpretados como oportunidades de mudança e melhoria coletiva e não como justificativa para atitudes burocráticas unilaterais. Nesse sentido, o espaço para a participação se amplia, mas também se torna delicado. Assim, o grande desafio se dá em torno da definição do escopo de uma avaliação, de seu objetivo (ou objetivos), de quem deve ser envolvido e como, enfim, a construção de um processo de avaliação e decisão que leve a resultados efetivos.

É evidente que para isso, além de experiência e intuição refinada, é preciso desenvolvimento metodológico, é preciso dissecar os componentes dos procedimentos que constituem as avaliações, desde suas acepções mais básicas. É preciso também entender e delimitar o uso dos métodos tradicionais de avaliação e examinar alternativas para fortalecer a presença de participação e aprendizado.

Uma interpretação mais operacional do significado de participação não o limita apenas à idéia de envolvimento de diversas pessoas, trata-se mais da negociação de diferentes perspectivas e entre valores e interesses muitas vezes conflitantes. Portanto, um processo de avaliação que pressuponha o convívio de diferentes métodos, desenhados originalmente para avaliar diferentes dimensões da realidade, é também um tipo de

processo participativo. A participação traduz o envolvimento não só de diferentes atores, mas principalmente de diferentes interpretações acerca da realidade. E aí está um problema metodológico instigante. Como compor um processo de avaliação capaz de articular essas diferentes interpretações? Haveria um método capaz disso?

Em essência, mas não em forma, essa foi a pergunta colocada por um projeto de pesquisa desenvolvido pelo GEOPI e ainda não mencionado. Esse projeto foi o antecedente direto e certamente o mais importante para as proposições deste trabalho. Trata-se do projeto "Políticas Públicas para a Inovação Tecnológica na Agricultura do Estado de São Paulo: Métodos para Avaliação de Impactos de Pesquisa" financiado inicialmente pelo primeiro Edital de Políticas Públicas da FAPESP (1999 a 2003) e posteriormente pelo Fundo Verde Amarelo do MCT/Finep (2002 até 2005).

O desafio da construção de uma metodologia para avaliar impactos da pesquisa na agricultura foi uma proposta do GEOPI que contou com o apoio e a participação de uma ampla rede de pesquisadores interessados no tema, incluindo como parceiros a EMBRAPA e o IAC e a colaboração de pesquisadores de instituições como o Instituto de Economia da Unicamp, o IPT, o IBMEC/RJ, o BETA/ULP/Estrasburgo, dentre outras.

A motivação inicial da pesquisa era superar as limitações impostas pelas metodologias tradicionalmente aplicadas para a avaliação de impactos da pesquisa, em geral fortemente voltadas aos aspectos econômicos da inovação e baseadas no instrumental conceitual importado dessa disciplina. Importantes efeitos provocados pela inovação tais como impactos sociais, impactos ambientais, a criação de competências e mesmo certos efeitos econômicos não considerados pela interpretação econômica convencional precisavam ser considerados e incorporados. Mas como?

O caminho traçado pela pesquisa para responder a essa pergunta aproximou o grupo dos métodos multicritério de apoio à decisão, como alternativa para estabelecer um diálogo plausível entre as diferentes dimensões de análise do problema (cada uma com suas próprias bases conceituais) e as diferentes visões possíveis em relação à magnitude e relevância dos impactos. Atualmente, a primeira versão "estável" do método desenvolvido para avaliação de impactos passa por seu segundo teste empírico, a avaliação do Programa de Saneamento Básico - PROSAB – da Finep. O primeiro foi realizado na primeira fase do Projeto e avaliou dois programas do IAC, o Programa de Melhoramento Genético de Cana-

de-Açúcar (PROCANA) e o Programa de Produção de Borbulhas e Mudanças Sadias de Citros (executado em parceria com o Fundecitros). Conjuntamente a esses testes empíricos foi desenvolvido o software IMPACTOS, desenhado exclusivamente como ferramenta de apoio à execução de avaliações de impactos.

Meu interesse pessoal na teoria e aplicação desses métodos fez com que, desde o início da pesquisa, eu participasse ativamente desse processo de construção, de tentativa e erro e de reflexão acerca de suas possibilidades metodológicas. O ambiente acadêmico do DPCT, favorável à multidisciplinaridade, funcionou como o meio de cultura ideal para o cultivo de novas idéias. Os diversos tópicos abordados nos quatro capítulos que compõem este trabalho foram em grande parte selecionados a partir de um processo dialético de amadurecimento de idéias, impulsionado pelas críticas e sugestões dos muitos pesquisadores envolvidos na pesquisa e por visitas a diferentes fontes teóricas e empíricas.

Desse modo, este trabalho é um resultado natural do programa coletivo de construção de uma metodologia de avaliação de impactos. Porém, muito do que nele se discute e se conclui é fruto de um caminho independente de reflexão, que não segue necessariamente a opinião ou as preferências de todos que participaram do desenvolvimento da metodologia em si.

Para mediar essa reflexão, e para construir uma fundamentação mais sólida para o método de avaliação proposto, foram exploradas temáticas não tradicionalmente tratadas pelo Departamento de Política Científica e Tecnológica. Alguns exemplos incluem: 1) os fundamentos de lógica e álgebra das relações, tópicos necessários para entender e criticar as relações entre estruturas matemáticas e as propriedades dos métodos de avaliação; 2) os resultados dos teóricos da Escolha Social, uma vez que a participação de diferentes atores é uma característica central; 3) os métodos multicritério e 4) elementos de estatística descritiva (uni e multidimensional), inferência e amostragem. Além disso, somaram-se a essa agenda outros temas, esses do cotidiano do DPCT: filosofia da ciência, história da organização da C&T, processos de decisão, instituições sociais, inovação, dentre outros.

O plano de redação desta pesquisa divide o texto em duas partes, com dois capítulos cada uma.

A Parte I – Elementos e Desafios Metodológicos da Avaliação – consiste de dois capítulos que abordam o tema avaliação de modo relativamente abstrato, com ênfase em

aspectos conceituais e lógicos dos processos individuais e sociais que geram as avaliações e seus métodos. Juntos, eles perfazem 15 seções destinadas a expor, a partir de um esquema analítico geral, os fundamentos, a variação de formas possíveis e os principais desafios da avaliação. Algumas dessas seções dão destaque a problemas importantes em avaliações, como os limites lógicos à escolha social, a delimitação das relações de causalidade e atribuição, a legitimação via objetivação das avaliações e as inevitáveis influências subjetivas.

O Capítulo 1 – Distinção e Vínculo Entre Avaliar e Decidir: proposição de um esquema analítico para avaliações – introduz três hipóteses como pontos de partida para o estudo das avaliações como fenômenos sociais. A primeira hipótese relaciona as definições de atributo e critério para compor a noção de avaliação. Atributos e critérios são construções presentes no sujeito que avalia, os atributos são representantes do objeto avaliado e os critérios são estruturas de classificação mediada por juízos de valor. Avaliar consiste em interpretar um atributo segundo um critério. A segunda hipótese introduz a noção de decisão como evento complementar a uma avaliação. Uma avaliação só pode ser compreendida socialmente se implicar em alguma ação. Esse processo de materialização social de uma avaliação seria a decisão. A terceira hipótese afirma que todo método de avaliação é uma construção social. Desse modo, é possível explicar a legitimidade de certas formas e mecanismos de avaliação e distinguir a avaliação que ocorre apenas no indivíduo daquela que se conduz coletivamente (ou orientada para aprovação coletiva) mediada por algum método.

Esse esquema conceitual é discutido em detalhe no Capítulo 1 e alguns tópicos complementares são mencionados para ao mesmo tempo ilustrar o debate e introduzir questões importantes que orbitam o tema avaliação. Os principais pontos incluem as teorias sobre processos decisórios em organizações, a possibilidade de construção de métodos de escolha social, a influência das instituições sociais, o aprendizado individual e coletivo associado a avaliações, a distinção entre avaliação individual e coletiva e o desafio da participação. Constantemente, chama-se a atenção para o equilíbrio delicado entre fatos ditos subjetivos e fatos ditos objetivos. Diversas perspectivas são apresentadas e discutidas, abrindo espaço para que, no capítulo seguinte, essa tensão seja mais bem explorada no âmbito dos métodos de avaliação.

O Capítulo 2 – Avaliação e Seus Métodos: limites das formas objetivadas e as bases para participação e aprendizado – articula e aprofunda diversos tópicos sobre a construção de métodos de avaliação. Em especial, analisa-se criticamente os métodos que empregam estratégias de objetivação para se legitimarem como abordagens válidas – às vezes inclusive com a pretensão de serem cientificamente válidas. Tais métodos são escrutinados em suas propriedades e dificuldades operacionais no trato de objetos sociais.

A crítica é construída em dois níveis complementares. Primeiro, quanto as anomalias de adequação dos métodos quando aplicados aos objetos sociais. Segundo, quanto a impossibilidade de superar essas anomalias por conta da adoção de premissas que pertencem à lógica convencional, que necessariamente redundam em modelos matemáticos ou estatísticos objetivados e, portanto, inadequados para certos processos sociais de avaliação.

Entendidos esses pontos, uma forma de flexibilizar as premissas lógicas é discutida, a partir do exemplo dos métodos de apoio à decisão. As principais consequências formais e práticas são examinadas do ponto de vista de como as propriedades de um método estabelecido sobre lógica nebulosa (*fuzzy*), que demanda explicitamente a intervenção subjetiva, se ajustam aos pressupostos de participação e aprendizado.

Na Parte II – Avaliação na C&T em Rede e na Coordenação da Inovação – a discussão se volta aos diferentes contextos para os quais métodos de avaliação da ciência e da tecnologia foram desenvolvidos. Os principais métodos correntemente empregados para avaliar a C&T são apresentados em linhas gerais e um método especificamente desenhado para a avaliação no contexto da C&T em Rede é descrito detalhadamente. A segunda parte é composta novamente por dois capítulos, com 14 seções.

O Capítulo 3 – Avaliação da Ciência e da Tecnologia: o contexto de legitimação dos principais métodos de avaliação das atividades de ciência e tecnologia – apresenta em cinco seções três formas institucionais distintas da organização da C&T e os métodos de avaliação associados a cada uma delas.

Em primeiro lugar, a aqui chamada Pequena C&T, que reúne características fundamentadas no *ethos* mais puro da ciência e da produção desinteressada de conhecimentos. Essa forma institucional legitimaria métodos como a avaliação pelos pares

e os indicadores bibliométricos e cientométricos acerca do desempenho em termos de publicações e contingente de pessoas formadas, pesquisadores e professores.

A segunda forma institucional, a Grande C&T, seria caracterizada por operar segundo a racionalidade administrativa e a lógica da efetividade econômica. Os métodos de estimativas de retornos econômicos, de avaliação de impactos e de auditoria administrativa seriam típicos desse contexto.

Por fim, a C&T em Rede seria um arquétipo ainda em formação. Seria caracterizada por arranjos cooperativos entre atores bastante heterogêneos, coordenados e orientados à inovação, que operam em ambientes de grande mobilidade e incerteza. Nesse contexto, surgem métodos para identificar e medir essas estruturas de cooperação e as formas de circulação de conhecimento e competências. Alguns métodos adotam também uma postura pró-ativa de criação de condições para viabilizar a inovação, como é o caso das abordagens de *foresight*.

O Capítulo 4 – O Método de Avaliação em Múltiplas Dimensões (MDM): concepção e propriedades de um método para avaliação na C&T em Rede – apresenta e discute o método desenvolvido inicialmente para a avaliação de impactos dos dois programas tecnológicos do IAC acima mencionados. Na realidade, a abordagem proposta pelo método pode ser empregada para outros casos e para outros objetivos. Trata-se de um método que valoriza a participação, tolera a ambigüidade entre distintas interpretações sobre o mesmo fato, e que opera com múltiplas dimensões de análise.

O capítulo introduz e discute os aspectos formais do método, em seguida suas estratégias de mensuração e de agregação de medidas e, por fim, os aspectos práticos de sua operação. A apresentação do método se restringe ao caso de avaliação de impactos e alguns exemplos da aplicação prática são mostrados a título de ilustração.

Finalmente, após o Capítulo 4, segue uma seção de conclusões na qual se procura amarrar os principais pontos discutidos durante os quatro capítulos em torno da proposta metodológica apresentada e de suas possibilidades.

A relevância para a Política Científica e Tecnológica em desenvolver métodos de avaliação está justamente no fato que eles podem se tornar ao mesmo tempo instrumentos de diagnóstico e contabilização e instrumentos de intervenção junto aos processos de inovação. E, se isso é especialmente verdadeiro no contexto de produção de conhecimentos

que atualmente se desenha na sociedade, a relevância do tema pode se estender para outros âmbitos que não apenas os tradicionalmente envolvidos com C&T. Desse modo, espera-se que este trabalho contribua para o trato consistente dos problemas que suscitam as avaliações e suas conseqüências e que seja lido não apenas como contribuição à Política Científica e Tecnológica, mas a toda e qualquer área na qual avaliar seja importante e ao mesmo tempo desafiador, na qual a presença de objetos fluidos e dependentes de seus contextos sociais, como é o caso do conhecimento e das inovações tecnológicas e organizacionais, desautorize o emprego de abordagens demasiado retedoras e simplificadoras da realidade.

Parte I – Elementos e Desafios Metodológicos da Avaliação

Distinção e Vínculo Entre Avaliar e Decidir

Capítulo 1 – Proposição de um esquema analítico para avaliações

"Explicar" significa para a ciência que se ocupa das ações sociais
captar a conexão existente entre uma ação, já compreendida como fato,
e seu sentido subjetivamente motivado

Max Weber

Nas ciências sociais, e em especial em publicações voltadas à economia, às políticas públicas ou à política e à gestão de ciência e tecnologia, é comum a referência a avaliações e métodos para avaliações. No entanto, ao vasculhar a enorme massa de casos e aplicações à disposição, é difícil encontrar definições convincentes que dêem conta de explicar de forma geral e rigorosa do que se trata, em última análise, a atividade de avaliação. Por um lado, uma grande diversidade de formas e métodos é empregada e muitos objetos diferentes parecem motivar avaliações, produzindo uma enorme gama de aplicações e possibilidades. Por outro, as definições, quando são explicitadas, refletem concepções geralmente particularizadas a técnicas ou a aplicações específicas, quase sempre restritas a determinados recortes disciplinares ou a certos objetos e contextos.

Pouco se discute sobre a natureza geral do processo de avaliar. Tampouco parece haver preocupação sistemática acerca das conexões existentes e necessárias entre os aspectos conceituais e metodológicos que compõem tal natureza (ou naturezas) e as diversas possibilidades de aplicações concretas de avaliações. Avaliações são ao mesmo tempo instrumentos de investigação e de ação social, mesclam a produção de conhecimento com a produção de justificativas para a tomada de decisões. Portanto, merecem contínuo estudo e reflexão.

A primeira dificuldade para se tomar avaliações de modo amplo como objeto de estudo é a profusão de particularizações existentes. Em grande medida, elas refletem a natureza das disciplinas e dos objetivos por detrás das escolhas metodológicas. A opção de partir do conjunto completo de formas e métodos para, indutivamente, determinar as características e elementos comuns a todas as avaliações esbarra justamente nisso. A

variação de possibilidades está além da capacidade de recuperar cada forma e cada método, ainda mais por se tratarem de construções sociais, vinculadas a seus contextos de aplicação e de legitimação. Não se pretende, por tal motivo, revisar e expor exaustivamente todas as variações das definições possíveis para avaliação, embora algumas delas sejam mencionadas no decorrer do trabalho.

Também não parece interessante o caminho de se buscar definições generalizadas a partir de comparações entre definições ou entre métodos selecionados. Isso porque há muitas sobreposições terminológicas e conceituais nas definições encontradas na literatura. Desde aquelas mais simples, envolvendo uso diferenciado dos mesmos termos, até as mais difíceis, envolvendo diferentes recortes de análise como, por exemplo, entre o individual e o coletivo; entre premissas e métodos (a participação é um exemplo de elemento atualmente valorizado em avaliações e que muitas vezes ocupa uma zona cinzenta entre premissa e método); entre causas e efeitos; entre efeitos de primeira ordem e de ordens superiores ou entre modelos e realidade. Há também definições mais ou menos abrangentes, desde formas que isolam as avaliações como objetos independentes de seu contexto até as que estendem seu escopo de modo a considerar como sendo avaliação atividades que dela decorrem ou a ela dão suporte, como a tomada de decisão ou a pesquisa empírica; resultados que dela emergem, como a descrição ou a previsão de fatos ou objetivos sociais a ela associados, como o planejamento, a dominação ou o controle. Tudo isso torna extremamente penoso e arriscado selecionar e examinar exemplos típicos para, da comparação entre eles, inferir alguma definição mais abrangente.

Dito isso, é oportuno deixar explícito que este trabalho não vê muito sentido em procurar por conceitos definitivos ou por métodos que sejam intrinsecamente os melhores. Para uma avaliação, ser melhor ou pior depende do objetivo, do contexto, da aceitação social e das conseqüências que dela resultam. A qualidade de uma avaliação não é facilmente determinável e tampouco pode ser medida por parâmetros objetivos. Ela está claramente vinculada à maior ou menor adequação às suas condições de contorno, ou seja, à extensão em que cumpre sua função social.³ Não há como julgar avaliações sem um referencial particular de adequação social. Aperfeiçoar conceitualmente ou

³ A impossibilidade de um sentido "objetivamente justo" ou "metafisicamente verdadeiro" distinguiriam, segundo Weber (2004, p. 6), as ciências sociais das "ciências dogmáticas", incluindo a lógica, a ética e a estética, caracterizadas por adotarem *a priori* referenciais "justos" e "válidos" para examinar seus objetos. Toda ciência social deveria, por princípio, relativizar essa premissa.

metodologicamente uma avaliação pressupõe ampliar sua adequação a um determinado contexto social.

Assim, a comparação entre as mais diversas formas de avaliação não pode ser feita sobre bases comuns. Elas se referem necessariamente a diferentes contextos e são adequadas a diferentes situações. A crítica a esta ou àquela forma deve vir daquela ou desta preferência ou interpretação sobre adequação social, mas não das formas em si.

Descartada, portanto, a estratégia indutiva para encontrar definições para avaliação, resta a via dedutiva. Para tanto, mais importante que definições é a construção de um marco analítico para estudar as avaliações, capaz de fornecer uma interpretação geral para avaliações e, ao mesmo tempo, ser aplicável às particularizações mais comuns. Este é o objetivo central deste capítulo, obter por via dedutiva, um esquema analítico útil para estudar as avaliações.

O desenrolar da discussão, entretanto, irá revelar um segundo objetivo associado ao principal: testar a base conceitual introduzida para que, nos próximos capítulos, ela permita aprofundar o entendimento dos limites atuais da maior parte dos métodos de avaliação e orientar a busca por outras opções metodológicas possíveis. Assim, espera-se que esse esquema analítico ajude a consolidar ao final do trabalho diretrizes conceituais e metodológicas que estejam, ao mesmo tempo, intelectualmente bem fundamentadas e afinadas com os desafios identificados como pertinentes do ponto de vista metodológico e do ponto de vista das demandas colocadas pela sociedade contemporânea.

As diferentes definições enunciadas no decorrer deste capítulo são introduzidas por meio de três hipóteses. Na realidade, as definições apresentadas não pretendem ser nada além do que pontos de partida para o esquema analítico, abstrações conceituais para compreender melhor o que são as avaliações e como podem ser realizadas. Concretamente, cada avaliação pode ser definida segundo suas particularidades de aplicação, com grande variação de possibilidades.⁴ Para um marco analítico, por sua vez, nos interessa postular pontos comuns que distingam as avaliações de outras atividades. Não é o caso de estabelecer uma definição completa, e sim de arrazoar seus elementos fundamentais.

⁴ Uma revisão atualizada que ilustra bastante bem a profusão de formas de avaliação e de possibilidades de classificação pode ser encontrada em WORTHEN, SANDERS e FITZPATRICK (2004).

O Capítulo 1 está organizado em nove seções. Dessas, cinco são dedicadas a apresentar a construção do esquema analítico proposto para avaliações e quatro foram incluídas para testar a coerência das definições sugeridas, bem como introduzir discussões que são relevantes para o estudo da avaliação em geral e particularmente importantes para os capítulos subsequentes.

I. AVALIAÇÕES PRESSUPÕEM OBJETOS E SUJEITOS

No senso comum e na raiz etimológica a palavra *avaliar* está relacionada à noção de valor. Avaliar, portanto, consistiria em atribuir valor aos objetos, ou mais precisamente, em submeter determinados objetos à apreciação de determinados valores para se chegar a alguma conclusão a respeito de tais objetos.

O termo *avaliação*, na linguagem cotidiana, carrega indistintamente os significados associados ao ato ou ao efeito de avaliar um objeto. Desse modo, chamamos de avaliação tanto o processo de avaliar quanto o resultado obtido desse processo. Porém, em nome da precisão, ao nos referirmos a uma avaliação precisamos saber se falamos do processo de observação e valoração empregado ou se falamos dos resultados alcançados. Por esse motivo, neste trabalho, sempre que nos referirmos a uma avaliação, nos restringimos exclusivamente à sua execução, ou seja, às formas e aos métodos empregados para delimitar objetos e valorá-los. Ao nos referirmos aos produtos de uma avaliação deixaremos essa condição explícita ao dizer *resultados de uma avaliação*. Mais adiante, essa distinção entre processo e resultado será claramente estabelecida.

Inicialmente, é preciso examinar com mais detalhe a definição de avaliação dada pelo senso comum apresentada dois parágrafos acima. Ela contém implicitamente um elemento essencial. Trata-se da interação entre objeto e sujeito que é intrínseca a toda avaliação. A ação de avaliar pressupõe que exista um objeto bem definido para ser avaliado e que exista um sujeito capaz de executar um juízo de valor acerca do objeto.

Para se descrever rigorosamente uma avaliação é preciso que se descreva também rigorosamente o objeto e o sujeito, ou os objetos e os sujeitos, que dela tomam parte. Entretanto, esta é uma tarefa espinhosa, que remonta a séculos de debates. Objeto e sujeito são conceitos difíceis de manejar porque não são independentes. Ou seja, a definição de um

objeto depende de um sujeito que o defina e a definição de um sujeito força que o sujeito trate o próprio sujeito (ou a outro) como objeto.

É importante esclarecer adequadamente essa interdependência. Com o auxílio de algumas idéias básicas emprestadas da filosofia⁵ podemos introduzir conceitos que serão úteis no decorrer de toda a argumentação em direção a um entendimento mais profundo sobre as avaliações.

Em primeiro lugar, vamos examinar as relações entre objetos e sujeitos numa avaliação. Para se definir um objeto, costuma-se substituí-lo por seus atributos. Aristóteles afirmava que um objeto não pode ser descrito como coisa em si. Em outras palavras, a essência de um objeto não estaria acessível, apenas o que os sentidos humanos podem captar é o que somos capazes de perceber dos objetos. Antes dele, Sócrates afirmava que a distinção entre essência e aparência era justamente o que distinguia conhecimento de opinião. Enquanto da primeira emanaria a verdade eterna e imutável, as conclusões tiradas a partir das aparências poderiam até ter valor prático, porém não teriam valor em termos de verdade. Entretanto, como sustentou Aristóteles, a argumentação sistemática, mesmo que baseada em atributos do mundo sensível poderia levar à verdade. Seu método consistia na aplicação sucessiva de estruturas lógicas, os silogismos,⁶ com as quais se podia deduzir formulações que preservam o conteúdo de verdade das premissas que as originavam. Desse modo, duas hipóteses corretas levariam a uma conclusão correta. Uma hipótese incorreta levaria a uma conclusão incorreta. A busca pelo conhecimento verdadeiro seria então possível, via descoberta das hipóteses corretas que, em última análise, seriam portadoras indiretas da essência dos objetos e do mundo.

Desde então, filósofos debatem sobre a possibilidade ou a impossibilidade de se conhecer a essência de algo, ou em palavras mais atuais, sobre a possibilidade de um conhecimento objetivo, puro e independente do sujeito. O princípio da incerteza na física quântica e a relatividade na física astronômica são exemplos de como esse debate é persistente. Nesses dois exemplos não é nem a opinião do sujeito que seria perniciosa à

⁵As obras utilizadas para revisão e referência acerca dos autores e conceitos da filosofia clássica citados nesta seção foram RUSSEL, (1960), RUSSEL (2002), ABRÃO (1999) e PEREIRA (1999).

⁶ Um típico exemplo de silogismo: "Todo homem é mortal. Sócrates é homem. Então Sócrates é mortal". Observe que o fato de um silogismo ser falso não significa que ele está errado: "Todos os homens são livres. Prisioneiros são homens. Portanto, prisioneiros são livres". A estrutura do silogismo está correta, o absurdo da conclusão vem da falsidade da primeira hipótese. Ou seja, o resultado incorreto não invalida o método; pelo contrário, indica como aperfeiçoar o conhecimento.

verdade, mas a própria existência do sujeito seria um fato condicionante dos fenômenos e da possibilidade de se afirmar algo sobre eles. Ambas teorias permanecem aceitas e mostram que há limites para a idealização da noção de objeto. Para a física atual, um objeto nada mais é do que uma convenção aceita intersubjetivamente.⁷

Em todo caso, hoje em dia é fato intersubjetivamente aceito que, ao falar sobre um objeto, falamos na realidade sobre seus atributos. Isto é, definimos um objeto a partir de abstrações que correspondem a certas propriedades desses mesmos objetos, mas que definitivamente não são os objetos. Os atributos dos objetos são os "objetos" que de fato podem ser apreendidos empiricamente e somente em relação a eles podem ser obtidas medidas. Assim, toda descrição de um objeto passa irremediavelmente pela escolha e construção subjetiva de atributos. O que conhecemos sobre os objetos somente faz sentido a partir do sistema cognitivo humano; por mais que com esse conhecimento seja aparentemente possível fazer previsões sobre o comportamento dos objetos em si, tal conhecimento não existiria sem homens capazes de interpretá-lo.

Os atributos são necessários para identificar e comparar objetos. Em um atributo – e apenas por meio de um atributo – pode-se perceber diferentes estados de um objeto, descrevê-los e, em certos casos, realizar medidas e fazer comparações quantitativas.

Escala são também estruturas conceituais abstratas, utilizadas para fazer medidas. Uma escala pode ser obtida a partir de conjuntos como o dos números naturais (ordem completa de elementos discretos), o dos números reais (ordem completa densa) ou o de certos adjetivos de um idioma (uma escala nominal pode representar, por exemplo, diferentes intensidades de um atributo abstrato como "beleza"). Toda escala é um conjunto completamente ordenado e uma escala só é aplicável a um atributo se a variação dos estados possíveis do mesmo se mostrar compatível com a noção de ordenação total. A escolha da escala leva em consideração a natureza do atributo medido e o grau de precisão desejado ou possível para a medida.

Por exemplo, quando afirmamos que "as preferências dos indivíduos são mais bem representadas por escalas ordinais que por escalas cardinais" trata-se de um resultado de uma avaliação que tomou como objeto as possíveis escalas para medir um atributo do

⁷ Termo sugerido por Popper em 1934. Para o autor, "a objetividade dos enunciados científicos reside na circunstância de eles poderem ser intersubjetivamente submetidos a teste" (POPPER, 1975, p. 46). Um objeto é reconhecido como tal quando seus atributos são racionalizados do mesmo modo por todos os sujeitos presentes.

comportamento dos indivíduos. O atributo das escalas (nosso objeto) que interessou no caso foi sua forma matemática. A escala para esse atributo podia ser o grau de adequação das formas matemáticas candidatas para representar preferências e a avaliação escolheu as escalas ordinais como melhor resposta. Segundo esse exemplo, o objeto da avaliação, as escalas, são objetos abstratos que, por sua vez, são usadas para representar um certo atributo (as preferências) de certos objetos concretos (os indivíduos) e são medidas por meio de uma escala nominal. O exemplo também ilustra que uma escala pressupõe ordem, inclui as possíveis medidas para o atributo e que pode ser discreta ou contínua.

Prosseguindo esta seção, resta ainda examinar os modos pelos quais o sujeito atribui valor aos objetos avaliados. Além da apreensão do objeto por meio de seus atributos, o sujeito interfere com seus valores para realizar juízos acerca do objeto a partir das descrições ou medidas que dispõe.

Essa valoração pode, todavia, ser realizada por meio de juízos diretos ou indiretos. Os juízos são chamados diretos quando o resultado da avaliação é dado por gosto do sujeito e são chamados indiretos quando uma valoração já está embutida na escala escolhida para medir o atributo. Em ambos os casos, emitir um juízo significa classificar os objetos, a partir de seus atributos, como certos ou errados, bons ou ruins, preferidos ou preteridos, ou em outros conjuntos de chegada satisfatórios para o problema de avaliação considerado. Essa estrutura de classificação é chamada *critério* em nosso esquema analítico. O juízo indireto acontece quando os atributos empregados na avaliação já possuem estrutura de critério e o juízo direto quando o sujeito projeta um critério pessoal sobre um atributo qualquer, mesmo que esse possua uma estrutura ordenada que poderia eventualmente funcionar como critério. Assim, um mesmo estado de um mesmo atributo pode provocar juízos diferentes em dois sujeitos distintos.

Da mesma forma que se imaginou a possibilidade de conhecer os objetos em si, muito se discutiu sobre a existência de valores humanos universais que pudessem ser critérios invariantes para avaliar e decidir. Diversas formulações éticas, ao longo da história, buscaram cumprir esse papel.

Sócrates identificava o bem com o conhecimento, de modo que buscar a verdade e a essência do mundo era o mesmo que praticar o bem. Assim, o bem não estava nas opiniões dos homens, e sim objetivado na ordem oculta do universo. O homem bom seria aquele que

desse mais valor à busca do entendimento dessa ordem do que à realização de seus desejos e opiniões. No século XX, a ética socrática, base comum da maior parte das éticas religiosas e da ética científica no ocidente, foi amplamente contestada tanto por filósofos humanistas e existencialistas como pelo individualismo e consumismo típicos da sociedade industrial. O contraponto de Nietzsche à concepção socrática, advogando pela centralidade dos desejos e vontades humanas na formulação de uma ética alternativa à dominante,⁸ abriu caminho para a percepção que preceitos éticos e morais também são construções sociais e, portanto, não são imutáveis e não se pode falar em juízos melhores ou piores, corretos ou incorretos *a priori*.

É de se esperar que as mesmas características percebidas de um objeto levem a avaliações distintas, dados sujeitos pertencentes a tradições éticas distintas. Sujeitos imbuídos de padrões éticos diferenciados terão juízos também diferenciados sobre determinados atributos de um objeto e suas variações. Sendo construções sociais, esses padrões são sujeitos a diferenças culturais, materiais, espaciais e temporais que caracterizam o desenvolvimento de um indivíduo. Hoje, no início do século XXI, essas diferenças são ainda mais significativas dado que muitos problemas passaram a ser avaliados simultaneamente por atores sociais também muito distintos, tornando esse aspecto de especial importância para quem se preocupa com processos de avaliação. Inclusive as diferenças entre línguas são importantes. Línguas diferentes contêm distintos conceitos sobre objetos e modos diferenciados de se referir aos eventos. Se a realidade fosse um contínuo, as línguas poderiam ser entendidas como recortes discretos dessa mesma realidade e seu funcionamento, e esses recortes não seriam totalmente coincidentes para todas línguas, e para cada uma delas, em todos os tempos e lugares.⁹

Com isso, espera-se cumprida a intenção de mostrar que em torno da noção geral de avaliação, conforme tida pelo senso comum, gravitam persistentemente algumas questões filosóficas de difícil e improvável resolução. Mais do que ambicionar resolvê-las, para a construção de um marco analítico para o estudo de avaliações é preciso tê-las em conta e tentar, com isso, escapar de concepções simplistas sobre conceitos e práticas de avaliação.

Finalmente, antes de iniciar a apresentação das hipóteses, cabe um último

⁸ Cf. NIETZSCHE (1999) em Para Além do Bem e do Mal, de 1886.

⁹ As línguas também são construções sociais e, como tais, estão em contínua transformação. Os significados atribuídos a determinados fatos no passado nem sempre são os mesmo no presente, mesmo para uma mesma língua (RUSSEL, 2002).

comentário sobre um persistente problema que assola aquele que se debruça sobre a avaliação como seu objeto de estudo. Quem realiza a avaliação é um sujeito. Mas para quem estuda a avaliação, o sujeito que realiza uma avaliação é também um objeto. Para o analista, sujeito e objeto de uma avaliação são objetos de sua própria avaliação. Há um risco em definir atributos para o objeto sujeito porque esses atributos deveriam valer também para o analista, mas como ele poderia saber se eles estariam corretos se não pode avaliar a si mesmo apenas como objeto? Mesmo supondo que pudesse, para isso necessitaria emitir juízo externamente a seus juízos "congelados" como objeto, o que é evidentemente uma contradição. Tomar o sujeito de uma avaliação como objeto é um recurso analítico que impossibilita uma conclusão a partir dos próprios elementos do problema, uma vez que provoca uma estrutura recursiva auto referenciada que sempre precisará de juízos externos para se realizar em avaliação. O risco de ficar preso entre os infinitos reflexos provocados pelas noções de sujeito e objeto, como se fossem dois espelhos frente a frente, impõe um limite claro a uma definição rigorosa e exaustiva para avaliação. É preciso evitar essa armadilha para que o esquema analítico tenha alguma utilidade.

II. ATRIBUTOS E CRITÉRIOS

O esquema analítico proposto neste capítulo não tem a pretensão de servir como teoria geral para avaliações. Entretanto, ele é construído a partir de hipóteses que procuram, a custo da particularização do sentido de alguns termos utilizados na linguagem corrente, propiciar a explicação de diferentes aspectos relativos ao processo de avaliação e suas conseqüências.

Uma dessas particularizações, já realizada, é sobre o próprio termo *avaliação* que na linguagem comum significa tanto o ato quanto o efeito de avaliar. As hipóteses do esquema analítico proposto a seguir consideram avaliação apenas como o processo de avaliar. Outro termo que será introduzido adiante em forma particularizada é *decisão*. Para efeito de interpretação das avaliações, os juízos ainda não são as decisões, eles são parte do processo de avaliação, do ato de avaliar. Só depois que esse termina é que as decisões acontecem. A separação entre avaliação e decisão pode parecer uma construção um tanto artificial num

primeiro momento, mas é especialmente útil para a interpretação de uma série de fatos que envolvem avaliação.

Uma terceira particularização é a representação de objetos e sujeitos por atributos e critérios, conforme foi introduzida na seção anterior. Assim, pode-se dizer que atributos e critérios são elementos que fazem parte de toda e qualquer avaliação. Ou de modo inverso, se uma avaliação não contiver atributos e critérios não será uma avaliação.

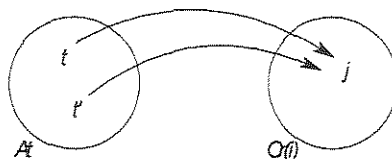
Desse modo, toda avaliação envolve partir de atributos e impor juízos sobre esses atributos, transformando-os em critérios. Uma avaliação pode, portanto, ser interpretada como uma função que leva atributos a critérios. Podemos enunciar, então, o primeiro princípio que define o esquema analítico.

Princípio 1: Avaliar é interpretar um atributo em uma estrutura de critério.

$$A: At \rightarrow Cr_i$$

Supõe-se que pelo menos uma caracterização do estado t de um atributo pertença ao conjunto At de todos seus estados possíveis (At não é vazio). O atributo avaliado é completamente descrito pelos estados possíveis contidos em At e At é um conjunto de estrutura qualquer, não pressupondo ordem (mas podendo eventualmente apresentar ordenação parcial ou total). O conjunto Cr_i contém uma estrutura, no mínimo parcialmente ordenada e limitada acima e abaixo,¹⁰ formada pelos juízos possíveis (j) de um indivíduo i .

O Princípio 1 é bastante geral, porém alguns fatos decorrem. A avaliação A é definida como uma função, portanto, para cada $t \in At$ é possível apenas um $j \in Cr_i$. Por outro lado, para diferentes $t \in At$, o mesmo $j \in Cr_i$ pode ser obtido. Esse tratamento rigoroso apenas formaliza o que já foi dito de modo discursivo na seção anterior.



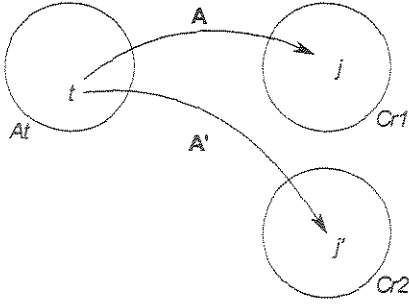
¹⁰ Essa estrutura corresponde à estrutura de um reticulado. As propriedades matemáticas dessa definição são devidamente exploradas no Capítulo 2.

O conjunto At pode ser tomado como representando apenas um atributo, ou quantos forem o caso. Na formulação mais geral, At pode ser o conjunto de todos os estados possíveis para todos os atributos de todos os objetos possíveis. Como nunca $At = \{\emptyset\}$, não é possível a avaliação prescindir de pelo menos uma caracterização de pelo menos um atributo.

A eventual escolha de escalas e suas opções para construir o conjunto At , como foi comentado na seção anterior, é resultado de um processo preliminar de avaliação (geralmente buscando intersubjetividade). Porém, para fins de definição, o conjunto de partida At para a função avaliação é considerado como dado. Evidentemente, esse corte é necessário para evitar que o esquema analítico tenha que abarcar todo o processo, de extensão variável e muitas vezes imponderável, de avaliações e decisões encadeadas que de fato compõem uma avaliação.

O conjunto de chegada Cr_1 pode ser modelado de diversas maneiras, porém apresenta mais restrições estruturais que At . Ao avaliar um atributo o indivíduo chega a um juízo pertencente a uma ordenação preexistente que ele utiliza como base para comparação e subsequente ação. Opções incluem {falso; verdadeiro}, $[0, 1]$, {feio, indiferente, bonito} ou os próprios estados dos atributos ordenados segundo os valores ou pressupostos éticos e conceituais do sujeito, desde que limitados acima (o juízo mais positivo) e abaixo (o juízo mais negativo).

Devido à estrutura de função definida para uma avaliação, é vedada a possibilidade de se obter mais de um juízo para um mesmo estado de um atributo. Desse modo, a ocorrência de juízos distintos para um mesmo t revela instâncias distintas de avaliação. Ou seja, $A: At \rightarrow Cr_1$ com $j \in Cr_1$ e $A': At \rightarrow Cr_2$ com $j' \in Cr_2$. O novo contra-domínio Cr_2 pode representar um outro indivíduo ou o mesmo indivíduo em outro momento, ou ainda um novo critério criado a partir do mesmo atributo. Por comodidade e sem perda de generalidade, vamos por hora identificar as diferentes instâncias de avaliação com diferentes indivíduos.



O conjunto de todas as instâncias possíveis $A, A', \dots$, é chamado A , *conjunto das avaliações*. O conjunto de todos os juízos j possíveis a todos os indivíduos (a união de todos os critérios) é chamado J , *conjunto dos juízos possíveis*.

$$\bigcup_{n=1}^r Cr_n = J$$

O conjunto comum a todos os juízos j das diferentes instâncias é chamado I , *conjunto dos juízos comuns*.

$$\bigcap_{n=1}^r Cr_n = I$$

A estrutura mínima requerida para um critério não é necessariamente preservada em J , mas o é em I .

Se em uma avaliação $A, j \in I$, a avaliação $\neg A: At \rightarrow I$ é chamada *avaliação coletiva* e é única. Mais adiante vamos explorar com mais detalhe esse ponto. Entretanto, vale destacar que a possibilidade de uma avaliação coletiva fica restrita à existência e ao uso de um critério comum por todos os indivíduos, seja por coincidência seja por ser invariante ou imposto a todos eles.

III. DECISÃO

A noção particularizada de decisão apresentada nesta seção é essencial para complementar o modelo analítico para avaliações. Embora não faça estruturalmente parte de uma avaliação, mostrada acima como uma função que leva atributos a critérios, a decisão pode ser identificada com a materialização de uma avaliação, delimitando-a no tempo por meio de uma ação. Dito de uma forma muito sucinta, uma decisão ocorre assim que se conclui uma avaliação, é a materialização do juízo em ação.

A decisão é, portanto, decorrente da avaliação. Desse modo, nesse esquema uma decisão só pode ser apreendida depois que ocorre uma ação (ou uma não ação, que não deixa de ser também um tipo de ação). Sem um sinal socialmente percebido, um observador não pode saber que houve uma avaliação. Mas depois do sinal, ou da suposição do sinal, a avaliação pode ser compreendida pela decisão.¹¹ Uma *ação* pode ser caracterizada desde pela manifestação de algum sinal involuntário até por estratégias sofisticadas de comunicação ou intervenção por parte do indivíduo. Somente um juízo socializado em forma de ação é considerado aqui como decisão.

Princípio 2: Uma decisão é uma ação produzida em decorrência de uma avaliação.

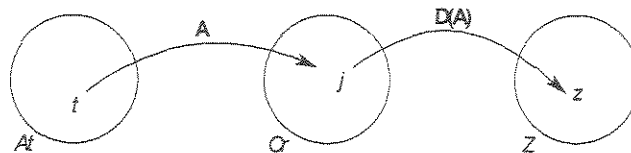
Esta definição vale para cada avaliação $A \in \mathcal{A}$. A partir do Princípio 1 podemos escrever

$$\neg D(A) : Cr_i \rightarrow Z_i$$

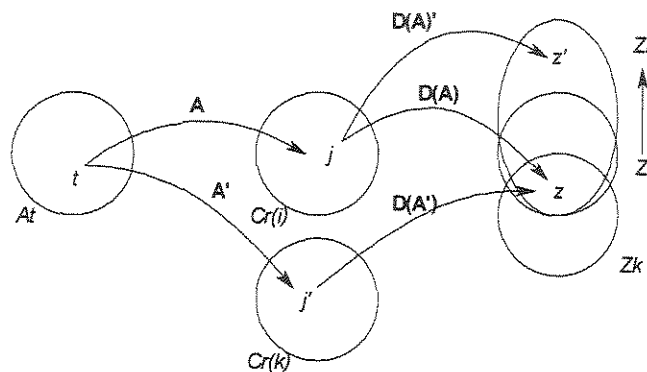
Segundo essa função, para cada $j \in Cr_i$, corresponde uma ação z no contra-domínio

¹¹ De fato, um observador externo tem acesso às ações concretas que pode perceber e aos atributos dos objetos que ele supõe terem sido considerados na avaliação. A inferência, nessa situação, de quais foram os critérios empregados pelo indivíduo para avaliar e agir se aproxima da noção de *compreensão* weberiana. Um indivíduo, por sua vez, pode objetivar sua avaliação revelando seus critérios ou empregando juízos indiretos. A explicação a que Weber se refere na epígrafe deste capítulo (WEBER, 2004, p. 6) envolve não só compreender as motivações associadas às estruturas de juízos dos indivíduos, mas também revelar o universo dos atributos percebidos e os modos como avaliações e decisões ocorrem para se tornarem ações sociais. Diversas relações e ordens sociais concorrem para definir padrões de conduta e reduzir as opções disponíveis. A forma abstrata do esquema analítico exposto neste capítulo não inclui propositalmente essas restrições para que se possa com ele simular situações especiais, de interesse para certos problemas metodológicos em avaliações.

Z_i , conjunto de todas as ações z possíveis ao indivíduo i .



Novamente, segundo as propriedades das funções, dado um juízo j , apenas uma ação z é possível para o indivíduo i . Juízos distintos podem levar a uma mesma ação, nada impede que um indivíduo k chegue à ação z por meio de um juízo j' . De todo modo, decisões diferentes indicam que vieram de avaliações diferentes ou que o indivíduo alterou seu leque de ações. Assim, $\mathbf{D}(\mathbf{A}')$ representa uma nova instância por conta de outra avaliação e $\mathbf{D}(\mathbf{A})'$ por conta da alteração do conjunto de chegada Z , relativo ao indivíduo i .



O conjunto de todas as ações z possíveis a todos os indivíduos é chamado Z , *conjunto das ações possíveis* e não pressupõe ordenação, por definição.

$$\bigcup_{i=1}^n \angle_i = \angle$$

O conjunto de todas as ações z comuns a todos os indivíduos é chamado S , *conjunto das ações comuns*. Também não pressupõe ordenação.

$$\bigcap_{n=1}^{\infty} Z_n = S$$

A partir dessas definições e daquelas enunciadas na seção anterior, pode-se descrever quatro situações especiais de decisão. Para qualquer conjunto com $n \geq 2$ indivíduos:

1) Se $j \in I$ e $z \in S$, a decisão $\neg D(A)_c: I \rightarrow S$ é chamada *decisão coletiva*. Para ocorrer uma decisão coletiva é preciso que exista um juízo e uma ação comuns a todos os indivíduos. A instância $D(A)_c$ é única e, nesse caso, S possui a estrutura ordenada de I .

Observe que se, por hipótese, $j \in I$ e $z \in Z$, a decisão será coletiva somente se existir $S \neq \{\emptyset\}$, o que reduz a situação ao caso anterior, z deve necessariamente pertencer a S . Ou seja, se a avaliação levar a um juízo comum ela é única e a decisão será também necessariamente única e coletiva. Com $S = \{\emptyset\}$ não pode haver decisão coletiva.

Sempre que $j \in I$, se existir uma ação z , ela será única e representará a interseção das ações possíveis para os indivíduos de algum subconjunto do universo de ações possíveis aos indivíduos. Seja W um subconjunto de Z , se $j \in I$, então a ação $w \in W$ é uma *decisão localmente coletiva* para os indivíduos representados em W .

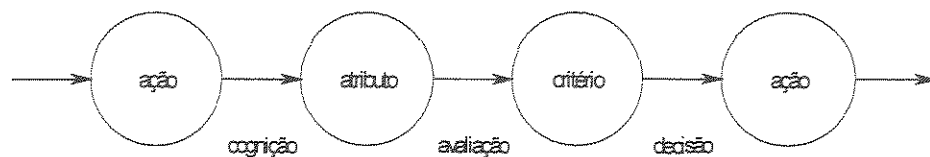
2) Se $j \in J$ e $z \in Z$, a decisão é chamada *decisão trivial*. Todas as instâncias para D podem valer. As ações resultantes podem não fazer sentido pleno a determinados indivíduos, nos casos em que não compartilhem do mesmo universo de juízos e/ou de ações possíveis. As condições $j \in J$ e $z \in Z$ são as mais frouxas possíveis e tornam a situação da avaliação/decisão totalmente indeterminada. Disso decorre que se não se puder limitar os juízos possíveis e as ações possíveis não há como prever avaliações e decisões. Portanto, avaliações e decisões para serem únicas precisam de condições mais restritas, como as mostradas no item anterior.

3) Se $j \in J$ e $z \in S$, a decisão é chamada *decisão limitada*. Nesse caso, faz-se restrição no conjunto de chegada das decisões. O resultado da decisão cai sempre em um subconjunto de ações comum a todos os indivíduos. Como os juízos possíveis não foram limitados, as decisões ainda podem ser muitas, mas pelo menos estão sempre dentro do

universo de possibilidades comuns (e cognoscíveis) a todos os indivíduos.

4) Se $\forall j \in J_i$ e $\forall z \in Z_i$, a decisão é dita *ditatorial*. As restrições colocadas indicam que os juízos possíveis e as ações possíveis se limitam às do indivíduo i . A decisão é sempre a ação z realizada pelo indivíduo i .

Finalmente, uma regra que decorre dos Princípios 1 e 2 é que toda avaliação é *ex post* a alguma mudança percebida pelo sujeito em algum objeto e *ex ante* a uma nova decisão. Essas mudanças no universo sensível dos sujeitos são resultantes de decisões anteriores.



Desse modo, todos os conceitos trabalhados até aqui: a) a delimitação de um objeto por meio de seus atributos (que só são perceptíveis porque mudam ou podem mudar); b) a avaliação e c) a decisão constituem um só processo concatenado no qual recorre percepções/medidas, juízos e ações para progressivamente tecerem a complexidade do universo de percepções, valores e movimentos que caracterizam os sujeitos em interação com o meio natural e com outros sujeitos.

IV. PROCESSOS DECISÓRIOS OU DE AVALIAÇÃO?

O Princípio 2 é, como foi dito, uma particularização do senso comum sobre decisão e, como foi mostrado, leva ao entendimento de decisão como um evento que decorre de uma avaliação provocando uma ação que delimita uma célula analítica que pode ser reproduzida inúmeras vezes durante a aplicação de um método de avaliação. Da mesma forma, o que muitos autores chamam de *processo de tomada de decisão* envolve quase sempre uma teia desses ciclos, que no universo empírico pode se mostrar bastante complicada. Uma consequência do Princípio 2 é que, segundo esse esquema analítico, para se compreender rigorosamente uma decisão se torna necessário considerar toda a trajetória

de avaliações e decisões concatenadas que ocorreram antes de tal decisão. O mesmo vale para uma ação socialmente percebida.

Para ilustrar melhor os conceitos introduzidos e marcar as distinções em relação a outros esquemas que fatalmente acabam usando as mesmas palavras com interesse analítico diverso, é apropriado examinar, por exemplo, algumas teorias sobre o chamado processo de tomada de decisão. A revisão de Ham e Hill (1984, cap. 5) sobre teorias que discutem decisões no processo político conta como alguns importantes autores interpretam como e porque elas ocorrem. Pelo esquema proposto por Ham e Hill, há duas correntes principais de interpretação para a tomada de decisão no âmbito das organizações públicas e privadas. De um lado, a corrente que adota os chamados modelos racionais (com propósito prescritivo) e, do outro, a corrente dos modelos incrementais (de caráter descritivo).

Hebert Simon, representante importante do primeiro grupo, considera a decisão como uma escolha entre alternativas de ação de modo a melhor atingir determinados objetivos, tendo como referência um dado sistema de valores. A racionalidade associada a uma decisão não é necessariamente objetiva. Ou seja, a noção de decisão racional não significa que apenas decisões objetivamente tomadas, no sentido de serem reprodutíveis por cálculos que independam dos sujeitos, sejam legítimas. Pelo contrário, para Simon, racionalidade é um conceito relativo, a racionalidade de uma decisão depende dos valores e objetivos de quem decide e, portanto, deve ser sempre descrita ao lado de adjetivos que esclareçam seus diferentes sentidos (racionalidades objetiva ou subjetiva, consciente ou deliberada, organizacional ou pessoal) (SIMON, 1962, p. 73). Em particular, toda decisão objetivamente racional envolveria: 1. conhecer de modo amplo as possíveis estratégias de ação para se atingir um objetivo; 2. considerar todas as conseqüências que decorreriam de cada estratégia; 3. escolher, usando como critério seu sistema de valores, uma dentre todo o conjunto de alternativas (SIMON, 1962, p. 77).

Esses três passos, segundo a terminologia adotada neste capítulo, nada mais seriam que 1. identificação de objetos para avaliação; 2. definição de atributos levando em conta a projeção temporal dos efeitos que poderiam ocasionar; 3. aplicação do(s) critério(s) do decisor para ordenar as alternativas e, em seguida, produzir uma ação em função da alternativa melhor colocada. Ajustando a terminologia, o modelo de decisão objetivamente racional de Simon é compatível com o marco analítico proposto.

O objetivo de Simon era obter um modelo geral para decisões, de modo a prescrever sistemas de decisão que fossem baseados na racionalidade das organizações e não na racionalidade dos indivíduos que as constituem. A seleção das melhores alternativas de ação poderia, desta forma, ser objetivada segundo e somente por valores mais gerais da sociedade e de suas organizações sociais, tanto de caráter público como privado.

Para o autor, a eficiência de uma organização depende dela se libertar dos indivíduos que a compõem, cultivando suas próprias metas e sua própria racionalidade decisória. Devido às limitações da racionalidade humana, na prática pouco afeita à racionalidade objetiva, é que uma teoria administrativa tornar-se-ia necessária, para objetivar a racionalidade no âmbito das organizações. Desta forma, ele sugere, para fins práticos, o conceito de *racionalidade limitada*.¹² O processo de tomada de decisão poderia ser realizado em condições não ótimas, considerando apenas um conjunto restrito de alternativas e conseqüências, e mesmo assim atingir resultados satisfatórios. Simon, que na década de 1950 foi também um dos mais importantes precursores dos sistemas de inteligência artificial, procurava por um modelo de decisão com características prescritivas e autônomas para viabilizar racionalidades decisórias em instâncias coletivas, e que minimizassem o papel dos indivíduos. Para Simon, não importa o indivíduo, importam as organizações e, no fundo, os indivíduos atrapalham as organizações. São as organizações que precisam decidir e, restringir as opções de escolha é vantajoso por dois motivos. Primeiro, por tornar o processo operacional. Segundo, por limitar o escopo de alternativas plausíveis (e seguras) à organização, induzindo um conjunto de ações do tipo *S*, compartilhado por todos os indivíduos pertencentes à organização e relativamente imune a comportamentos oportunistas de qualquer um desses indivíduos.

Ham e Hill (1984) chamam esse programa de racional-liberal e pelo menos duas críticas são a ele freqüentemente dirigidas. Na realidade, as diferenças entre os valores de uma organização e dos indivíduos que a constituem não são homogêneas, nem constantes, nem previsíveis, especialmente quando o que se entende por organização inclui grandes corporações privadas ou instâncias governamentais. Se esse descompasso tomar proporções

¹² Apesar dessa ser a tradução correntemente utilizada para o original *bounded rationality*, não é muito precisa e pode gerar confusão. Não é apenas a racionalidade do indivíduo que é limitada, no sentido dele não ser totalmente capaz de avaliações objetivas, também as suas opções de escolha são efetivamente reduzidas por conseqüência dele estar inserido e obedecer às regras da organização. Talvez a tradução para *racionalidade cerceada* fosse mais adequada. Entretanto, feita a ressalva, mantém-se o termo consagrado.

muito amplas pode não fazer mais sentido falar em metas de uma organização como critérios para decisão, pois as metas dos indivíduos teriam influência decisiva. A própria organização necessariamente se altera, evolui ou perece. Na prática, não haveria uma separação tão nítida entre a organização e seus indivíduos constituintes. Uma segunda vertente de crítica identifica a incerteza do ambiente de atuação da organização e dos indivíduos como um fator fundamental para a possibilidade de racionalidade objetiva. Em períodos que a incerteza dos mercados aumenta nem mesmo a organização seria capaz de racionalidade objetiva. Assim, imaginar decisões que ocorrem de forma otimizada (em termos racionais) estaria mais para uma atitude normativa que representaria, no máximo, situações de exceção, do que para uma teoria que efetivamente descreva a tomada de decisão como ela ocorre na prática.

Outro modo de entender as decisões é adotado pela corrente incrementalista, representada aqui por Charles Lindblom. Em 1959, Lindblom lança o conceito de *comparações sucessivas limitadas* para tentar descrever como realmente as decisões ocorrem nas organizações sociais. Para Lindblom, não é possível separar o processo de decisão em uma organização dos processos políticos que a permeiam. Para o autor, o processo político faz necessariamente contraponto aos processos ditos racionais (LINDBLOM, 1980). O modelo incremental é caracterizado por uma estrutura ramificada, partindo das situações existentes e propondo modificações incrementais a elas. O processo incremental pressupõe que quem toma as decisões pode mudar de rumo de acordo com as consequências sentidas a cada mudança incremental; não há a escolha otimizada em função de um resultado claramente definido, como nos modelos racionais. Uma variação do modelo (o incrementalismo desarticulado) é considerada por Ham e Hill (1984, p. 88-89) a melhor descrição de como as decisões são realmente tomadas. Tal formato envolve o exame apenas de alternativas familiares e há fragmentação tanto em relação às consequências consideradas quanto em relação à delimitação dos objetivos. Uma ação é originalmente direcionada a um problema, sua implementação é tentada, ela é alterada, tentada em sua forma alterada, tentada novamente e assim por diante, com a participação desarticulada de muitos atores.

Um modelo assim é descritivo e não é capaz de prever autonomamente quais deveriam ser as decisões.

A definição adotada na seção anterior para decisão, vinculando-a a um processo recorrente de cognição, avaliação e decisão, é também incrementalista em sua essência.

Em geral, para se concretizar ações sobre temas mais complicados, isto é, aqueles que envolvem muitos objetos e muitos juízos, vários estágios intermediários de avaliação e decisão precisam ser percorridos. Desse modo, exceto para os casos mais simples, as decisões práticas são na realidade concatenações de muitos processos de avaliação entremeados, por definição, por suas correspondentes decisões. Isso coincide com a interpretação incremental do processo decisório. O processo até uma decisão sobre uma questão complicada ou complexa¹³ não é algo monolítico, e sim um intrincado jogo interativo que pode ramificar em diversas direções e que avança passo a passo, com possíveis saltos, retornos e descontinuidades.

Uma ação pode modificar o estado dos objetos percebidos pelos sujeitos e motivar novas avaliações, que levarão a novas decisões, e assim por diante. Os juízos utilizados a cada etapa não necessariamente guardam coerência entre si, especialmente se não houver um quadro de referência *a priori*, e apenas o sujeito da avaliação pode ter controle disso – e muitas vezes nem ele o tem. Isto está de acordo com a imprevisibilidade da decisão apontada pelos incrementalistas. O comportamento dos sujeitos parece não ser modelável e, no entanto, são esses mesmos sujeitos que, em última análise, fazem a avaliação que resulta na decisão.

A discussão sobre limites e formas de racionalidade é ampla na literatura de disciplinas como economia, administração, ciência política e psicologia.¹⁴ O pressuposto de racionalidade objetiva é central para a teoria econômica e para certos modelos comportamentais usados nas demais disciplinas. No próximo capítulo, mostra-se que tal premissa guarda estreita ligação com uma estrutura algébrica específica necessária para se operar objetivamente os modelos teóricos construídos. Eliminar a referência abstrata de

¹³ Os adjetivos *complicado* e *complexo* são utilizados neste trabalho referindo-se a características diferentes. Um sistema só é complexo quando seus elementos se influenciam mutuamente, isto é, possui relações não-lineares do tipo $a = f(b)$ e $b = f(a)$ simultâneas. Quando há complexidade, é impossível determinar a evolução do sistema analiticamente, sendo necessário simplificá-lo para resolvê-lo logicamente. Por sua vez, um sistema complicado é todo aquele composto por intrincadas relações. Esta mesma distinção é adotada, entre outros, por Bresciani Filho (1999) e Loveridge (2002). Uma relação pode ser, portanto, complicada sem ser complexa ou ser complexa sem ser complicada.

¹⁴ Cf., por exemplo, a coletânea editada por Cook & Levi (1990) e Vanberg (1994) para economia, McCubbins & Thies (1996) para ciência política.

racionalidade objetiva significa também eliminar a possibilidade de se construir teorias sociais com capacidade de previsão.

Por isso, esse é um ponto delicado. De um lado, vem de longe a crítica à ingenuidade dessa premissa e de sua inadequação como postulado para teorias sociais. De outro, o argumento de preservar a capacidade prescritiva de uma teoria parece irresistível a muitos. Do ponto de vista pragmático da racionalidade organizacional, Simon (1962, p. 60) já afirmava que sua análise "não é uma descrição de como decidem os administradores reais, e sim uma descrição de como decidem os *bons* administradores".

Independentemente das virtudes ou fragilidades da idealização da racionalidade nas teorias sociais, o esquema analítico para avaliações apresentado neste capítulo precisa ser capaz de acomodar a aparente volatilidade dos valores dos indivíduos e suas múltiplas racionalidades. Avaliações ocorrem a todo o momento, com ou sem racionalidade objetiva, com ou sem teoria. Em outras palavras, como explicar com os elementos do esquema analítico apresentado até aqui as diferentes formas de racionalidade que os indivíduos podem apresentar?

A partir do esquema analítico exposto, uma seqüência do tipo *avaliação, decisão, avaliação, decisão etc.* pode ser identificada com *aprendizado*. Tanto a recorrência de seqüências de avaliação e decisão como o aprendizado são processos evolutivos e irreversíveis mediados pelos indivíduos. Em cada indivíduo, a sucessão de *aprendizado* que se acumula é única e fornece exemplos que se tornam parâmetros que moldam seus juízos subseqüentes. A capacidade de aprendizado e a unicidade dos padrões desse aprendizado são propriedades que garantem a impossibilidade de determinação (ou modelização) da estrutura cognitiva e de julgamento dos indivíduos. Logo, a existência de racionalidades distintas nas decisões significa simplesmente que os indivíduos são estruturalmente diferentes enquanto sistemas geradores de decisão. A aproximação a um comportamento estável que possa ser identificado com alguma racionalidade em particular é algo que só é possível se os indivíduos compartilharem de referenciais cognitivos bastante similares.

A esta altura, pode causar estranheza o fato de que, segundo o marco analítico proposto, apenas indivíduos possam decidir. Segundo as definições adotadas, a decisão é uma prerrogativa do indivíduo. Fica claro que dadas as mesmas condições dois indivíduos somente poderão decidir da mesma maneira se ambos compartilharem do mesmo processo

de avaliação. Nesse caso, cada indivíduo realiza consigo próprio medidas e juízos do mesmo modo e as duas avaliações geram duas decisões que por decorrência são idênticas.¹⁵ O rigor em definir que cada indivíduo avalia e decide segundo ele próprio é importante para distinguir essa situação daquela em que a avaliação acontece como fenômeno social. A avaliação socializada se distingue da que o indivíduo realiza porque comporta formas e métodos bem definidos e acontece apartada dos indivíduos. Ela se diferencia da avaliação *stricto sensu* definida acima porque ocorre no plano social. Porém, há um nexo geral segundo o qual a avaliação socializada influencia a avaliação do indivíduo, conforme será discutido adiante.

Uma definição geral de avaliação precisa de algum modo acomodar esses dois níveis de uso do conceito. A avaliação socializada pode ser entendida como um processo de legitimação social das decisões individuais. A distinção de dois níveis de avaliação, aquela que ocorre no indivíduo e a executada (ou explicitada) socialmente é outro fator que ajuda a explicar a multiplicidade de valores nos decisores. Os valores explicitados socialmente não precisam ser os valores que de fato os indivíduos empregam.

Todavia, antes avançar para o exame de uma terceira hipótese do marco analítico, necessária para distinguir os dois tipos de avaliação, é oportuno explorar um pouco mais o alcance das definições sugeridas até o momento. Na prática das avaliações, o suposto de causalidade é alvo de inúmeras discussões teóricas e metodológicas. Na próxima seção, a noção de causalidade é enunciada e interpretada de modo abstrato com o auxílio do marco analítico proposto, com o intuito de destacar mais alguns pontos importantes desta discussão geral sobre avaliação.

V. A NOÇÃO DE CAUSALIDADE

A noção de causalidade é o cerne da interpretação lógica dos eventos. Uma estrutura de explicação típica consiste em estabelecer uma determinação lógica entre eventos distinguindo aqueles que são causas daqueles que são efeitos. No universo físico observável – e portanto passível de constatação empírica – uma *teoria* é todo enunciado capaz de

¹⁵ Essas asserções são válidas para a descrição de uma célula analítica de avaliação e decisão. É claro que para um processo real de avaliação ou decisão, diferentes rotas ou diferentes métodos podem implicar nas mesmas decisões, ou nas mesmas ações finais. Analiticamente, diferentes caminhos representam diferentes tramas obtidas com muitas células de avaliação e decisão, independentemente do resultado final ser ou não o mesmo.

determinar os efeitos observados a partir de causas hipotéticas. Em outras palavras, trata-se de um enunciado que afirma que se certos eventos ocorrerem então determinados eventos conseqüentes necessariamente ocorrem.¹⁶

O método científico aplicado às ciências naturais consiste em procurar as hipóteses que melhor expliquem certos fatos recorrentemente observados. Uma vez aceita uma hipótese como explicação plausível, novas hipóteses podem ser tentadas para explicá-la, e assim por diante. O fato das hipóteses iniciais de um sistema explicativo não serem explicadas, ou seja, não serem causadas, não afeta a validade do sistema, o importante é sua capacidade de explicar e corroborar os efeitos previstos.¹⁷ A ciência procura encontrar e explicar regularidades e padrões de mudança nos atributos dos objetos, e a causalidade é o mecanismo que permite isolar e prever no tempo e no espaço os fenômenos.

A decisão conforme definida na seção III é causada pela avaliação, porém numa relação de causa-efeito totalmente contingente, por conta da imprevisibilidade e liberdade pressupostas nos indivíduos. A noção de decisão pode, portanto, ser entendida como uma generalização da idéia de causalidade, uma vez que é mais ampla que esta. Ou, de modo inverso, a causalidade pode ser entendida como um caso particular de decisão, para o caso de decisões previsíveis.

Para os sujeitos, toda mudança observada, seja regular ou não, implica em avaliação, decisão e ação. E a tal mudança é causada por ação anterior. Desse modo, pode-se sempre dizer que alguma decisão causou as mudanças observadas por um sujeito, não importando

¹⁶ Segundo as palavras de Weber, a explicação causal "significa que, de acordo com uma determinada regra de probabilidade, qualquer que seja o modo de calculá-la – e somente em casos raros e ideais isso ocorre a partir de dados mensuráveis, a um determinado processo observado segue outro processo determinado" (WEBER, 2004, p. 11).

¹⁷ Esta definição está de acordo com o pensamento de Popper no que tange ao caráter eminentemente dedutivo das teorias científicas. Popper rejeita a possibilidade de se obter enunciados verdadeiros a partir da indução (generalização) de fatos empíricos e, portanto, particulares. Para ele, uma teoria científica trata de fatos universais que são por definição impossíveis de verificação completa; os testes empíricos fornecem sempre verificações parciais porque representam observações que são apenas algumas instâncias da classe universal de fenômenos a que pertencem. Desse modo, eles não "provam" uma teoria e não podem levar a hipóteses gerais, apenas "corroboram" ou não os efeitos deduzidos de hipóteses testadas. É preciso ter claro que a "validade" de uma teoria não está nos fatos empíricos que a corroboram, mas no próprio sistema lógico que permite deduzir os efeitos das causas. Uma hipótese é assumida como verdadeira se os efeitos previstos forem corroborados pelos fatos empíricos, isto é, se o enunciado *Hipótese* $\Rightarrow$ *Efeito* se sustentar (POPPER, 1975). Russel também chama a atenção para a importância das hipóteses no método científico. Afirma que o processo de elaboração de hipóteses é intuitivo e por isso geralmente é mal compreendida sua função para a ciência. Não há nada que impeça a construção de qualquer hipótese, mesmo que contrarie o senso comum, e seja aparentemente descabida, sua validade é função apenas de sua capacidade de "salvar as aparências" (RUSSEL, 2002).

quem teria tomado tal decisão.

Nos fenômenos naturais, as mudanças observadas podem, portanto, ser interpretadas como resultantes de decisões fixas, imutáveis, independentes da existência de um sujeito as realize. O objetivo original da investigação científica de encontrar as *leis naturais* do universo, os princípios que determinariam todas as mudanças passadas e futuras, é em sua essência uma tentativa de compreender e racionalizar todas essas "decisões".

Pode parecer artificial insistir na idéia de decisão para interpretar os fenômenos naturais, mas historicamente foi justamente o contrário: a particularização da explicação da natureza como causalidade, como encadeamento de fatos plenamente determinados, que modificou a interpretação dos eventos do universo sensível e introduziu a explicação científica racionalmente objetiva.

Nem sempre foi assim. Por exemplo, por muito tempo a impossibilidade de determinar a causa inicial legitimou a explicação divina para a decisão sobre o estado inicial do universo, sendo que os critérios utilizados para chegar a essa decisão seriam inacessíveis aos homens. O determinismo dos fenômenos naturais seria, portanto, decorrência de um juízo divino invariavelmente correto e onipresente, suas decisões no mundo físico a todo o momento se repetiriam dando a aparência de ordem e previsibilidade. O fato dos critérios divinos serem repetidas vezes os mesmos indicava a correção e a perfeição desses juízos. E quando eventualmente não o fossem, indicavam o poder divino de decidir sobre os eventos.

Se o determinismo é sinal da perfeição da natureza, o homem, sendo por suposto imperfeito, teria a ilusão do livre arbítrio, ou melhor, a incapacidade de saber *a priori* as decisões corretas. Ou seja, teria a ilusão de interferir no fluxo dos acontecimentos decidindo segundo sua própria avaliação dos fatos, segundo suas próprias medidas e juízos. Isso justifica o argumento que impõe então determinados preceitos éticos ou religiosos – supostamente ditados por divindades – para anular o livre arbítrio e dotar as decisões dos indivíduos de correção. Como consequência, o comportamento humano ganha previsibilidade e critérios para que se meça a correção das decisões. Seria inclusive possível estender esse argumento para explicar como diversas éticas levaram, em várias épocas, a formas diferenciadas de valorar decisões como corretas ou incorretas.

Segundo o Princípio 2, a avaliação é necessária para o indivíduo decidir. Assim,

para entender uma decisão é preciso compreender a avaliação que a precede. Por outro lado, do ponto de vista da natureza, uma sequência de eventos aparentemente prescinde de juízos para ocorrer. Entretanto, para que um indivíduo entenda as decisões da natureza, ele precisa interpretá-los segundo conceitos que são uma construção humana. O homem necessariamente faz julgamentos para decidir quais atributos considera relevantes medir para encontrar o mecanismo de decisão oculto. E tal mecanismo contém de fato uma avaliação implícita, há um juízo embutido na aceitação da correção da teoria explicativa. Para os eventos objetivos da natureza parece haver uma explicação objetiva, mas na realidade ambos, objetos e explicação, são condicionados por juízos de validade intersubjetivos.

Isso remete ao fato que, em todo caso, há simultaneidade entre a verificação empírica de uma conclusão causal de uma teoria objetiva e uma decisão sobre sua veracidade que utiliza um critério de validade lógica, que ocorre no indivíduo. Independentemente do que motiva a sucessão de eventos no universo físico, a verificação dessa ocorrência por um observador é um processo que envolve avaliação e decisão, e sem esse processo, o indivíduo seria incapaz de ter consciência de uma mudança.

Desse modo, por um lado faz sentido supor toda sucessão de eventos do universo físico como uma manifestação de uma *decisão natural* que, mesmo que possua algum tipo de avaliação que a preceda, é irrelevante para que o observador tome dela consciência. Em compensação, a interpretação de qualquer mudança no universo físico só é possível com um critério de validade que parte do indivíduo, e por consequência, de uma avaliação sua.

Para o indivíduo, determinadas conclusões a que ele próprio chega aparentam ser mais determinadas enquanto outras aparentam ser mais arbitrárias. Para uma decisão do indivíduo ser considerada correta, seja ela aparentemente arbitrária ou não, é preciso que outros indivíduos assim a considerem. A correção de decisão, e mesmo seu valor como causalidade é, portanto, um critério socialmente definido.

Desse modo, as regras de avaliação que levam a uma decisão considerada correta representam um mecanismo para legitimar uma decisão. Os indivíduos que explicitam as regras de suas próprias avaliações as fazem compreensíveis a outros indivíduos e tornam suas decisões consequências claras de seus "princípios". Se suas regras não coincidem com aquelas socialmente explicitadas e aceitas a avaliação não é racionalizável e a decisão é,

socialmente, lida como arbitrária.

VI. AVALIAÇÃO COMO LEGITIMAÇÃO SOCIAL

Até aqui a avaliação, e sua consequência necessária no indivíduo, a decisão, foram descritas a partir de seus componentes internos, formando uma célula básica que pode ser repetida muitas vezes e agrupada em tramas complicadas para analisar a consecução de um determinado processo real de avaliação ou decisão. Por sua vez, esta seção avança na interpretação da avaliação como fenômeno social. O caráter social dos processos de avaliação pode ser introduzido a partir de um novo princípio.

Princípio 3: Todos os métodos de avaliação são construções sociais.

Foi postulado que, independentemente do modo que uma avaliação se apresente do ponto de vista da função que leva atributo a critério, ela em sequência implica uma decisão. Entretanto, embora o indivíduo geralmente tenha consciência de que decide, isto é, de que de algum modo age, ele nem sempre tem clareza de como decide, ou seja, de qual processo – ou processos – de avaliação utiliza para chegar a determinada ação.

Alguns desses processos são tão intrínsecos ao indivíduo que as decisões que deles resultam são equivalentes a decisões naturais, do tipo causa e efeito, comentadas na seção anterior. Por outro lado, outros levam a raciocínios complicados que incluem muitas etapas com decisões intermediárias e a necessidade de critérios gerais de validade que permitam um desenrolar coerente das etapas seguintes. Critérios de validade aceitos por diversos indivíduos colaboram para a estabilidade social porque permitem a previsibilidade dos juízos e portanto das decisões quando determinadas condições estiverem presentes.

Ao conseguir clareza sobre um processo de avaliação e conseguir explicitá-lo e comunicá-lo a um outro indivíduo, incluindo a exteriorização de seus critérios de validade, o sujeito torna a avaliação intersubjetiva, ela adquire um método e se torna um instrumento para a legitimação social da decisão tomada pelo indivíduo.

O Princípio 3 afirma que todos os métodos de avaliação, entendendo-se por métodos os mecanismos pelos quais os indivíduos executam as avaliações, são criados e aprendidos

pelos indivíduos a partir do convívio com outros indivíduos. Portanto, pertencem a uma esfera externa a cada indivíduo, são codificados e compartilhados socialmente. São construções sociais na medida que correspondem aos critérios coletivamente aceitos para legitimar ações. Essa distinção esclarece os dois níveis, mencionados anteriormente, nos quais podem se dar as avaliações.

A avaliação consciente só faz sentido se interpretada socialmente. As ações óbvias, no sentido de se aproximarem das decisões naturais, possuem avaliações tão diretas e regulares que podem perfeitamente passar inconscientes. As decisões lógicas, ou seja, aquelas apoiadas em princípios de validade socialmente aceitos da racionalidade objetiva precisam ser aprendidas.

O sentido de aprender aqui é o mesmo definido há pouco. O indivíduo adquire a capacidade social de avaliar, ou seja, de identificar, mensurar e julgar objetos concretos e abstratos à sua volta a partir de uma sucessão evolutiva de avaliações e decisões que se inicia de modo rudimentar e adquire progressivamente sofisticação de acordo com o meio social com o qual interage. Assim, ele adquire método.

O caráter evolutivo desse processo significa que ele é caracterizado por incerteza e embate entre concepções mais ou menos divergentes que emergem do aprendizado e são constantemente submetidas ao teste da aprovação social. A variável aprovação social não é necessariamente binária e tampouco necessariamente definitiva. É um resultado de síntese das avaliações realizadas por outros indivíduos, de modo que possa subsistir ambigüidade sobre a aprovação ou não sobre algo aprendido pelo indivíduo.

O aprendizado sobre como avaliar é condicionado em grande medida pelas práticas de avaliação correntes em determinada sociedade. Um método de avaliação pode persistir por muito tempo como construção social se a identificação dos atributos for sedimentada, por exemplo, em uma determinada forma lingüística e os juízos orientados por um determinado preceito ético ou cultural. Entretanto, uma vez questionada e superada por outro método, passa a existir aprendizado que não pode ser revertido senão pela eliminação dos sujeitos que o detêm, assim como de seus produtos de comunicação. Valem, portanto, as propriedades de irreversibilidade e cumulatividade para o aprendizado social de métodos de avaliação e decisão.

Para que se legitime socialmente ações que contrariam o critério socialmente aceito

de correção é preciso primeiro explicitar os elementos das avaliações que levaram a elas e, em seguida, convencer tal sociedade da pertinência do novo método ou critério. Apenas mostrar quais atributos e quais critérios estão por trás de uma decisão não garante que ela se legitime. Do mesmo modo, para legitimar uma decisão não é necessário que se mostre todas os atributos e todos os critérios por trás de decisão, pode-se mostrar apenas o conveniente, a parte socialmente aceitável. O importante é que um indivíduo só aceita a decisão de outro se ele compartilhar de um mesmo sistema de avaliação ou se julgar (segundo seus próprios critérios) que deve concordar com tal decisão.

Uma importante consequência disso tudo é que uma avaliação com método pode ser uma atividade coletiva. Retomamos o raciocínio até aqui delineado para explicar a avaliação coletiva. Definimos a decisão como sendo uma prerrogativa individual. A cada decisão corresponde também uma avaliação individual, consciente ou não, com um método socialmente construído. Havendo método, nada impede que sejam previstos mecanismos de participação coletiva na avaliação. Pelo esquema analítico proposto, não faz sentido falar em decisão coletiva, porém processos de avaliação coletiva podem influenciar as avaliações individuais e, por consequência, as decisões individuais.

Desse modo, o fato de o resultado de uma avaliação coletiva ser aceito pelos indivíduos revela novamente que ou esses indivíduos compartilham em grande medida do mesmo sistema de avaliação ou avaliam pessoalmente que é estratégico ou prudente aceitar tais resultados. Novamente, a aprovação social de uma avaliação (e sua consequente decisão) é algo circunstancial e definitivamente não pode ser representada por um critério binário de verdade ou falsidade. É insatisfatório tomar as noções de aceitação e rejeição sociais de uma avaliação como sendo mutuamente excludentes e ao menos uma delas necessária. Pode haver um vasto conjunto de resultados frente aos quais os indivíduos sejam indiferentes ou demonstrem hesitação.

A implicação entre uma avaliação coletiva e uma decisão não é direta e tampouco necessária. A menos que as condições apresentadas acima se verifiquem, não é possível garantir que uma avaliação coletiva torne-se uma decisão coletiva socialmente ótima. Uma avaliação coletiva é necessariamente uma construção sofisticada que provoca a interação de diferentes indivíduos com códigos de procedimento codificados e supostamente universais mas que nem sempre são facilmente e homogeneamente interpretados pelos processos

individuais de avaliação. É, portanto, em contraste com a avaliação individual, um processo político de convencimento acerca da validade de decisões e não um mecanismo objetivo de decisão. A avaliação coletiva forçaria a adequação da avaliação individual produzindo "verdades" sociais, enquanto a avaliação individual controlaria a decisão do indivíduo frente a essas "verdades". Como a primeira não implica efetivamente a decisão, ao indivíduo resta sempre a possibilidade de discordar.

Com isso, completa-se a proposição do esquema analítico, alvo deste capítulo. Conforme previsto de partida, há ainda duas aplicações escolhidas para testar sua coerência. Em ambas, problemas relevantes das ciências sociais serão examinados à luz das definições adotadas para avaliação e decisão.

VII. O PROBLEMA DA ESCOLHA SOCIAL

Pelo que foi discutido na seção anterior, não faz sentido pensar em decisão ou escolha social. Entretanto, esse foi um objetivo extensamente perseguido por inúmeros autores e é especialmente relevante para a discussão de métodos coletivos de avaliação. O problema da escolha social, inicialmente tratado por matemáticos e economistas, consiste em pesquisar os melhores métodos para agregar as opiniões dos indivíduos de modo a determinar a decisão (escolha) coletiva capaz de melhor refletir as opiniões individuais. O tratamento dessa questão apresentou desde a polêmica entre Borda e Condorcet contornos eminentemente normativos.¹⁸

O pensamento sobre escolha social se inicia com o tratamento sistemático do problema do voto por Borda e Condorcet, respectivamente em 1781 e 1785.¹⁹ A ambos interessava encontrar o método ótimo para interpretar o resultado de uma eleição. Para tanto, o problema que se colocava podia ser definido assim: dadas diferentes alternativas a

¹⁸ Mueller, na exaustiva revisão que assina sobre escolha social, mostra que nos últimos vinte anos o campo progressivamente abandonou o caráter normativo a favor de investigações empíricas. Entretanto, afirma que os famosos resultados teóricos das décadas de 1950 e 1960 são ainda válidos e, de fato, consistem na principal motivação dos estudos atuais, que procuram evidências empíricas para avançar em direção a novas perspectivas. Cf. Mueller (1997, p. 8-9). Por esse motivo, esta seção se atém a comentar apenas os resultados mais relevantes do campo, as investigações de Condorcet e Borda e o Teorema da Possibilidade de Arrow.

¹⁹ Portanto, às vésperas da Revolução Francesa de 1789. O artigo seminal de Jean-Charles de Borda chamava-se *Mémoire sur les élections au scrutin*, publicado no *Mémoires de l'Académie Royale des Sciences*, 1781, pp. 657-65 e a resposta do Marquês de Condorcet *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*, 1785 (ARROW, 1963, pp. 93-94).

uma sociedade, composta de indivíduos independentes, qual a forma de eleição que resultaria numa escolha socialmente ótima, ou conforme Condorcet, fosse capaz de encontrar a escolha correta?

Moulin (1988) explica com detalhes a controvérsia entre Condorcet e Borda quanto à solução do problema. Ambos concordavam que o método mais comum de eleição, a escolha do candidato preferido pelo eleitor e a decisão por maioria simples, não era adequado na presença de mais de dois candidatos. Argumentavam que o método poderia levar a resultados indesejáveis. Para entender o que eles queriam dizer, basta imaginar uma situação em que os eleitores ao invés de explicitarem apenas a escolha do candidato que consideraram o melhor fossem convidados a indicar em qual ordem de preferência todos os candidatos se localizavam. Os resultados hipotéticos mostrados abaixo correspondem a um exemplo elaborado por Moulin (1988, pg. 228).

3	5	7	6
<i>a</i>	<i>a</i>	<i>b</i>	<i>c</i>
<i>b</i>	<i>c</i>	<i>d</i>	<i>b</i>
<i>c</i>	<i>b</i>	<i>c</i>	<i>d</i>
<i>d</i>	<i>d</i>	<i>a</i>	<i>a</i>

A interpretação é a seguinte. De um universo de 21 eleitores e quatro alternativas *a*, *b*, *c* e *d* a primeira coluna reúne 3 eleitores que ordenaram suas preferências como sendo $a > b > c > d$, a segunda coluna 5 eleitores que consideraram que $a > c > b > d$ e assim por diante. Por maioria simples, a alternativa *a* seria a escolhida com 8 votos. Entretanto, ela é também a alternativa classificada como sendo a pior por todos os demais eleitores. E mais, é a pior por maioria absoluta, 13 votantes. Claramente, afirmariam Condorcet e Borda, o resultado é insatisfatório e novos métodos precisavam ser propostos.

Para Condorcet, o procedimento correto seria comparar todos os candidatos dois a dois e eliminar aqueles superados por outras alternativas. Desse modo, *a* supera as demais alternativas para 8 eleitores, mas é preterido por qualquer alternativa para 13 eleitores, portanto, deve ser eliminado. Outra maioria prefere *c* a *b* (11 contra 10) e, novamente, *c* a *d* (14 contra 7). Disso resulta que a alternativa que corresponde à escolha ótima seria o

candidato c e não o candidato a eleito pelo método pluralista convencional.²⁰

Por sua vez, Borda sugere outro procedimento. A Regra de Borda consiste em considerar cada ordenação de alternativas como uma hierarquia à qual o último membro (menos preferido) é valorado com nota zero, o penúltimo com nota 1, o antepenúltimo com nota 2 e assim por diante até assinalar ao primeiro colocado $n-1$ pontos. A melhor escolha seria a alternativa que mais somar pontos globalmente. Se aplicarmos essa regra ao exemplo acima, teremos: $a = 6$, $b = 8$, $c = 7$ e $d = 3$. Pelo método de Borda, a alternativa a seria também classificada como uma escolha inferior. Entretanto, o método indicaria a alternativa b como sendo a escolha ótima, em contraste com a proposta de Condorcet.

Young (1997) conta que a controvérsia entre os dois nunca foi solucionada e, de fato, cada abordagem deu origem a diferentes métodos para escolha social. Os métodos de comparações dois a dois sucessivas e os métodos de pontuação (*scoring*) são os exemplos mais diretos. Desde então, muitas variações desses métodos e outros procedimentos foram propostos, porém sem conclusão sobre qual linha seria a mais correta.²¹ A controvérsia entre Condorcet e Borda é de fato o primeiro resultado negativo quanto a possibilidade de um método ótimo para a escolha social.

Esse dilema ficaria definitivamente explícito com o famoso Teorema da Possibilidade de Arrow, cuja primeira formulação data de 1951.

Arrow (1963) descarta de partida a possibilidade de se obter uma medida cardinal (quantificada) das preferências dos indivíduos de modo que se possa comparar ou somar tais preferências. Com isso rechaça integralmente procedimentos do tipo Regra de Borda.

²⁰ Se não houver uma alternativa que supera todas as demais, isto é, um *vencedor*, surge a situação conhecida como Paradoxo de Condorcet, ou paradoxo do voto, na qual as comparações levam a um ciclo não-transitivo. Isto é, quando, por exemplo, dadas 3 alternativas a tenha maioria sobre b , b sobre a e c sobre a .

²¹ Moulin (1988, pg. 237-240) descreve uma série de resultados teóricos mais recentes que concluem que a abordagem de Borda é consistente em relação a algumas propriedades consideradas desejáveis, enquanto a de Condorcet não. Entre elas estão as propriedades de 1) Reforço (*Reinforcement*): se dois grupos disjuntos de eleitores N_1 e N_2 escolhem o mesmo candidato a de um conjunto A de opções, então o eleitorado $N_1 \cup N_2$ deve também eleger a dado A ; 2) Participação: se um candidato a é eleito de um conjunto A por um eleitorado N e um votante i externo a N tem a como preferência, então $N_1 \cup \{i\}$ deve escolher a ; e 3) Continuidade: se N_1 elege a de A e N_2 elege b de A é possível replicar N_1 m vezes de modo que o eleitorado $(mN_1) \cup N_2$ escolha a . O Teorema de Young (de 1975) mostra que apenas os métodos de pontuação (tipo Borda) obedecem simultaneamente as propriedades de Reforço, Continuidade, Anonimato (independência dos eleitores, se eles trocarem seus votos a eleição não se altera) e Neutralidade (independência das alternativas, se os nomes das alternativas forem trocados para todos os eleitores o resultado não se altera). Os métodos tipo Condorcet com mais de 3 alternativas não obedecem a propriedade de Participação e podem, em situações específicas desobedecer as demais. Todavia, métodos envolvendo a comparação sucessiva de alternativas e seleção por maioria são muito utilizados por sua simplicidade são considerados mais afins ao espírito das decisões democráticas.

Para ele, a preferência dos indivíduos apenas pode ser representada pela ordenação das alternativas disponíveis para escolha. Arrow definiu como problema de escolha social encontrar um método capaz de produzir uma ordenação síntese das ordenações realizadas pelos indivíduos. Para isso, além da recusa a comparações cardinais, Arrow supôs outras condições que correspondem às propriedades que tal método deveria obedecer.

A escolha social pode ser realizada por quatro meios. Primeiro, pela convenção. A escolha dos indivíduos pode ser condicionada por uma tradição ou moral que os faça decidir todos do mesmo modo, consciente ou inconscientemente. Num outro extremo, a escolha pode ser a escolha de um ditador, ou de um grupo ditador que imponha a decisão à revelia dos demais indivíduos. Apesar dessas formas de decisão social não serem incomuns, na prática, um método para agregar as preferências dos indivíduos não deveria induzir, segundo Arrow, nenhuma dessas situações. As situações típicas pressupostas pelo autor são as escolhas sociais que resultam do voto ou das transações de mercado. Analisando a ambas, Arrow mostra que transações de mercado podem ser entendidas como um caso particular da situação de voto.

O indivíduo é assumido como tendo plena liberdade de ordenar as alternativas disponíveis como bem entender. A premissa de racionalidade exigida se limita à suposição de que os indivíduos reconheçam as alternativas disponíveis e sejam capazes de ordená-las. Formalmente falando, a ordenação exigida é do tipo parcialmente ordenada, ou seja, uma ordenação na qual para cada par de alternativas possa se estabelecer uma relação de preferência ou uma relação de indiferença e valha a propriedade de transitividade. Essas mesmas propriedades são esperadas para a ordenação social das alternativas. O método procurado deve ser capaz de fornecer uma ordenação completa das alternativas. Além disso, se todos indivíduos em um dado momento alterarem seu voto melhorando a posição da alternativa x em suas ordenações então a escolha social deve ou refletir essa alteração ou, pelo menos, permanecer inalterada (propriedade de associação positiva ou monotonicidade).

Arrow considerou as alternativas disponíveis para escolha como objetos genéricos, podendo representar desde candidatos ou produtos num mercado até vetores contendo todos os diferentes atributos aplicáveis a descrição de cada estado social. Parte do princípio de que o método deve ser aplicável a escolhas envolvendo quantas alternativas se deseje. Por

fim, espera-se que o método não seja sensível a alternativas externas ao pleito considerado nem que a ordenação seja alterada pela retirada de alguma alternativa (propriedade da independência a alternativas irrelevantes).

Definido rigorosamente o problema e as propriedades desejadas, o Teorema da Possibilidade demonstra que para três ou mais alternativas é impossível sustentar todas as condições simultaneamente. Ou seja, não existe um método capaz de gerar uma ordenação síntese a partir das ordenações dos indivíduos que preserve simultaneamente as condições impostas de liberdade individual e racionalidade, associação positiva, independência a alternativas irrelevantes e escolha não ditatorial.

Esse resultado foi devastador. O próprio autor adiantava a principal conclusão, dado que a maior parte das premissas é dificilmente removível, por serem premissas estruturais da racionalidade esperada para um método, restava unicamente a possibilidade de uma escolha racional pela supressão da premissa de escolha não ditatorial, a única não necessária à construção de um método com comportamento aceitável (lógico-racional).

Na sequência, vieram outros trabalhos que mostraram outros resultados surpreendentes e quase sempre pessimistas em relação à possibilidade de escolha social. Entre eles se destacam os trabalhos de Amartya Sen. Seu mais famoso resultado demonstra a incompatibilidade entre a liberdade de escolha do indivíduo e o princípio econômico do equilíbrio geral simbolizado pelo postulado de que uma sociedade realiza suas escolhas segundo o princípio de Pareto²² (SEN, 1997 p. 25). Pattanaik (1997, p. 225) expõe com clareza o ponto em que atualmente se encontra a pesquisa sobre escolha social. Em suas palavras:

A questão que naturalmente se coloca é o que exatamente se conseguiu com este grande e diverso conjunto de resultados negativos. A verdade é que nós ainda não temos nenhuma "solução" para o problema de como as avaliações de bem estar social devem ser feitas ou de como a sociedade pode fazer suas escolhas. Entretanto, os resultados de impossibilidade da teoria da escolha social contribuíram para nosso entendimento acerca de vários aspectos desse problema complexo. Eles nos fizeram conscientes de muitas das armadilhas existentes em nossa busca por critérios satisfatórios para a avaliação social do bem estar.

²² O equilíbrio de Pareto é atingido quando um indivíduo não pode melhorar seus ganhos sem levar à redução do ganho dos demais indivíduos.

É no exato sentido que aponta Pattanaik que o trabalho dos autores da escolha social é útil ao estudo da avaliação. Se entendermos toda avaliação formalizada como sendo uma avaliação socialmente construída e esperarmos que haja participação de indivíduos para a consecução de uma avaliação coletiva, então as restrições levantadas pelos autores da escolha social não podem ser ignoradas. Os limites apontados pelos resultados mencionados acima precisam ser levados em conta e controlados.

Não por acaso, o marco analítico que foi definido para avaliações exclui, de partida, a possibilidade de decisão social. Desse modo, as questões metodológicas foram recolocadas segundo um ponto de vista menos ambicioso. Os métodos de avaliação coletiva seriam formas de apoiar ou influenciar as decisões e não as decisões em si. Não há a necessidade de encontrar respostas corretas conforme acreditava Condorcet, nem mesmo de obedecer critérios rígidos de racionalidade, conforme defendeu Arrow. Isso porque a avaliação coletiva não deve ser vista como racionalizadora ou objetivadora do processo político, ela é de fato uma forma de organizar e mediar um processo político, que pode ser mais ou menos racionalizado, mas não uma substituta da decisão.

Ainda no espírito de aprender com as armadilhas já desarmadas, esta seção se encerra com mais uma rodada sobre as concepções dos autores clássicos da escolha social. Suas premissas e abordagens são de grande importância para identificar as raízes de diversas discussões metodológicas (e também de alguns preconceitos) que persistem até o presente.

Em seu trabalho, Condorcet acreditava que não se tratava apenas de encontrar a melhor escolha, mas a escolha correta (MUELLER, 1997). Para ele, os indivíduos podiam acertar ou errar uma escolha, dado que podiam ser mais ou menos aptos a empregarem sua racionalidade para decidir. Mas a escolha correta existia e era só uma questão de encontrar um método robusto aos erros dos menos aptos racionalmente. Por isso, a agregação das ordenações de alternativas dos indivíduos era para ele, no fundo, um problema probabilístico de eliminação do erro e maximização da racionalidade coletiva.

Borda, nesse sentido, propunha uma alternativa que subordinava o imperativo racional à escolha que parecia representar o melhor compromisso entre diferentes juízos. Condorcet nunca aceitou tal solução, porque para ele não importavam os valores, e sim o procedimento com a maior probabilidade de representar a escolha racional correta. As

soluções dadas por Condorcet congelam o componente de juízo presentes nas avaliações e isso traz uma série de implicações teóricas e práticas. Primeiro, faz-nos supor que a avaliação social teria a mesma estrutura que a avaliação individual e que se considerarmos os indivíduos racionais, ou mesmo imperfeitamente racionais, a decisão de cada um conteria parte da verdade e a soma de todas as decisões conteria, de algum modo, a verdade completa e o ótimo social. Congelado o juízo, também decorre que todos os indivíduos identificam e interpretam os objetos postos a escolha do mesmo modo. Para todos os autores que lidam com escolha social, aliás, esse é um ponto que chama atenção. A capacidade dos indivíduos identificarem as alternativas *a* e *b* nunca é posta em dúvida, há muita discussão sobre as estruturas de preferência²³ e sobre a capacidade dos critérios utilizados para distinguir entre *a* e *b*, mas nunca sobre a pertinência de *a* ou *b* para aqueles que avaliam.

Arrow também não aceita a solução de Borda, baseada na comparação quantitativa das preferências (identificadas como comparações entre *utilidades*), embora tenha consciência que o problema não é probabilístico. Entretanto, mesmo sem considerar as estruturas de preferência dos indivíduos acessíveis, ele as considerou igualmente racionais e objetivas. Além disso, postulou que uma suposta estrutura de preferência social também deveria ser igualmente racional. A delimitação de ambas estruturas de preferência a uma racionalidade padrão as torna estruturalmente isomórficas e cria um ideal de referência que, conforme ele mesmo demonstrou, não se sustenta.

Arrow não deixa espaço para se entender a avaliação coletiva e a avaliação que ocorre no indivíduo como fenômenos essencialmente distintos. Em contraste, essa distinção é essencial no esquema analítico proposto nas seções anteriores, sendo a avaliação coletiva apenas capaz de influenciar, por meio de métodos que emergem codificados das experiências e das relações entre muitos indivíduos, a diversidade de possibilidades de mecanismos válidos para justificar as decisões que ocorrem em cada indivíduo.

VIII. AVALIAÇÃO E INSTITUIÇÕES

Conforme foi discutido na seção VI, uma avaliação pode ser entendida como um

²³ Cf. Seções XIII e XIV, no Capítulo 2.

processo de intermediação entre indivíduos e sociedade. A impossibilidade de se construir um método racional capaz de objetivar essa intermediação não significa que não existam nexos entre as decisões dos indivíduos e a existência de forças sociais que as tornem em algum sentido previsíveis. Arrow propositadamente considera o que ele chama de *convenções* como irrelevantes ao problema da escolha social. Tais convenções poderiam ser interpretadas como pressões sociais tão poderosas a ponto de serem capazes de reduzir o universo de opções dos indivíduos a uma única decisão. Tal decisão seria a única socialmente válida ou legítima e emergiria naturalmente como o resultado agregado das avaliações de todos aqueles pertencentes à tal sociedade.

Entretanto, nem todas as convenções ou instituições são uniformemente aceitas nem imunes a controvérsias. Há, portanto, situações intermediárias entre os tipos considerados por Arrow. Uma avaliação de um indivíduo pode estar – e em geral está – situada em outras posições no contínuo entre a liberdade plena e a determinação total. O ponto importante a destacar é que a influência social, em maior ou menor grau, sempre estará presente nas avaliações realizadas pelos indivíduos, seja por motivação própria para legitimar as decisões perante outros indivíduos seja por decorrência do método utilizado. Nesse último caso, sendo as formas e métodos das avaliações construções sociais, isso equivale a dizer que uma avaliação, uma vez formalizada e amplamente aceita, torna-se uma espécie de instituição social, capaz de promover ordem e previsibilidade.

Seguindo a definição bastante abrangente dada por Scott (2001, p. 48), “instituições consistem em elementos culturais-cognitivos, normativos e regulatórios com alto grau de resiliência que juntos proporcionam estabilidade e significado ao comportamento social. As instituições são transmitidas por vários tipos de portadores – incluindo sistemas simbólicos, sistemas relacionais, rotinas e artefatos – que operam em níveis múltiplos de jurisdição, desde o nível mundial até as relações interpessoais localizadas. As instituições, por definição, conotam estabilidade, mas ao mesmo tempo são sujeitas de processos de mudança, tanto incrementais quanto de descontinuidade.”

Vale a pena explorar um pouco mais essa definição. Nela estão presentes conceitos que precisam ser mais bem esclarecidos. Primeiro, os três grupos de elementos constituintes das instituições, elementos culturais-cognitivos, normativos e regulatórios. Essa divisão corresponde a três pilares estruturais que dão sustentação a uma instituição. Para se

compreender uma instituição são necessários os três. De modo sucessivo, são como esferas de significação para uma instituição, do geral ao específico e do abstrato ao concreto. Os elementos culturais e cognitivos se referem aos modos de apreensão da realidade e à racionalidade subjacente à tradução simbólica dessa realidade. Os elementos normativos, por sua vez, correspondem aos valores e normas sociais construídos a partir do esquema anterior. Finalmente, os elementos regulatórios dizem respeito aos mecanismos práticos para constrangimento, controle ou indução comportamental. Numa dada instituição, esses três tipos de elementos são expressos simultaneamente. A propriedade de resiliência de uma instituição é a expressão da conexão intrínseca que esses três pilares apresentam e sua estabilidade será função da consistência que essa conexão demonstrar. Instituições bem constituídas – no sentido de sua consistência interna – tornam-se "naturais" e muitas vezes apresentam grande longevidade em um dado sistema social, conferindo racionalidade e previsibilidade comportamental. Certas regras e mecanismos de controles se tornam tácitos nas relações sociais e influenciam as percepções, especialmente quando ancoradas no pilar cultural-cognitivo ou no normativo. O âmbito regulatório geralmente é mais bem codificado e, portanto, mais facilmente discutível.

Ainda segundo Scott, os portadores das instituições exprimem os mecanismos pelos quais as instituições se configuram e são socialmente transmitidas, cobrindo um espectro que vai desde aqueles mais abstratos até os mais concretos. Incluem os sistemas simbólicos, os sistemas relacionais, as rotinas e até mesmo certos artefatos. Os sistemas simbólicos são os portadores mais abstratos, incluem por exemplo, representações, crenças, modelos, classificações e lógicas. Sistemas simbólicos estão em toda parte, são não só delineadores do contexto cultural de uma avaliação, mas responsáveis também pelos esquemas mentais e conceituais dos indivíduos envolvidos. Sistemas relacionais dizem respeito ao comportamento de atores sociais, os constrangimentos e benefícios derivados das relações de poder, indução de formas similares em diferentes organizações (isomorfismos estruturais) e estruturas de governança. Enquanto portadores, sistemas relacionais contribuem para a criação de ligaduras que, a partir de processos de interação entre instituições, geram novas instituições e enfraquecem ou reforçam antigas. Rotinas são portadores que transmitem hábitos e procedimentos tácitos ou em vários graus de codificação. Rotinas podem ser identificadas com a propriedade de inércia que

organizações muitas vezes apresentam; enquanto sistemas relacionais podem forçar mudanças institucionais, sistemas simbólicos e rotinas geralmente favorecem a estabilidade institucional. Finalmente, os artefatos são as formas mais concretas de transmissão de instituições. Nem sempre é evidente que artefatos tecnológicos carregam em si elementos culturais-cognitivos, normativos ou regulatórios, entretanto, características como beleza, desenho e qualidade são perfeitamente enquadráveis nesses elementos. O conceito marxiano de fetiche da mercadoria é um exemplo típico de como artefatos podem transmitir e legitimar instituições. O quadro 1.1 abaixo traz diversos exemplos de tipos de portadores referentes aos três pilares institucionais. É possível notar um gradiente da esquerda para direita e de cima para baixo saindo dos portadores de instituições mais abstratos e indo para aqueles mais concretos.

Quadro 1.1: Exemplos de portadores de instituições nos diferentes pilares institucionais (SCOTT, 2001).

<u>Portadores</u>	<u>Pilares das Instituições</u>		
	<i>Cultural-Cognitivo</i>	<i>Normativo</i>	<i>Regulatório</i>
<i>Sistemas Simbólicos</i>	categorias, tipificações, esquemas	valores, expectativas	regras, leis
<i>Sistemas Relacionais</i>	isomorfismos estruturais, identidades	regimes, sistemas de autoridade	sistemas de governança, sistemas de poder
<i>Rotinas</i>	padrões comportamentais	empregos, papéis, obediência moral	protocolos, procedimentos operacionais padronizados
<i>Artefatos</i>	objetos com valor simbólico	objetos dentro de convenções e padrões	objetos dentro de especificações determinadas

Há claras conexões entre avaliações e instituições. Esse tema pode ser explorado por diversos ângulos. Primeiro, há a constatação quase óbvia de que instituições condicionam, ou limitam, em alguma extensão os juízos dos indivíduos. Nesse ponto, já em 1899 Veblen afirmava que as decisões dos indivíduos sobre assuntos econômicos consideravam diversos outros valores que apenas a racionalidade econômica objetiva.²⁴

²⁴ Theodore Veblen, figura destacada da chamada Escola Econômica Institucionalista, em seu livro *A Teoria da Classe Ociosa* de 1899 argumentava que as decisões das elites econômicas eram fundamentadas muito mais pelo desejo de *consumo conspicuo* e de impressionar seus pares mostrando-lhes a posse de muitos excedentes

Um segundo modo de perceber a ligação entre avaliação e instituições é pelo mecanismo de aprendizado. O objetivo dos primeiros institucionalistas no início do século XX ao enfatizarem a importância das instituições era justamente contrapor os esforços teóricos de cientistas sociais e economistas que buscavam teorias científicas *stricto sensu* para explicar o comportamento dos indivíduos em interação, isto é, teorias capazes de medir e prever o comportamento do homem em sociedade. Os institucionalistas argumentavam que tais esforços seriam demasiado reducionistas e simplificadores da realidade e que a sociedade não podia ser estudada em laboratório como um objeto físico. Também Max Weber, nessa época, tinha clareza que em se tratando de objetos sociais, "tanto o pesquisador quanto o objeto em estudo atribuem significados aos eventos" (SCOTT, 2001). Ou seja, o objeto aprende. Assim, deixar de fora o contexto dado pelas diversas instituições que afetam tanto pesquisador como objeto de estudo era não apenas temerário como metodologicamente inaceitável.

A partir desses alertas, ou apesar deles, muitas abordagens metodológicas para avaliações que lidavam com objetos sociais interpretaram o fator aprendizado como um problema. Desse modo, em muitos casos, produziram-se mecanismos para evitar que o aprendizado ocorresse e permitir que os métodos pudessem ser validados à moda dos experimentos científicos típicos das disciplinas físicas e biológicas. Apenas mais recentemente, alguns métodos aplicados a avaliações coletivas passaram a aproveitar a oportunidade da avaliação para explicitamente promover o aprendizado. Essa é linha metodológica efetivamente adotada pelo método proposto e discutido no Capítulo 4.

Pelo que foi definido na Seção IV, o aprendizado é um efeito individual, resultado da concatenação entre avaliações e decisões no tempo. Entretanto, se subirmos um nível e tomarmos a concatenação de avaliações coletivas teremos algo como um aprendizado coletivo. Não é absurdo identificar esse aprendizado coletivo com motivação para a criação, manutenção ou modificação de certas instituições.

Além disso, uma vez estabelecida como rotina, uma avaliação pode se tornar um processo institucionalizado, com regras, códigos de conduta e significado social. Porém, nem toda avaliação é institucionalizada como prática social e essa não pode ser uma característica definidora de avaliação.

do que por cálculos racionais de maximização orientados ao crescimento da produção. Cf. VEBLEN, T. (1983, p. 35-48).

IX. ATRIBUTOS PARA CARACTERIZAR AVALIAÇÕES

No decorrer deste capítulo, diversos elementos descritivos úteis para analisar processos de avaliação e decisão foram apresentados e discutidos. Alguns surgem naturalmente a partir das definições elaboradas, outros são conseqüências mais ou menos diretas. Esta última seção realiza um apanhado dos elementos que podem ser úteis como atributos para caracterizar uma avaliação. Tais atributos prestam-se ao estudo do objeto *avaliação* e obviamente não são exaustivos nem contêm embutidos todos os juízos que um avaliador precisa empregar para interpretá-los.

Dada uma avaliação, isto é, uma situação envolvendo a interpretação de atributos por critérios, a primeira distinção a se fazer é se se trata de avaliação individual ou coletiva. Apenas a avaliação individual leva diretamente à ação, simbolizada pela decisão – e ocorre independentemente do indivíduo ter consciência do processo de avaliação, ou em outras palavras, do método utilizado.

A avaliação coletiva não é a avaliação do indivíduo e não implica na decisão, apesar de poder, evidentemente, influenciar a avaliação individual. Uma avaliação coletiva necessariamente possui método explícito e envolve a participação de diversos indivíduos.

Em ambas, os objetos avaliados podem ser concretos ou abstratos, uma vez que apenas seus atributos são considerados para a aplicação dos critérios. São os contrastes entre os possíveis estados dos atributos que leva a diferentes juízos. Atributos são sempre abstratos e, portanto, construções subjetivas individuais ou sociais (intersubjetivas).

Em qualquer avaliação ocorre aprendizado. O aprendizado pode ser um subproduto explícito ou não, desejado ou não, planejado ou não, mas sempre presente. Quem aprende são os indivíduos e as organizações. Do ponto de vista coletivo, seja em organizações, seja socialmente, o aprendizado se identifica com a criação ou modificação de instituições.

Do ponto de vista dos métodos utilizados, isto é, dos mecanismos de avaliação socialmente construídos que levam a ações socialmente previsíveis e legítimas, é possível estabelecer ainda alguns atributos para comparar avaliações.

Na avaliação individual, apesar de existir casos nos quais aparentemente se prescinda de método, é mais verossímil considerar que métodos sempre estão presentes, de modo inconsciente ou consciente. O fato de um método não estar formalizado ou do

indivíduo não ser capaz de codificá-lo não significa que ele não exista. Ele pode ser inconscientemente utilizado, enquanto construção social embutida pela cultura, linguagem ou valores éticos.²⁵ Método de avaliação é todo mecanismo utilizado para selecionar atributos de objetos, selecionar o tipo de juízo de valor e aplicar esses juízos aos objetos por meio das percepções ou medidas obtidas.

É evidente que pode haver graus diferenciados de precisão e rigor formal nos métodos. Isso é também atributo para classificar as avaliações segundo seus métodos. O próximo capítulo explora em mais detalhes algumas possibilidades de variação formal em avaliações e suas consequências.

Numa avaliação coletiva, ampliam-se as possibilidades de variações de método e de sistematização processual. A participação de diversos indivíduos pode ser dar em diversos momentos e com diferentes ênfases. Em geral, indivíduos são empregados como instrumento de medida e, menos freqüentemente, como fornecedores de juízos. É metodologicamente difícil incorporar juízos dos participantes sem incorrer numa situação de impossibilidade (de escolha social). Por isso, considerar juízos dos indivíduos não pode ser visto como matéria-prima para um “juízo coletivo objetivo” porque isso implicaria numa decisão coletiva. Numa avaliação coletiva, os juízos individuais são também objetos. Os métodos para avaliações coletivas são, portanto, necessariamente mais complicados. Eles precisam mediar diferentes níveis de avaliação e “absorver” diferentes métodos individuais para operar. Em última análise, uma avaliação coletiva consiste em coordenar o aprendizado coletivo para, a partir deste, induzir decisões nos indivíduos.

Desse modo, nem toda avaliação participativa é uma avaliação coletiva, em seu sentido pleno. Se a participação consistir apenas em fornecer medidas ou tomar conhecimento de um processo de avaliação, é provável que não se atinja o grau de construção social compartilhado que uma avaliação coletiva pode proporcionar. Uma avaliação participativa, em sua acepção restrita, é ainda muito próxima a uma avaliação

²⁵ A respeito disso — e corroborando outros elementos assumidos pelo marco analítico proposto — é emblemático o comentário de Weber no ensaio *A Objetividade do Conhecimento*, de 1904. Afirma ele que “levar até uma decisão não constitui uma tarefa possível para a ciência [social], mas antes para a pessoa dotada de vontade. Esta confronta e escolhe os valores em causa segundo a sua própria consciência e a sua própria concepção do mundo. Por certo que a ciência pode ajudá-la a perceber que qualquer ato e também, segundo as circunstâncias, a ausência de um ato signifiquem, pelas suas consequências, tomar o partido de determinados valores. E, simultaneamente, significa, fato expressamente e com freqüência ignorado na atualidade, tomar partido contra outros valores” (WEBER, 1991, p. 6).

individual em termos de método. Mais adiante, essa distinção é trabalhada com maior detalhamento.

Para encerrar este capítulo, cabe um breve comentário sobre atributos de pouca relevância para classificar avaliações. Na literatura sobre o tema avaliação, existe uma profusão de atributos que não foram diretamente considerados no esquema analítico apresentado. Pode-se citar as distinções entre métodos qualitativos e quantitativos, abordagens *ex ante* e *ex post*, avaliações rápidas (*rapid appraisal*) ou em profundidade, reducionistas ou holísticas, e assim por diante. Todas essas distinções são desnecessárias para a análise da avaliação como fenômeno social. Elas ilustram apenas variações de método ou de orientação temporal que, se utilizadas como ponto de partida, pouco colaboram para captar o sentido social e a motivação individual, subjacente às avaliações e decisões, que efetivamente são importantes, inclusive para compreender os métodos e pensar alternativas de procedimento. A caracterização de uma avaliação e a discussão sobre métodos de avaliação não deve partir desses atributos, com o risco de não se atingir o cerne da questão.

Este capítulo apresentou e procurou justificar a pertinência e a utilidade de um esquema analítico para entender e caracterizar avaliações. Ele foi construído tendo em mente uma série de problemas recorrentes dos métodos de avaliação, especialmente daqueles com pretensões coletivas e de participação. Diversas construções teóricas foram consideradas e procurou-se ajustar as definições de modo que todas pudessem ser acomodadas.

O próximo capítulo mergulha na discussão sobre as características e os limites dos métodos de avaliação, especialmente daqueles que utilizam como estratégia de legitimação social a objetivação de seus resultados a partir da tentativa de anular a interferências subjetiva dos indivíduos nos julgamentos de valor. Essa classe de métodos se destaca por englobar numerosas formalizações e modelos de análise correntemente empregados em avaliações. Ela merece ser visitada para que se compreenda em maior profundidade as raízes da racionalização objetiva nela presentes como premissas. Só assim os efeitos sobre os métodos de avaliação quando da flexibilização dessas premissas poderão ser rigorosamente descritos.

Avaliação e Seus Métodos

Capítulo 2 – Limites das formas objetivadas e as bases para participação e aprendizado

So you want facts only, no values, because they are illegitimate.

But you can't have facts without values. Touché!

Amartya Sen

No Capítulo 1, formulou-se um esquema analítico geral para interpretar e estudar avaliações. Tal esquema foi construído a partir da sugestão de três princípios, cada qual introduzindo uma definição de partida útil para apreender o significado social do que é avaliação. Foi assumido que: 1) avaliar é interpretar um atributo por meio de um critério; 2) decidir é executar uma ação a partir de uma avaliação; e que 3) todos os métodos de avaliação são construções sociais. O Capítulo 1 discutiu ainda as implicações dessas definições e diversos conceitos derivados foram apresentados, incluindo aprendizado, avaliação coletiva e diferentes formas de decisão. Além disso, a pertinência do marco analítico foi testada, com a concomitante revisão de alguns dos mais importantes tópicos relacionados à teoria e prática de avaliações, tais como o conceito de causalidade, o processo racional de tomada de decisão, a proposição de métodos para a escolha social e o papel das instituições.

O marco analítico construído é propositadamente genérico, tanto para dar conta da diversidade de possibilidades metodológicas e de aplicação abertas às avaliações quanto para fornecer os elementos necessários para a crítica de determinadas premissas metodológicas bastante disseminadas e arraigadas em práticas mais institucionalizadas de avaliação. É o caso da restrição dos métodos de avaliação apenas àqueles baseados em critérios objetivados de validade. Essa restrição, argumenta-se, é também uma construção social e absolutamente desnecessária para a construção de métodos de avaliação válidos, bastando para isso haver situações socialmente pertinentes que demandem outro tipo de legitimação que não exclusivamente a fornecida por essa estratégia.

Na extensão em que métodos de avaliação são construções sociais, o esforço dos indivíduos em codificá-los e explicitá-los é, em última análise, um esforço para legitimar

suas decisões socialmente. A metodologia escolhida para uma avaliação constitui, nesse sentido, um recurso de legitimação social dessa avaliação.

Este capítulo examina aquela que é por suposto uma das mais importantes estruturas de legitimação para as avaliações: o enquadramento das avaliações na classe dos problemas científicos e, portanto, o uso do método científico para avaliar e do critério de validade científica para legitimar decisões.

Dentro da diretriz de enquadramento das avaliações como problema científico há ainda, evidentemente, um enorme espaço para variações metodológicas e matizes conceituais, de modo que o tratamento do assunto requer um grau de abstração relativamente alto. As implicações formais do enquadramento científico são de especial interesse para desenvolver o argumento central do capítulo, qual seja, a defesa de que alterações formais de base são necessárias para a proposição de métodos de avaliação melhor adequados a contextos específicos de aplicação, especialmente àqueles que envolvem objetos sociais e aos quais as abordagens objetivadas mostram-se insuficientes.

O que interessa a este capítulo é a crítica da racionalidade objetiva usual que subsiste por trás de muitos métodos de avaliação. Algumas das conclusões adiantadas neste parágrafo não são evidentes, porém ilustram concisamente o que resulta dessa crítica. Nos interessa examinar como os métodos de avaliação podem evoluir para novas relações entre seus elementos objetivos e subjetivos, avançando além das premissas de máxima objetivação usuais. Entretanto, não faz sentido procurar essa evolução apenas pelo caminho de oposição à formalização nos métodos e de negação sumária à quantificação, partindo para a defesa de métodos ditos “qualitativos”. Porque abandonar a quantificação de atributos em favor da qualificação não significa abandonar os critérios e a lógica objetiva usual. Muitos métodos qualitativos são ainda objetivados quando, mesmo sem números, lêem os atributos como critérios e, para produzir conclusões, obedecem aos mecanismos lógicos usuais. Por outro lado, se ser qualitativo, ou flexibilizar a objetivação, for entendido como a ausência de método, qualquer possibilidade de racionalidade e de legitimação social se esvanece. Como decorrência do que foi argumentado no capítulo anterior, esconder o método, mistificar ou personalizar o processo tornando medidas e juízos totalmente idiossincráticas impede a legitimação do resultado de uma avaliação, e desta como resultado de um mecanismo racional socialmente aceito. Por tudo isso, negar a construção

formal de métodos não é uma boa direção da investigação e é evitada neste capítulo. É necessário continuar buscando apoio na racionalidade, na formalização e no rigor matemático (mas não necessariamente em números) como meios de socializar intersubjetivamente os procedimentos de uma avaliação. A novidade, por seu turno, são as possibilidades abertas ao se trabalhar não apenas em uma matemática, fundamentada em apenas uma lógica, mas também em outras, construídas sobre outras premissas lógicas, tão leves ou particulares quanto se deseje. É preciso considerar métodos para situações que não requeiram capacidade de previsão e repetição e que a racionalidade exigida possa ser menos rígida e uniforme, caso típico das avaliações coletivas de objetos sociais, por exemplo. Como e porque fazer isso, o que se ganha e o que se perde com isso são os assuntos deste capítulo.

O que foi resumido no parágrafo acima necessita de diversos esclarecimentos pois há certas assertivas aceitas tacitamente pelo senso comum acerca de formalização, quantificação, abordagens qualitativas e mesmo sobre o significado de objetividade e subjetividade em métodos de avaliação que precisam ser cuidadosamente demovidas para delimitar o escopo da discussão metodológica que está em pauta. O Capítulo 2, concebido para essa tarefa, consiste em seis seções, em continuação às do capítulo anterior. As seções X e XI tratam do que aqui se convencionou chamar de *avaliação objetivada*. Primeiro são apresentados os meios pelos quais certos elementos do método científico podem ser transpostos para as avaliações, a fim de minimizar o papel do sujeito. Em seguida, discute-se, a partir dos problemas apontados pelos próprios defensores desse tipo de abordagem, os limites metodológicos mais evidentes. A seção XII retoma as definições do Capítulo 1 para delimitar com clareza quais são então os caminhos a serem trilhados para escapar dessas armadilhas metodológicas em avaliação. As seções XIII e XIV são dedicadas a mostrar, a partir do exemplo do estudo das preferências e dos métodos de apoio à decisão multicritério, que existem estruturas formais e métodos que, ao violarem o padrão científico usual, viabilizam métodos de avaliação com outras propriedades. Finalmente, a seção XV retoma e organiza os principais argumentos deste capítulo a favor de diretrizes metodológicas que incorporem mais naturalmente a participação e o aprendizado nas avaliações e procura recolocar o debate sobre cientificidade e racionalidade em avaliações.

X. AVALIAÇÃO OBJETIVADA

Adota-se a denominação *avaliação objetivada* para designar toda avaliação que pressupõe o uso do método científico e aplica o critério lógico de validade sobre medidas objetivadas de atributos. Essa estratégia resulta em diversas vantagens para a legitimação de uma avaliação. O cânone científico é identificado por muitos como o único meio de se obter avaliações corretas, objetivas, livres de opiniões de veracidade duvidosa e, portanto, legítimas. A crítica a esse ponto de vista não é trivial, porque em sua essência ele não pode ser considerado errado. O método científico tem justamente o mérito de objetivar os problemas e validar soluções teóricas que se tornam assim amplamente aceitas como corretas. O uso do método científico é inegavelmente uma ótima opção para avaliações, para todas aquelas a que ele pode ser aplicado. A questão é até que ponto a abordagem científica de objetivação pode ser generalizada para todas as avaliações.

Como foi visto, uma vez que elementos subjetivos são parte integrante de qualquer avaliação, fica claro que reside justamente aí o ponto de tensão que cria dificuldade para as tentativas de objetivação. A discussão metodológica usual gira em torno desse fato, os métodos de avaliação objetivados têm histórica dificuldade em lidar com o subjetivo, ou melhor, de evitar o subjetivo. Fazendo alusão à discussão sobre paradigmas na ciência (KUHN, 2000), essa situação, e a maior parte dos debates por ela incentivados, é um típico exemplo de anomalia discutida dentro dos esquemas conceituais e das crenças metafísicas aceitos pelo cânone em vigor (chamado por Kuhn de *ciência normal*). Todos percebem os elementos subjetivos presentes numa avaliação como um problema a ser contornado, como uma falha, uma fraqueza dos métodos. O subjetivo não pode ser aceito como intrínseco nas avaliações científicas porque ele impede sua legitimação. Entretanto, a crítica insiste em afirmar que os métodos de avaliação precisariam melhorar sua capacidade de domar o subjetivo. Ora, esse tipo de crítica reforça ainda mais a distorção que está na própria base do paradigma da avaliação objetivada: a sobreposição de atributos e critérios e a crença metafísica que identifica qualquer elemento subjetivo como essencialmente imponderável, portanto, ilegítimo numa avaliação.

Os métodos construídos a partir dessas premissa apresentam todos eles o traço comum de procurar eliminar ou reduzir a um mínimo controlável o papel dos juízos

estruturados em critérios independentes dos atributos e de outros elementos subjetivos numa avaliação. Do ponto de vista da objetivação científica, esse é o caminho para se obter correção e legitimidade nos resultados. Do ponto de vista do esquema conceitual que definimos para avaliações, isso é perfeitamente aceitável, mas representa apenas um caso particular que pressupõe grande homogeneidade entre os indivíduos em relação a seus valores e características cognitivas, de modo que possam identificar os fatos e os atributos relevantes do mesmo modo, e em relação à aceitação da estrutura de juízo objetivada na avaliação. Quando essas condições são encontradas, a objetivação científica pode com sucesso ser adotada para se fazer avaliação. Novamente, há alguma coincidência entre essa situação e a noção kuhniana de paradigma. Um paradigma, “de um lado, indica toda constelação de crenças, valores, técnicas, etc. (...), partilhadas pelos membros de uma comunidade determinada. De outro, denota um tipo de elemento dessa constelação: as soluções concretas de quebra-cabeças que, empregadas como modelos ou exemplos, podem substituir regras explícitas como base para a solução de problemas da ciência normal” (KUHN, 2000, p. 218). A primeira parte da definição de Kuhn consiste em certas condições institucionais necessárias para legitimar um conjunto de interpretações e regras como socialmente válidas, como é o caso das premissas de uma avaliação objetivada. A segunda ilustra o poder de influência que aplicações exemplares de certas abordagens ou teorias têm sobre as soluções propostas para os novos problemas que surgem, o que também pode ser encontrado no caso da avaliação. Certos métodos de avaliação, especialmente de avaliação objetivada, ao se tornarem casos exemplares passaram a constituir modelos para toda e qualquer avaliação e, apesar de serem problemáticos em determinadas situações, continuaram legítimos, inclusive para as aplicações problemáticas.

A crítica construída a partir desta seção não trata, entenda-se bem este ponto, de negar o emprego do método científico e tampouco da relevância das avaliações objetivadas. Já se adiantou que elas continuam a ser uma excelente opção para todos os casos em que fizerem sentido, ou seja, quando as condições institucionais ou a característica do objeto avaliado assim permitirem. O problema reside nas extensões indevidas, realizadas a objetos não compatíveis com suas aplicações exemplares, no sentido apontado pela segunda parte da definição de Kuhn. Para esses casos, a crítica precisa mostrar como evitar a lógica objetiva numa avaliação e conceber outros casos exemplares que aceitem a idiossincrasia

do subjetivo como algo natural, e mais importante, como algo a ser internalizado e não eliminado em um método de avaliação. Afinal, há casos de avaliação em que a inadequação de um método e dos resultados não pode ser atribuída à aplicação incorreta e nem mesmo ao método em si. A inadequação vem desde as premissas adotadas, e essas se convertem em estruturas formais inadequadas à situação alvo de avaliação.

Porém, antes de avançar nesse sentido, é importante examinar como a subjetividade é cerceada do ponto de vista metodológico pela avaliação que se pretende cientificamente objetiva. Pelo menos quatro estratégias complementares se destacam.

Conforme a famosa recomendação de Galileu, segundo quem "tentar resolver uma questão científica sem o auxílio da matemática é tentar o impossível. Temos que medir o que é mensurável e tornar mensurável o que não pode ser mensurado" (KLEIN, 1974), a primeira estratégia é restringir o alcance dos métodos de avaliação àquilo que se pode medir. A medida, nesses casos, recai sobre atributos selecionados e com propriedades intersubjetivamente conhecidas. Em geral, adota-se escalas que já contenham juízos embutidos, ou a quais é fácil tratar como critério, como no caso dos valores monetários. Medir os objetos do ponto de vista de seu atributo unidades monetárias consiste numa forma socialmente objetivada de avaliação, na qual o juízo de valor está naturalmente implícito e a distinção entre atributo e critério se perde.

Uma segunda característica da aplicação do método científico em avaliações é o uso de deduções. Toda dedução corretamente executada, no sentido lógico, preserva o valor de verdade das premissas utilizadas. Ao utilizar premissas aceitas pelo senso comum, ou obtidas por indução empírica de fatos evidentes a todos, um método de avaliação pode chegar a conclusões que devem ser, por legitimidade lógica, também aceitas. Tal "raciocínio científico" em avaliações nem sempre pode ser corroborado por fatos, como ocorre nas ciências naturais, uma vez que os objetos a que se referem podem ser objetos sociais únicos e não reprodutíveis, por exemplo por existirem somente no passado ou no futuro da experiência sensível dos indivíduos. Desta feita, certas avaliações se aproximam rapidamente da retórica. Apesar de logicamente estruturadas, não são nem podem ser cientificamente testadas. Sua legitimidade provém, portanto, da forma e não de seu conteúdo empírico. Problemas lógicos que decorrem desse tipo de abordagem, como o problema da atribuição causal que superestima causas passadas para explicar os efeitos

presentes ou subestima os efeitos futuros a partir de causas presentes, são inerentes do emprego de esquemas dedutivos rígidos em avaliações que contenham objetos sociais.

Uma terceira característica do tratamento científico em avaliações é assumir o pressuposto de que para todo problema passível de avaliação existe uma, e apenas uma, resposta correta. Assim, a maior ou menor dificuldade para se chegar a essa resposta pode ser explicada pela incapacidade de identificar e medir corretamente todos os fatores necessários (atributos) e suas conseqüências (atribuições causais). A saída convencional para esses limites é admitir que associado a toda medida e toda avaliação há um erro, sistemático ou aleatório. O erro sistemático pode, em tese, ser controlado, pois se origina de um defeito constante de mensuração ou da presença de uma variável não conhecida de antemão e que precisaria ser introduzida no esquema conceitual que serve como modelo para a avaliação. Já o erro aleatório só pode ser controlado com o auxílio de modelos probabilísticos, a partir dos quais se pode inferir aproximações da verdade oculta. Torna-se natural, nesse sentido, interpretar a variação dos juízos dos indivíduos como erro aleatório, como pressupunha, por exemplo, Condorcet.

Por fim, como quarta estratégia, a racionalidade científica adota a segunda parte do conselho de Galileu e constrói novas medidas, para atributos que num primeiro momento não eram mensuráveis. Essa estratégia enquadra uma série de esforços teóricos e metodológicos e, dentre as quatro, é a que melhor ilustra o mecanismo de resposta científico para as críticas de inadequação de certos métodos de avaliação. Ela se diferencia da primeira estratégia porque requer a introdução de novos conceitos. Com eles, passa-se a considerar, e modelar sobre métricas manipuláveis matematicamente, aqueles atributos que não eram até então reconhecidos como relevantes ou necessários a uma avaliação. Com isso, o arsenal conceitual e metodológico se expande. Em avaliação, muito da variação de métodos e aplicações existentes decorre desse mecanismo, chegando a compor, segundo a interpretação de alguns autores, diferentes paradigmas de avaliação, com escolas que apresentam grandes variações metodológicas e até mesmo conflitos explícitos entre si.²⁶ Ao não extrapolar o paradigma mais amplo da objetivação das avaliações, a maior parte desses esforços contribui incrementalmente para seu fortalecimento e legitimação, mantendo a discussão restrita a quais atributos são ou não relevantes e de que modo encaixar neles

²⁶ Cf. Stern (1993) para um comentário particularizado às avaliações de ciência e tecnologia. Worthen *et al.* (2004) também ilustram esse fato para os métodos de avaliação orientados a programas sociais.

medidas e modelos de causa e efeito. Exemplos clássicos de novos atributos cujas medidas foram introduzidas em diversos métodos de avaliação incluem as diversas variações de medidas para utilidade, a teoria dos jogos,²⁷ a psicologia cognitiva e a psicometria.²⁸ Em todos esses exemplos, procura-se abstrair e objetivar o comportamento individual, produzindo modelar critérios para determinar e justificar avaliações que, pelo menos nos contextos de suas aplicações metodológicas, são considerados como medidas objetivas isentas de "subjetivismo". Dizer que uma avaliação é "subjetiva", nesse contexto, se torna uma qualificação que ganha conotação pejorativa.

As estratégias de objetivação científica, que são características das avaliações objetivadas, aparecem em geral combinadas e se auto reforçam. A título de ilustração, segue a apresentação de elementos metodológicos gerais presentes em toda uma família de métodos de avaliação de programas sociais, segundo a terminologia utilizada por Cook e colaboradores (COOK et al., 1977).

Essa apresentação interessa particularmente aos objetivos deste capítulo, uma vez que representa um importante esquema metodológico padrão. Como já foi anunciado acima, interessa examinar situações em que ocorre inadequação desse esquema. Esta seção se limita a explicar os elementos da avaliação objetivada de objetos sociais sem, entretanto, discuti-los criticamente. Isso será feito na próxima seção.

Para o esquema geral de avaliação apresentado a partir de agora, a denominação programa social é atribuída a um amplo espectro de intervenções, necessariamente planejadas, que de algum modo impactam os indivíduos de uma determinada comunidade. Para se fazer a avaliação de um programa, três condições são pré-requisitos e precisam ser testadas previamente quando a sua existência: 1) um programa delimitado claramente; 2)

²⁷ Cf. Seções XIII e XIV, respectivamente, para as descrições sucintas das medidas de utilidade e utilidade esperada (base da teoria dos jogos).

²⁸ A psicometria ganhou impulso após o artigo de Likert publicado em 1932 (*A technique for the measurement of attitudes*). Rensis Likert, especialista em estatística e psicologia, postulou certos princípios construtivos para que uma escala semântica pudesse ser convertida em um contínuo numérico. Assim, seria possível a medida direta de atributos psicológicos e empregar pessoas como instrumentos de medida. No Capítulo 4 essa técnica é descrita e exemplificada. Um exemplo importante da psicologia cognitiva está no trabalho de Tversky e Kahneman de 1974 (*Judgement under uncertainty: Heuristics and biases*). Os autores demonstram a presença de *framings* (ancoragens) comportamentais que levam a avaliações distintas para os mesmos atributos em função dos riscos percebidos no contexto dado para a avaliação. Propõem ainda que os indivíduos empregam modelos heurísticos que seguem certos padrões e, portanto, poderiam ser determinados experimentalmente.

objetivos e efeitos claramente especificados; 3) um esquema racional de causalidades ligando programa a objetivos e efeitos.

Segundo esse ponto de vista, a avaliação ideal seria equivalente a um experimento científico. "A avaliação é, antes de tudo, um processo que aplica procedimentos científicos para acumular evidências confiáveis e válidas sobre como e em que medida determinadas atividades produzem determinados efeitos ou resultados" (RUTMAN, 1977, p.16). Desse modo, a intervenção de algum evento (o programa ou componente dele) é chamada *tratamento*. A avaliação em si consiste em averiguar os efeitos do tratamento por meio de acompanhamento sistemático de grupos de controle e das medidas das unidades de estudo (pessoas, outros atores, variáveis). A "avaliação compara um programa ou um conjunto de programas, assim como foram definidos pelos administradores ou responsáveis por políticas, com a realidade: as atividades e os efeitos que realmente ocorreram ou ocorrem" (WHOLEY, 1977, p. 41). As conclusões são obtidas por meio de inferências estatísticas da probabilidade das relações de causa e efeito presumidas.

Para que os experimentos sejam efetivos é necessário que as unidades experimentais sejam submetidas aleatoriamente aos tratamentos. Quando há impedimentos práticos para isso, são assumidos os *quase-experimentos*, nos quais a alocação dos tratamentos ocorre de modo não aleatório, porém ainda sob rígido controle, de modo que algumas inferências causais ainda sejam possíveis.

Uma inferência causal depende de mostrar que uma causa é relacionada com um efeito presumido, que a variação na causa precede a variação no efeito e que nenhuma outra alternativa de relação é possível além da medida obtida em decorrência da causa modificada. Dadas essas condições, ainda assim o tratamento probabilístico pode levar a conclusões erradas. Além do intervalo de confiança das medidas, que indica a acurácia dessas medidas (cujas variações são explicadas pela ocorrência de flutuações aleatórias nas condições de mensuração e na capacidade de observação do sujeito avaliador), costuma-se considerar outros dois tipos de erro, o Erro do Tipo I, que significa o falso positivo e o Erro do Tipo II, que representa o caso de falso negativo. Um falso positivo ocorre quando a inferência indica como resultado provável algo não verdadeiro e o falso negativo quando indica como não provável algo verdadeiro. Esses erros indicam falta de precisão nas medidas ocasionada por erros sistemáticas, provenientes da inadequação do instrumental de

medida, de problemas amostrais ou de problemas de modelagem, estes últimos provocados pela falta de verificação de todas as possíveis explicações concorrentes. Talvez outras seqüências de dedução se mostrem também plausíveis ou até mesmo mais verossímeis.

Para uma conclusão estatística ser considerada válida, é preciso existir uma associação altamente provável de uma relação de causa e efeito modelada previamente. A *validade interna* da associação estatística entre as medidas é considerada a demonstração de uma causalidade real, desde que a manipulação dos dados realizada obedeça às propriedades das escalas e das estruturas dos conjuntos de medidas, adequação esta chamada *validade de construto*. A *validade externa* significa que os resultados podem ser generalizados e comparados com outros casos e períodos.

Mesmo uma inferência estatística estruturalmente válida está sujeita a erros aleatórios e sistemáticos dos tipos I e II. Entretanto, para aumentar a validade estatística das conclusões, ou seja, a validade interna, pode-se:

- a) pesquisar amostras maiores, o tamanho ideal depende da magnitude esperada do efeito em relação à variância esperada;
- b) diminuir as fontes externas de erro circunscrevendo o estudo a populações homogêneas e usando instrumentos padronizados para medidas;
- c) identificar fontes externas de variância, por meio de análise de covariância e análise fatorial;
- d) aumentar a confiabilidade nas medidas, aprimorando os instrumentos de mensuração e, quando possível, aumentando o número de replicatas;
- e) padronizar a implementação do tratamento, de preferência proporcionando um alto grau de exposição ao tratamento.

Aumentar a validade de construto significa aprimorar o modelo presumido de uma relação de causa e efeito, considerando as variáveis utilizadas, as propriedades das escalas adotadas para essas variáveis e as relações hipotéticas entre elas. Para isso, o procedimento consiste em verificar, por meio de um pré-teste se diferentes formas de medir uma variável se correlacionam e quais variáveis distintas não covariam. A análise de fatores, a análise de correspondência ou testes como o α de Cronbach podem ser usados para esses fins.²⁹

²⁹ Cf. Pereira (1999), Hair *et al.* (1998) e Dillon & Goldstein (1994) para maiores detalhes sobre essas e outras técnicas. Elas não são detalhadas neste trabalho.

Para aumentar a validade externa, a amostragem aleatória precisa ser expandida para considerar elementos da população que compõe o universo generalizado, é preciso considerar um grupo heterogêneo de pessoas e outras condições de contorno e horizontes de tempo. De modo complementar, pode-se ainda comparar a generalização com situações similares para compatibilizar as variáveis.

Para completar a apresentação, é oportuno reproduzir o esquema de execução sugerido para uma avaliação de programas sociais. Segundo Rutman (1977), a execução de uma avaliação consiste de vários passos, em geral obedecendo a uma seqüência como a proposta abaixo:

1) Determinar os propósitos da avaliação, objetivos explícitos e implícitos das diversas partes interessadas.

O primeiro passo de uma avaliação consiste em identificar claramente os objetivos da avaliação (geralmente distintos dos objetivos do objeto avaliado) e evitar diversos tipos de agendas ocultas, que podem introduzir ruídos e distorções nos resultados. Desse modo, é preciso evitar, no plano do desenho metodológico, que as relações de causa e efeito presumidas sejam modeladas com deliberada superficialidade ou viesadas para mostrar bons ou maus resultados. Na condução dos trabalhos, é preciso evitar que ocorra ocultação de falhas durante a investigação e uso político de resultados intermediários. No plano administrativo, tanto a avaliação mandatória realizada como ritual inócuo como a avaliação para justificar e legitimar decisões já tomadas precisam ser evitadas.

2) Especificar as variáveis.

Nesta etapa, já eliminadas as interferências que poderiam obstruir a transparência do modelo de avaliação, são identificados os elementos necessários para construí-lo. É preciso descrever em detalhes o objeto, quais são seus componentes e sua organização, quais são seus objetivos e os efeitos presumidos. Além disso, é preciso considerar seus antecedentes, ou seja, fatores independentes do programa mas que compõem seu contexto e as restrições ou condições que o afetam. Por fim, as variáveis de intervenção, aquelas que podem colaborar ou atrapalhar a consecução dos objetivos.

3) Obter as medidas.

Para se efetuar as medidas, é necessário obter indicadores, definidos sobre escalas mensuráveis, que captem adequadamente o comportamento das variáveis dos modelos. A

confiabilidade dos resultados depende dessa adequação. Para garantir que a variação dos dados obtidos seja real, isto é, não venha de fatores externos e inconsistências é preciso controlar as fontes de inconformidade. Entre tais fontes se destacam: fatores subjetivos dos respondentes como humor, fadiga e motivação; fatores subjetivos do observador; condições de observação que desviem a medida correta; problemas com instrumentos de medida (questões mal elaboradas, por exemplo); problemas de processamento das medidas.

4) Testar as considerações do desenho metodológico.

Uma vez obtidos dados confiáveis, resta testar a validade das inferências de causa e efeito presumidas nos passos 1 e 2. Como já comentado acima, há três dimensões de validade, a validade interna, a validade de construto e a validade externa, às quais se aplicam as considerações já adiantadas. Desta etapa, a avaliação fornece como resultado a mais provável descrição da realidade que o desenho metodológico é capaz de obter e suas possibilidades de generalização. Esse resultado, expresso em forma de modelos de causalidade discrimina variáveis independentes (causas) e dependentes (efeitos) e os níveis de plausibilidade das relações entre elas. Parâmetros estatísticos de probabilidade aparecem associados a todas as medidas e a todas as correlações observadas.

5) Aplicar os achados.

Por fim, os resultados de uma avaliação estão prontos para serem utilizados como referência para a tomada de decisão. Entretanto, Rutman (1977) alerta que é comum os resultados de uma avaliação serem subtilizados. São identificadas causas como falta de comunicação, falta de envolvimento no planejamento e na condução da avaliação, *timing* inadequado das respostas e fraca disseminação como sendo típicas desse tipo de situação. Uma causa importante apontada é a separação das tarefas de planejamento, administração e avaliação em processos distintos dentro das estruturas organizacionais. Conseqüentemente, dificilmente se consegue lograr que as avaliações provoquem mudanças estruturais nas organizações.

Como recomendação geral, esse esquema de passos é de grande utilidade prática para qualquer avaliação, especialmente em relação aos conselhos presentes nos comentários de cada etapa. Alguns deles, como é fácil antever, remetem a problemas estruturais da racionalidade científica de objetivação quando aplicada a objetos sociais.

XI. ANOMALIAS DA AVALIAÇÃO OBJETIVADA³⁰

O interesse de dotar as avaliações de objetos sociais de um corpo teórico e metodológico consistente, apoiado na objetividade e em premissas científicas, não é recente. Há inúmeras manifestações de métodos de avaliação objetivados em economia, em administração, em sociologia, em ciências políticas, em antropologia e assim por diante, com grande variação em seus procedimentos e focos. Praticamente toda teoria social pode ser associada a algum tipo de avaliação objetivada. Há casos nos quais uma teoria conhecida é utilizada para modelar uma avaliação e há outros nos quais algum método conhecido de avaliação é utilizado para demonstrar uma teoria.

O desenvolvimento da avaliação objetivada em sua forma mais rigorosa confunde-se com o próprio desenvolvimento das técnicas de amostragem e inferência estatística. Desde o sucesso na previsão do resultado das eleições presidenciais de 1936 nos EUA, quando com uma amostra cientificamente construída de apenas 1.500 pessoas os recém fundados institutos Gallup, Roper e Crossley adiantaram a vitória de Franklin Roosevelt, o interesse no aperfeiçoamento e aplicação de técnicas científicas de avaliação cresceu e elas rapidamente suplantaram os métodos típicos da época, centrados na entrevista a personalidades e especialistas selecionados (SUDMAN *et al.*, 1996, p. 5).

Contudo, foi devido ao contexto do Welfare State e dos programas de transferência tecnológica e de treinamento a países pobres capitaneado pelos EUA nos anos 1960 que a demanda por avaliações se expandiu definitivamente. As organizações que suportavam programas sociais precisavam demonstrar quantitativamente seus desempenhos e havia ceticismo quanto à real efetividade de muitos programas e divergências entre avaliações.

O tratamento estatístico rigoroso para determinar os impactos e a efetividade desses programas surgiria então como a melhor solução. As práticas de *evaluation research* passaram a ser crescentemente aceitas como insumo para o planejamento e a tomada de decisão e, segundo Rutman (1977), o sentimento era de que muitos dos impedimentos para

³⁰ Nesta seção, o termo *anomalia* é usado segundo o conceito de Kuhn. Uma anomalia ocorre quando, de alguma maneira, a natureza viola as expectativas paradigmáticas da ciência normal (KUHN, 2000, p. 78). Apesar de uma anomalia ser em geral reconhecida pelos cientistas, geralmente ela não é suficiente para desafiar um paradigma. Apenas quando um número expressivo de anomalias ou alguma anomalia considerada mais séria passa a ser amplamente reconhecida é que se viabiliza a construção de um paradigma concorrente.

a tomada de decisão objetivamente racional poderiam ser suplantados pelo uso dessas técnicas de avaliação.

Entretanto, os próprios entusiastas do rigor científico em avaliações logo perceberam que havia problemas, que suas metodologias não funcionavam como deveriam, em tese, funcionar. Diversos autores do *mainstream* da avaliação objetivada identificaram problemas em suas abordagens. Dentre as principais críticas, era comum relacionar os problemas identificados à falta de rigor nos procedimentos, insuficiência das técnicas ou mal desenho metodológico.

Campbell (1979) é um dos autores que reconhece que a avaliação de programas sociais, especialmente aquela identificada como abordagem quantitativa, possuía muitos problemas. Todavia, ele enxergava nisso estímulo e justificativa para avançar ainda mais na rota de desenvolvimento metodológico dada pelos pressupostos de objetividade. Campbell fornece uma instrutiva lista dos problemas comuns que enfrentavam as avaliações desenhadas como experimentos e que precisariam ser superados (CAMPBELL, 1979, pp. 49 e 50):

- “Experimentos que implementam um único tratamento em um único caso dado são profundamente ambíguos sobre o que causou o que;
- o processo é geralmente negligenciado em um desenho experimental focado nos resultados gerais de um tratamento complexo, e saber somente os efeitos pode levar a implicações equivocadas quanto a replicação e melhoria do programa;
- programas grandes são geralmente ambíguos quanto a seus objetivos e indicadores;
- há precariedade na rigidez dos sistemas de medida, limitando a coleta de dados àquelas dimensões antecipadas *a priori*;
- mudanças no objeto no decorrer da avaliação torna a comparação experimental entre *inputs* e *outputs* impraticável;
- programas sociais são geralmente implementados de modo que são pobres do ponto de vista de desenho experimental;
- mesmo sob situações controladas, experimentações são tediosas e facilmente equivocadas;
- a experimentação é muito lenta para ser politicamente útil.”

Além disso, o autor acrescenta que uma avaliação controlada experimentalmente geralmente negligencia evidências contextuais relevantes, cria uma exagerada dependência em relação a apenas algumas abstrações quantificadas e ignora evidências qualitativas contraditórias. Por outro lado, argumenta que inferências são intrínsecas ao contexto social da implementação de um problema. Elas não seriam uma ficção resultante do desenho do experimento, inferências são construídas socialmente independentemente de haver ou não uma avaliação sistemática. Assim, um bom desenho experimental efetivamente ajudaria a aliviar as distorções de um contexto particular, socialmente idiossincrático, e revelar os efeitos verdadeiros, livres de ruídos (CAMPBELL, 1979).

A lista de Campbell é instrutiva porque, ao apresentar problemas percebidos na execução de avaliações, revela na realidade anomalias para as quais a abordagem objetivada do problema não conseguia – e não podia – obter solução. Em primeiro lugar, o problema da complexidade, que impede distinguir nitidamente variáveis dependentes e independentes. Em seguida, a impossibilidade de definir o objeto da avaliação, seja porque fosse composto de inúmeros outros objetos (muitos deles pessoas que, como sabia Simon, não necessariamente possuem os mesmos objetivos que o programa), seja porque o objeto simplesmente aprendia e se alterava durante a avaliação. Além disso, Campbell se lamenta do fato que os programas sociais não serem implementados como experimentos e não considera legítimo fazer concessões metodológicas em nome do pragmatismo político. Para ele, o que se ganha com o rigor supera o que se perde pela rigidez metodológica.

Outro autor que reconhece as dificuldades da abordagem objetivada é Rutman (1977). Ao descrever os procedimentos gerais para avaliações, faz referência a diversas armadilhas que podem comprometer o bom desempenho do desenho metodológico. Rutman identifica, por exemplo, importantes fatores que podem prejudicar uma inferência causal. Entre eles, efeitos de aprendizado devidos à repetição de testes, efeitos de maturação devido às mudanças naturais da passagem do tempo (por exemplo, envelhecimento da amostra), efeitos de seleção devido à desistência ou perda da amostra e efeitos de múltipla atribuição (outros eventos simultâneos que afetam as mesmas variáveis). Além disso, esses fatores podem interagir e complicar ainda mais a inferência causal. Do ponto de vista da capacidade de generalização do modelo (validade externa), o autor destaca que o resultado particular que seria usado para induzir o resultado geral pode ter sido influenciado por

interações capazes de superestimá-lo ou subestimá-lo. São elas: a interação de tratamentos (o resultado inclui a influência de mais de um programa), a interação entre teste e tratamento (a avaliação influencia o programa e o torna fora do padrão, impedindo a generalização), a interação entre seleção e tratamento (a amostragem é incorreta ou se modifica de modo não aleatório durante o tratamento) e a interação entre teste e fatos (a avaliação capta apenas parcialmente os fatos).

É interessante perceber que as dificuldades relatadas não são encaradas como intransponíveis, como elementos constituintes da própria natureza do objeto e, portanto, indissociáveis dele. Elas são percebidas e encaradas como problemas difíceis, mas contornáveis. Rutman inclui ainda na mesma lista dos problemas listados no parágrafo anterior problemas metodológicos do tipo instabilidade e baixa confiabilidade nas medidas, mudanças no instrumento de medida e escolha do método de regressão. Esses problemas não pertencem à mesma categoria dos anteriores, embora o autor pareça não se dar conta. Problemas metodológicos envolvendo a escolha de variáveis, indicadores e métodos de regressão podem ser contornados e evitados por meio de um desenho adequado, enquanto o aprendizado do objeto e a impossibilidade de isolar um sistema social para avaliá-lo não podem.

Como observa Kuhn (2000), uma característica importante de um paradigma é sua capacidade de resistir às anomalias. A insistência em usar parâmetros científicos objetivados para legitimar uma avaliação é exemplar nesse sentido. Não importa que o objeto não se enquadre no cânone – e na realidade ainda existem aqueles que insistem que *o objeto deveria se enquadrar* e chegam a propor absurdos como aleatorizar e padronizar programas que necessitam justamente ser específicos para funcionar ou manter grupos de controle, que ao não serem beneficiados, são prejudicados. Se, por princípio, a abordagem utilizada é tida como correta em sua essência, é difícil perceber os fundamentos de sua inadequação.

Algumas tentativas de superar as anomalias são emblemáticas. Uma primeira ilustração é dada pela oposição que se construiu entre as tradições metodológicas ditas qualitativas e quantitativas que, diga-se de passagem, foi bastante importante décadas atrás, e talvez ainda o seja em alguns círculos acadêmicos e burocráticos mais arraigados.

A distinção qualitativo *versus* quantitativo não é uma boa distinção, tanto que no capítulo anterior foi considerada pouco relevante como atributo para caracterizar avaliações. Não é boa porque qualificar algo não necessariamente evita medidas e inferências causais e quantificar não elimina a necessidade preliminar de qualificar. Conforme foi apontado no Capítulo 1, para um atributo pode-se associar uma medida, mesmo que a escala dessa medida seja qualitativa. O indivíduo só pode fazer um juízo se for capaz de captar a distinção entre duas situações em um atributo, a mudança uma vez reconhecida ou medida (tanto faz qualitativamente ou quantitativamente) permite ao indivíduo aplicar sobre ela um juízo, presente em seu conjunto de chegada Cr_i (ou adotar um juízo já implícito na própria escala associada ao atributo). Quanto às inferências causais, todo indivíduo *aprende* por meio de seus ciclos de avaliação e decisão e adquire estruturas conceituais do tipo causa e efeito que funcionam como modelos para aqueles eventos que sugiram repetibilidade (cf. seção V). Se a tradição quantitativa da avaliação objetivada exagera a mão e insiste em modelar como estruturas causa e efeito puras e objetivas programas sociais e seus impactos, isso não significa que uma tradição qualitativa prescindia da inferência, porque qualquer conclusão precisa de inferência. Os indivíduos necessariamente adquirem suas próprias interpretações das relações causais que ocorrem mesmo nos casos mais complicados como intervenções de programas sociais e seus impactos. Uma grande dificuldade para os avaliadores imbuídos de racionalidade objetiva é aceitar que as racionalidades desses indivíduos, quaisquer que sejam, também podem estar corretas. Essa aceitação não tem nenhuma conexão necessária com quantificação ou qualificação.

Entretanto, essa oposição foi muito explorada no meio especializado. A avaliação quantitativa seria aquela que adota experimentos aleatórios, quase-experimentos, análises multivariadas, *surveys* amostrais e assim por diante. Avaliação qualitativa incluiria etnografia, estudos de caso, entrevistas detalhadas e observação participante (REICHARDT & COOK, 1979). Havia também a tendência de apresentar essa oposição como uma disputa entre paradigmas. Bogdan e Taylor (1975, *apud* REICHARDT & COOK, 1979) a ilustraram por meio das características estereotipadas de dois supostos paradigmas concorrentes, no Quadro 2.1 abaixo.

Quadro 2. 1 – Características que opõem os Paradigmas Qualitativo e Quantitativo de avaliação (Bogdan e Taylor, 1975).

Paradigma Qualitativo	Paradigma Quantitativo
Defende o uso de métodos qualitativos	Defende o uso de métodos quantitativos
Fenomenológico – Entender o comportamento humano a partir do próprio quadro de referências do ator.	Positivismo-Lógico – busca os fatos e as causas de um fenômeno social com pouca atenção aos aspectos subjetivos dos indivíduos
Observação naturalista e não controlada	Medidas controladas e interferentes
Subjetivo	Objetivo
Próximo aos dados, perspectiva interna do problema	Independente dos dados, perspectiva externa
Orientado à descoberta, exploratório, expansionista, descritivo e indutivo	Orientado à verificação, confirmatório, reducionista, inferencial e hipotético dedutivo
Processo	Resultados
Validade. Dados "reais", "ricos" e "profundos"	Confiabilidade. Dados "duros" e replicáveis
Não generalizável, casos singulares de estudo	Generalizável, casos múltiplos de estudo
Holístico	Particular
Assume uma realidade dinâmica	Assume uma realidade estável

Antes de qualquer outro comentário, é preciso adiantar que esta divisão é problemática. O ideal da avaliação objetivada e científica está bem descrito pelo Paradigma Quantitativo, exceto pela caracterização do Positivismo-Lógico que está imprecisa. Como argumenta Popper (1975), o Positivismo-Lógico não é caracterizado por ser hipotético dedutivo, mas pelo contrário, esta corrente acredita que as explicações verdadeiras são primeiro obtidas por indução de observações experimentais, inclusive as hipóteses verdadeiras que são depois desdobradas dedutivamente e conservam seu valor de verdade. Ademais, não é exato que o Positivismo-Lógico não se preocupe com os sujeitos, ele procura deliberadamente objetivá-lo ou, como último recurso, evitá-lo.

Feita a ressalva, as demais qualificações mesclam objetivos metodológicos desejáveis do ponto de vista de uma avaliação objetivada, características dos instrumentos utilizados e algumas das limitações decorrentes, todas elas já comentadas anteriormente.

Já o quadro fornecido para o Paradigma Qualitativo é bem mais pitoresco. Salta aos olhos que ele foi deliberadamente construído em oposição ao primeiro. Além disso, reflete uma predileção implícita (ou quase explícita) de seus autores por esse último. Isso talvez mais indique a direção para onde eles apostariam avançar para superar as anomalias do que

apresente as características consolidadas de uma real alternativa à legitimidade da racionalidade cientificamente objetivada representada pelo Paradigma Quantitativo.

Como a discussão de caminhos para a superação da avaliação científica faz parte da discussão mais ampla deste capítulo, neste ponto do texto os itens apresentados para caracterizar a oposição ao Paradigma Quantitativo no Quadro 2.1 não serão comentados em detalhe, em favor de oportunidade mais adequada. Há, entretanto, um problema de fundo que deve ser apontado nessa controvérsia, essencialmente colocada como controvérsia entre métodos. Ao se sugerir uma distinção do tipo da ilustrada pelo Quadro 2.1 e estabelecê-la como oposição entre paradigmas, assume-se, de modo equivocado, que os métodos são diretamente e irrevogavelmente ligados a universos conceituais rígidos e inescapáveis. Se assim fosse, em decorrência, métodos qualitativos e quantitativos jamais poderiam ser usados conjuntamente. Além disso, todas as demais distinções entre, por exemplo, objetivo e subjetivo, particular e holístico, resultados e processos seriam sempre excludentes entre si. Misturar as coisas seria "ingenuidade" epistemológica porque invariavelmente a balança iria pesar para um ou para outro paradigma e os elementos complementares ficariam destituídos de significado.

Esse tipo de conclusão é evitado se a avaliação for encarada conforme a proposta do capítulo anterior, a legitimidade de um método vem de sua aceitação intersubjetiva, que é socialmente construída e pode ser mais ou menos institucionalizada. Assim, faz pouco sentido falar de método melhor ou método pior, há sim premissas e métodos mais ou menos adequados a determinadas comunidades e situações de aplicação. O que está em debate aqui é como certos conjuntos de métodos, identificados com diferentes premissas, em sua luta por legitimidade se colocam como alternativas exclusivas. Nesse sentido, mesmo o suposto Paradigma Qualitativo apresentado requisita para si hegemonia sobre qualquer outro, o que é, de certo modo, motivado pela convicção de estar mais próximo da verdade, de ser mais evoluído. Porém, não é assim que se deve interpretar um processo de mudança na legitimidade entre diferentes métodos de avaliação. Essa mudança ocorre naturalmente condicionada por mudanças no contexto social e não há qualquer sentido implícito de que evolui para melhor ou para pior, tal juízo é necessariamente conjuntural e dependente do observador.

Desse modo, é natural que aqueles que mantêm convicção sobre a correção de um tipo de abordagem considerem muito alto o preço a pagar para mudar de convicção, sentem que muita coisa importante seria perdida (e realmente elas são perdidas, como mostra Kuhn). Eles preferem concentrar seus esforços para superar as anomalias a partir de suas próprias armas. Por outro lado, aqueles que acatam uma nova abordagem, sabem que ela é inferior à anterior em muitos pontos e só lentamente oferece novas alternativas plausíveis para solucionar velhos problemas, muitos deles satisfatoriamente resolvidos pela antiga abordagem.

O exemplo dos quase-experimentos é típico dessa situação, novas ferramentas sobre velhas bases para tratar das anomalias. Boruch e Rindskopf (1977), conscientes da dificuldade de fazer avaliações de objetos sociais como se fossem experimentos científicos, adotam a designação *quase-experimento* para descrever uma aproximação à situação experimental típica de um laboratório e que seria desejada para métodos de avaliação. Neles, a aleatorização ainda traria importantes vantagens, independentemente do tipo de aproximação utilizado, uma vez que “permite empacotar adequadamente a ignorância em erros probabilísticos, aliviar a preocupação com os inúmeros fatores que podem afetar as medidas e evitar distorções nas avaliações, pois pressupõe regras estáveis para o tratamento dos dados” (BORUCH & RINDSKOPF, 1977). Esquemas sofisticados de aleatorização em quase-experimentos mitigam várias anomalias, porém com enormes custos e perdas na capacidade de inferência. Situações que justificam essa abordagem incluem aquelas em que a amostragem é condicionada por critérios externos (de alocação, políticos, ideológicos etc.), aquelas em que um tratamento é implementado universalmente e não há grupo de controle contra factual ou aquelas em que se investiga o impacto de uma mudança depois dela já ter ocorrido.

Essas situações correspondem a muitas situações reais nas quais avaliações são realizadas. Todas elas são problemáticas para uma abordagem que as interpretam como problemas científicos. Para ser efetiva, a avaliação precisaria acompanhar no tempo e no espaço um objeto social, isolando as variáveis previstas em seu modelo de interpretação. Em certos casos isso é possível, por exemplo, numa avaliação que objetiva o acompanhamento financeiro de uma organização ou a mudança no tempo de parâmetros que podem ser medidos sem maiores dificuldades como expectativa de vida, número de

alfabetizados, renda média e muitos outros. Em outros, como em avaliações de impactos e de bem-estar, isso se torna demasiado complicado.

Esta seção se encerra com a reprodução de uma definição recente da OCDE para avaliação. Para essa organização, "avaliações são julgamentos cuidadosos (*assessments*), que enfatizam a consistência e a utilidade de seus achados, acerca dos resultados de políticas públicas, organizações ou programas. Seu papel é melhorar a informação e diminuir a incerteza; entretanto, mesmo avaliações baseadas em métodos rigorosos apoiam-se significativamente em julgamentos. Os objetivos principais das avaliações são melhorar a tomada de decisão, a alocação de recursos e a legitimação (*accountability*). Isso pode se conseguido informando o público, informando os processos de tomada de decisão e encorajando o aprendizado organizacional. Avaliações são parte de um esquema mais amplo de gestão de desempenho. Elas podem suplementar e melhorar a gestão, mas não a substituir" (OCDE, 1999, p. 6).

Há nessa definição um pragmatismo que não corresponde à situação da década de 1970 descrita por Rutman em 1977. Como será visto no próximo capítulo, ocorreu desde então uma importante mudança de perspectiva sobre a função das avaliações nos processos de mudança técnica e social. Porém, enquanto do ponto de vista conceitual esta nova perspectiva já pode ser considerada madura, do ponto de vista metodológico ela ainda é frágil, limitando-se no geral a um uso eclético de métodos tradicionais.

Na próxima seção, o marco analítico proposto no Capítulo 1 é retomado para examinar outros possíveis caminhos metodológicos. Na seqüência, serão introduzidos alguns exemplos de abordagens matemáticas alternativas que se constituem em naturais candidatas à matéria-prima para métodos de avaliação que valorizem o subjetivo, a participação e o aprendizado.

XII. MÉTODOS OBJETIVADOS DE AVALIAÇÃO — APLICAÇÃO DO ESQUEMA ANALÍTICO

No esquema de análise introduzido no capítulo anterior para interpretar avaliações, parte-se do princípio que todos os métodos de avaliação são construções sociais. Em uma avaliação, ter método significa que o mecanismo que transforma atributos em critérios é conhecido e codificado, sendo, portanto, passível de comunicação e legitimação social.

O emprego de métodos baseados em premissas científicas de objetivação é sem dúvida uma estratégia eficiente de legitimação. Isso faz com que as avaliações aparentemente prescindam de convenções sociais ou éticas para que seus resultados sejam aceitos. Entretanto, não é demais insistir que fatores subjetivos estão sempre presentes em avaliações e organizações ou instituições sociais sempre exercem influência. O ideal de neutralidade sustentada por avaliações científicas não pode, por definição, ou melhor, por construção social, evitar os valores e escolhas que estão em sua base.

Toda instituição social e toda organização sinalizam os indivíduos sobre qual ou quais métodos ele deve considerar legítimos para subsidiar suas decisões. Assim como também limita os conjuntos de partida (atributos) e chegada (critérios) relevantes para as avaliações,³¹ a ponto de cristalizar certos métodos e critérios de adequação para as decisões em uma racionalidade própria da organização ou em regras de comportamento socialmente aceitas.

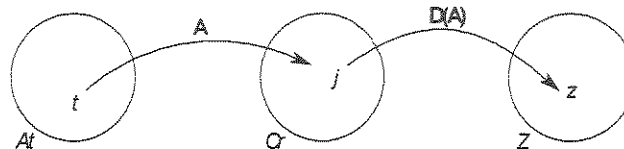
É claro, desde logo, que há vantagens em abordar uma avaliação como um problema científico. Há ganhos evidentes quando os objetos são estáveis e facilmente mensuráveis e quando os objetivos são bem definidos, seja por serem objetivos próprios de uma organização, seja por serem socialmente unânimes. Nesses casos, os fatores subjetivos e institucionais não são empecilho porque podem ser tomados como constantes para todo indivíduo envolvido. Há ganhos porque a avaliação pode ser precisa e independente de qualquer indivíduo em particular, o que a legitima como resultado *racionalmente objetivo e verdadeiro* do qual decorre naturalmente uma decisão.

O critério de validade, nesses casos, é o juízo de verdade fornecido pela lógica científica. A legitimidade é conseguida por um mecanismo de transferência, segundo o qual os indivíduos reconhecem a precedência do juízo dado como verdadeiro e o igualam a seus próprios juízos, passando a acreditar na função $A: At \rightarrow Cr$ que o mecanismo de transferência estabelece e a defender sua veracidade e universalidade. Esse mecanismo de legitimação é fundamental para a existência das organizações e para a coesão social.

O conjunto Cr é, na avaliação objetivada, diretamente o conjunto de todos os juízos possíveis para os estados de At . Cr pode ser construído de diversas maneiras, mas é sempre

³¹ Há nessa afirmação um nítido paralelo com a noção de cerceamento da racionalidade imposta pelas organizações aos indivíduos, segundo os conceitos de racionalidade organizacional e racionalidade limitada Simon (1962).

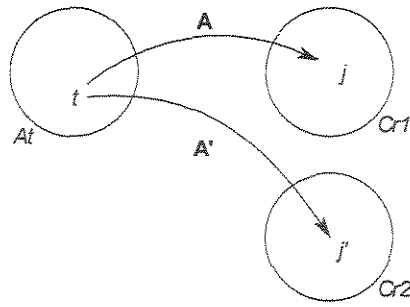
independente dos critérios pessoais que os indivíduos possam ter (o que não impede que os indivíduos o adotem em seguida como referência). Se Cr existe, então também se pode supor a existência de $D(A):Cr \rightarrow Z$, onde Z é o elenco de todas as ações possíveis e a escolha de uma alternativa z pode ser chamada de uma escolha objetivamente racional porque foi apoiada por um juízo verdadeiro, obtido objetivamente.



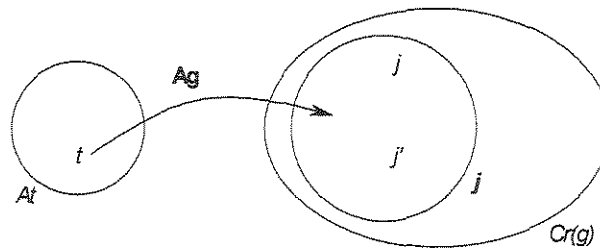
Entretanto, há casos não tão bem comportados que também requerem avaliações. Como foi amplamente discutido, quando se avalia um objeto social, ou quando há objetivos e valores conflitantes em pauta, métodos apoiados exclusivamente na racionalidade científica objetivadora se complicam.

Um objeto social – pode ser um indivíduo, uma organização, uma comunidade etc. – é essencialmente um objeto cuja mutação, por causa de sua capacidade de aprendizado, não pode ser satisfatoriamente modelada nem prevista, salvo ocasiões muito especiais. Isso mina as possibilidades de replicação e de universalização da maior parte dos resultados obtidos, mesmo que pela via experimental.

A existência de valores conflitantes, ou mesmo parcialmente conflitantes, significa que existe pelo menos dois diferentes modos de interpretar uma mesma medida (mesmo no caso de um objeto estável), do que decorre a impossibilidade de haver um método (mecanismo) único que dê conta de codificar essas duas situações simultaneamente preservando suas identidades. O diagrama a seguir, já reproduzido anteriormente, mostra a existência de duas instâncias de avaliação, A e A' , que levam aos juízos incompatíveis j e j' nos indivíduos 1 e 2. Na ausência de um critério externo de verificação, se for possível codificar A e A' ambos se tornam métodos válidos e incompatíveis.



Entretanto, é possível propor um método que generalize a incompatibilidade e forneça um novo juízo. O diagrama a seguir ilustra essa situação. A avaliação generalizada $Ag: At \rightarrow Crg$ leva a medida a um conjunto de juízos Crg que generaliza o anterior. Os elementos de Crg são composições dos elementos dos critérios dos indivíduos. Desse modo, o juízo j obtido de t é alguma composição dos juízos originais e não reflete necessariamente nenhum deles. Do ponto de vista da avaliação Ag o elemento $j = \{j, \dots, j'\}$ é único e os elementos originais não mais importam nem podem ser recuperados.



Em avaliações coletivas, a presença inevitável de valores conflitantes exige um método que force um resultado generalizado sobre o qual não se aplicaria decisão. Supor a possibilidade de $D(Ag): Crg \rightarrow Zg$ é supor a possibilidade da escolha social.³²

Uma avaliação coletiva se limita a induzir uma avaliação particular em cada instância participante e somente desta é que pode sair o impulso para escolher uma ação de cada repertório Z_i .

³² A função decisão foi definida no Princípio 2 como válida apenas para o indivíduo. No caso suposto $D(Ag): Crg \rightarrow Zg$, a função teria que operar a partir de um domínio Crg artificialmente construído e não pertencente, em princípio, a nenhum indivíduo. Seria, portanto, um caso incompatível com a definição estabelecida no Capítulo 1. Daí o porque dela não valer para avaliações coletivas.

Esclarecidos esses pontos, quais seriam então possíveis propriedades esperadas para outros métodos de avaliação?

Depois de tudo o que foi discutido neste capítulo, torna-se evidente a intenção de buscar possibilidades metodológicas que permitam tratar de forma diferenciada os casos problemáticos de avaliação de objetos sociais. Esses métodos precisam de propriedades tais que os torne adequados a tratar os casos que métodos restritos as abordagens objetivadas convencionais tratam mal.

Desse modo, a pergunta pode ser recolocada mais precisamente. Que propriedades um método precisa apresentar para dar conta de problemas com objetos sociais e múltiplos juízos?

Para se desenhar uma resposta a essa questão, é preciso antes mergulhar no universo das estruturas lógicas e matemáticas que dão sustentação para os modelos conceituais que objetivam juízos e objetos sociais. O problema da objetivação dos juízos corresponde a uma longa tradição de estudos no tema da tomada de decisão e na proposição de métodos para a tomada de decisão. O principal recurso teórico desta tradição é o estudo e a modelagem de estruturas de preferências, problema que será tratado a seguir.

É somente a partir de um entendimento mais profundo dos pressupostos assumidos como propriedades características das preferências dos indivíduos que se pode discutir alternativas adequadas para se construir metodologias de avaliação de objetos sociais. Há uma correspondência intrínseca, muitas vezes não percebida, entre as propriedades lógicas dos métodos de avaliação objetivados e a rigidez que leva aos artificialismos típicos de seus procedimentos no trato de objetos sociais.

As duas próximas seções tratam desses assuntos. A explanação a seguir procura condensar o assunto e restringe-se somente ao estritamente necessário para a discussão, inclusive com a intenção de demarcar os aspectos realmente essenciais. A seção XIII apresenta os elementos formais básicos que constituem as estruturas matemáticas utilizadas para se modelar as preferências dos indivíduos e discute tópicos como a diferenciação formal entre atributos e critérios e as condições mínimas necessárias para se inferir julgamentos a partir dos modelos de preferência. A seção IX exemplifica essa discussão a partir do caso dos métodos multicritério de apoio à decisão, demonstrando como escolhas

bastante simples do ponto de vista formal impactam fortemente nas propriedades dos métodos e no significado de seus resultados.

XIII. A ESTRUTURA DAS PREFERÊNCIAS

Na seção VII, o problema da construção de métodos para a escolha social foi apresentado em linhas gerais, com destaque para suas controvérsias originais e para o resultado fundamental de Arrow. Nas seções IV e V diferentes interpretações para *decisão* foram discutidas, procurando-se marcar as diferenças entre as interpretações prescritivas e as interpretações descritivas dos fenômenos decisórios e a conexão existente entre as noções de decisão e causalidade. Nesta seção, o tema decisão volta à baila, agora sob uma perspectiva metodológica. Esta seção examina as estruturas matemáticas básicas que foram aplicadas a diferentes tentativas de explicar e modelar decisões.³³

Para que se possa examinar em profundidade as estratégias de modelagem das preferências e os métodos que as utilizam, é necessário comentar alguns aspectos importantes sobre estruturas matemáticas. Na seção II foi definido que, numa avaliação, o conjunto de partida At dos estados dos atributos se diferencia do conjunto de chegada Cr (critérios). Isso significa que esses conjuntos podem possuir propriedades estruturais distintas. De fato, o conjunto At é definido como uma estrutura que pode ser menos rígida do ponto de vista algébrico que a do conjunto Cr .

Enquanto Cr é sempre no mínimo um reticulado,³⁴ At pode ser também desordenado ou parcialmente ordenado (sem ser reticulado).

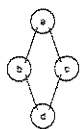
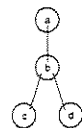
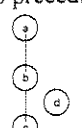
³³ A noção de decisão utilizada pelos teóricos do campo de tomada de decisões difere um pouco da noção apresentada como *função decisão* pelo Princípio 2 do Capítulo 1. Para a tradição do estudo dos problemas de decisão, está seria fundamentalmente um problema de escolha, orientado pelas preferências intrínsecas dos indivíduos. Para nosso esquema analítico a idéia de preferência está incluída no conjunto avaliação e decisão e é aproximadamente representada pela manipulação do conjunto Cr , dos juízos possíveis do indivíduo. Se for possível saber como ele manipula esse conjunto (ou se for possível evitar que ele o manipule livremente) é possível inferir a estrutura de preferências do indivíduo.

³⁴ Um reticulado é todo conjunto parcialmente ordenado $(U, \geq)$ tal que todo par de elementos de U possua um *sup* e um *inf* (abreviações de *supremum* e *infimum*) em U (BIRKHOFF, 1967, cap. 1 e NGUYEN & WALKER, 2000, pg. 20). Um reticulado, na prática, pode assumir um vasto número de configurações. A compreensão da estrutura dos reticulados é facilitada pela inspeção dos diagramas de Hasse, no qual as bolas representam os elementos de U e as linhas a relação $\geq$ (sempre que $x \geq y$ a bola de x é desenhada acima da de y e unida a ela por um traço). Como exemplo, toma-se abaixo um conjunto $U = \{a, b, c, d\}$, ordenado de diferentes modos:

A ordenação parcial dos elementos $\{x_1, \dots, x_n\}$ de um conjunto qualquer é dada por uma relação $\geq$ tal que para qualquer par de elementos $(x_i, x_j) \in \geq$ (par ordenado), valem simultaneamente as propriedades:

- P1. $x_i \geq x_i$ (a relação é reflexiva);
- P2. Se $x_i \geq x_j$ e $x_j \geq x_k$ então $x_i \geq x_k$ (a relação é transitiva);
- P3. Se $x_i \geq x_j$ e $x_j \geq x_i$ então $x_i = x_j$ (a relação é anti-simétrica).

Desse modo, o par ordenado $(At, \geq)$ indica um conjunto parcialmente ordenado, pois a inclusão de uma relação com as propriedades P1, P2 e P3 induz uma ordem parcial no conjunto At . Nota-se que, sendo um conjunto parcialmente ordenado nem todos os elementos de At são necessariamente incluídos na relação. Se isso ocorrer, isto é, se para qualquer par de elementos $(x_i, x_j) \in At$ valer $x_i \geq x_j$ ou $x_j \geq x_i$ o conjunto será completamente ordenado. Um exemplo típico de um conjunto completamente ordenado é o conjunto dos números naturais (com a relação $\geq$ assumindo o significado de “maior ou igual a”) e um exemplo de um conjunto parcialmente ordenado é o conjunto das cores (com a relação $\geq$ assumindo o significado de “mais claro que”).

-
- i.  Esta configuração é a de um conjunto totalmente ordenado: vale a cadeia $a \geq b \geq c \geq d$. É também um reticulado porque todo par de elementos sempre possui um *sup* e um *inf*. Toda cadeia limitada acima e abaixo é um reticulado. Em geral, os critérios possuem esta forma de ordenação.
- ii.  Este arranjo parcialmente ordenado é também um reticulado. É parcialmente ordenado porque entre os elementos b e c não vale a relação $\geq$, porém para todo par de elementos existe um *sup* e um *inf*.
- iii.  Este arranjo parcialmente ordenado não é um reticulado. Os elementos c e d ocupariam ambos a posição de *inf* no conjunto. Trata-se de uma situação de ambigüidade quanto ao ínfimo, uma vez que ambos são precedidos por b mas não são comparáveis entre si pela relação $\geq$.
- iv.  Este arranjo parcialmente ordenado também não é um reticulado. O elemento d não se relaciona com os demais e sempre será *sup* ou *inf* em relação a si próprio quando incluído em algum par de U .

Foi assumido que o conjunto Cr dos critérios é sempre um reticulado. Avaliar significa justamente operar uma função A que relacione todos os elementos de At em $Cr = \{At, \blacktriangleright\}$, onde o símbolo $\blacktriangleright$ significa a relação “é preferível a” (ou “mais correto”, “mais valoroso” etc.) ou numa escala independente de valoração $Cr = \{Cr, \blacktriangleright\}$.

Assim, ao aplicar seus juízos sobre um conjunto At de atributos, o indivíduo obtém em Cr uma ordem de preferências por exemplo do tipo $x_1 \blacktriangleright x_2 \blacktriangleright \dots \blacktriangleright x_{n-1} \blacktriangleright x_n$ que, em princípio, independe da existência de uma ordem similar (isomórfica) $x_1 \geq x_2 \geq \dots \geq x_{n-1} \geq x_n$ em At .

A ordem em Cr não precisa ser formada pelos mesmos elementos presentes em At . A estrutura do critério pode ser, por exemplo, booleana (constituída pelos juízos “verdadeiro” ou “falso”), de modo que *verdadeiro* $\blacktriangleright$ *falso* e a função A associe cada $x_i \in At$ a uma das duas alternativas. Pode ser também definida por julgamentos do tipo “sim”, “não”, “talvez”, “não sei”, de modo que não se opere a relação $\blacktriangleright$ entre “talvez” e “não sei”, mas haja uma estrutura de reticulado no conjunto parcialmente ordenado com “sim” como *sup* e “não” como *inf*.

A escala utilizada para modelar um critério pode ser discreta ou contínua, porém é sempre limitada acima e abaixo, isto é, existe um elemento simbolizando o mais preferível (denota-se *sup*(Cr)) e um elemento simbolizando o menos preferível (denota-se *inf*(Cr)).³⁵

A noção de que as decisões dependem das estruturas de preferência dos indivíduos é bastante antiga. O conceito de *utilidade* é um exemplo importante.³⁶ Segundo ele, cada indivíduo associa a toda opção de ação a uma utilidade $u(a)$, sendo essa utilidade definida na escala representada pelo intervalo real $[0, 1]$ de modo que se $u(a) \geq u(b)$ então $a \blacktriangleright b$. Nesse caso, a utilidade é modelada como sendo um atributo completamente ordenado dos objetos $a, b, \dots, n$ com rebatimento direto para uma estrutura de preferência isomórfica.

³⁵ Essa exigência faz com que toda escala para um critério possua de fato a estrutura matemática de um reticulado. A denominação *escala* é em normalmente sinônimo para um conjunto totalmente ordenado. Neste trabalho, entretanto, adota-se o termo *escala* também para conjuntos parcialmente ordenados, desde que sejam reticulados. Os limites acima e abaixo são necessários porque se medidas de preferência pudessem tender ao infinito isso traria problemas sérios no trato matemático dessas medidas (afinal a metade de infinito é ainda infinita, mesmo contendo menos elementos).

³⁶ A formulação matemática da utilidade (ofelimidade) como variável para explicar as decisões econômicas dos indivíduos é devida a Pareto, em 1909 (PARETO, 1984, v. 1, p. 84). Entretanto, a idéia de medir e comparar os gostos dos indivíduos numericamente remonta a 1789, no escopo da filosofia utilitarista de Jeremy Bentham (COHON, 1978, p. 165).

Esse tipo de modelo, ao imputar já aos atributos uma estrutura matemática isomórfica à estrutura dos critérios, representa um modo sofisticado de se construir um atributo. Por causa da estrutura matemática adotada, a avaliação se reduz simplesmente a alterar o conteúdo semântico da relação de ordenação obtida. Assim, mais utilidade é diretamente avaliado como mais preferência. Entretanto, justamente pelo fato da estrutura matemática ser sofisticada, é difícil encontrar atributos que obedeçam simultaneamente a todas suas propriedades (P1, P2, P3, ordem total e a existência de *sup* e *inf*) e possam, portanto, ser modelados com sucesso por essa estrutura. As premissas de racionalidade da teoria econômica neoclássica são questionáveis justamente por esse motivo: o comportamento dos indivíduos não é compatível com as propriedades da estrutura matemática empregada.

Curiosamente, a crítica ao conceito de utilidade geralmente centra fogo em sua não verificação empírica ou, por motivos metafísicos ou ideológicos (livre arbítrio ou liberdade individual), no fato do atributo comparar as escolhas de diferentes indivíduos a partir de uma mesma medida de utilidade supostamente universal e independente desses mesmos indivíduos, mas não em sua estrutura formal. Arrow, por exemplo, defendia que apenas era possível comparar as preferências de indivíduos a partir das diferentes ordenações de opções que cada um produzisse, sem quantificá-las. Uma ordenação não quantificada é uma estrutura matemática apenas um pouco mais fraca que o intervalo denso $[0, 1]$. Porém, a estrutura que Arrow utilizou é ainda uma ordem completa, valendo as propriedades P1, P2 e P3 e a existência de *sup* e *inf*. O que ele assumia como condição mínima de racionalidade era que as preferências dos indivíduos (e o resultado da agregação dessas preferências em uma escolha social) obedecessem a essas propriedades, o que nem sempre é verificável na prática e, conforme ele mostrou, impede formalmente a existência da escolha social.

Entretanto, do ponto de vista matemático é perfeitamente possível assumir estruturas mais fracas. Se, por exemplo, se considerar que o indivíduo não é capaz de comparar todas as alternativas de modo a estabelecer uma relação de ordem entre os atributos de certos pares de alternativas, pode-se modelar o conjunto dos atributos como sendo parcialmente ordenado. Mas se essa mesma estrutura parcialmente ordenada dos atributos não for um reticulado não é possível utilizá-la completamente como critério para o juízo. Se o indivíduo não conseguir estabelecer relações de preferência nas regiões extremas do

conjunto, restará ambigüidade que o impedirá de ajuizar a melhor ou a pior alternativa dentre todas e escolher. Na prática, isso ou impede o indivíduo de avaliar (e decidir), ou o força a adotar como critério apenas um recorte da ordem parcial do conjunto dos atributos ou o leva a reduzir tudo a algum reticulado mais simples conhecido (o reticulado booleano, por exemplo). A estratégia de adotar subjetivamente algum reticulado conhecido é a única possível para o caso extremo do conjunto dos atributos ser desordenado (por exemplo, para se avaliar um nome "bonito" dentro do conjunto de todos os nomes próprios).

A grande vantagem de se escolher um modelo matematicamente sofisticado para modelar os atributos do objeto avaliado é que as estruturas do tipo ordem total delimitada podem ser utilizadas para articular prescrições por meio de deduções lógicas. Do ponto de vista metodológico, esse é o caminho para se suprimir a intervenção subjetiva na avaliação e construir uma teoria para se explicar um evento (ação).

Vincular atributos e critérios às propriedades do reticulado booleano é requisito para se estabelecer relações causais entre diferentes eventos, uma vez que a estrutura lógica que permite deduções do tipo se..., então... é justamente a do reticulado booleano. Para se demonstrar isso com rigor, seriam necessárias algumas páginas de definições e um grau de formalismo que foge do objetivo deste capítulo.³⁷ O importante aqui é marcar a fronteira entre dois universos metodológicos, o prescritivo e o descritivo. Qualquer análise

³⁷ De modo muito esquemático, uma forma de apresentar as premissas da lógica clássica (lógica dedutiva booleana) é a partir de três axiomas: 1. Não-Contradição $\neg(\neg A \wedge A)$: significa que não é possível uma afirmação ser verdadeira e falsa ao mesmo tempo; 2. Terceiro Excluído $(\neg A \vee A)$: impede que haja qualquer outra valoração para uma afirmação que não seja "verdadeiro" ou "falso"; 3. Identidade $(x = x)$: postula que todo elemento é igual a si próprio (introduz o predicado "igual", indicado pelo símbolo $=$). Os símbolos $\neg$, $\wedge$ e $\vee$ significam respectivamente as relações de negação, simultaneidade (e) e alternativas não-excludentes (ou), as denominações "verdadeiro" e "falso" podem ser representadas pelos números 1 e 0, respectivamente, formando o conjunto $\{0, 1\}$. As relações definidas sobre tal conjunto binário constituem uma álgebra de Boole $(\{0, 1\}, \neg, \wedge, \vee)$ que, por sua vez, é um reticulado. Para se realizar operações lógicas, além da álgebra de Boole e os axiomas são preciso regras de transformação que permitem escrever novas sentenças (chamados teoremas). Essas regras são tais que conservam necessariamente o valor de verdade originalmente associado aos axiomas. Além disso, certas composições entre as relações $\neg$, $\wedge$ e $\vee$ podem dar origem a novas relações, sendo as mais importante aquelas abreviadas como $\Rightarrow$ (implica) e $\Leftrightarrow$ (se, e somente se). $A \Rightarrow B$ é logicamente equivalente a $\neg A \vee B$ e $A \Leftrightarrow B$ é logicamente equivalente a $(A \Rightarrow B) \wedge (B \Rightarrow A)$. A estrutura hipotético-dedutiva é proporcionada pelas regras, tais como a conhecida como *Modus Ponens*:

$$\begin{array}{c} A \\ A \Rightarrow B \\ \hline B \end{array}$$

Isto é, se a hipótese A e a relação causal $A \Rightarrow B$ são verdadeiras, então B também é verdadeiro. Para aprofundar nos conceitos e no instrumental da lógica booleana cf. MENDELSON (1987), SHOENFIELD (1967), TARSKI (1961), QUINE (1996). Para a caracterização da álgebra de Boole como um reticulado cf. BIRKHOFF (1967), MCKENZIE *et al.* (1987) e NGUYEN & WALKER (2000).

prescritiva precisa, em algum momento, modelar os critérios da avaliação como um reticulado booleano, ou seja, $Cr = \{0, 1\}$.

Essa constatação é fundamental, uma vez que qualquer estrutura mais frouxa associada aos atributos (por exemplo, conjuntos de estados desordenados ou arranjos em ordens parciais sem *sup* e *inf* definidos) impede a dedução prescritiva objetivada. Isso porque, nesses casos, não é possível por isomorfismo ou homomorfismo (preservação de ordem) vincular automaticamente à estrutura dos atributos valores de verdade ou falsidade de um critério booleano. Desse modo, fica sendo sempre necessária a intervenção do sujeito avaliador (decisor) para executar a função A de avaliação e produzir um juízo a partir dos atributos em questão.

Ao ser necessária a intervenção do indivíduo avaliador, perde-se o caráter dedutivo estrito que somente é possível se existir uma estrutura de ordem completa delimitada no conjunto de partida de um atributo que é transposta e preservada no conjunto de chegada *Cr*. Se o indivíduo precisar construir sua própria passagem para o reticulado booleano em *Cr* o processo torna-se idiossincrático e, nesse sentido, diferentes indivíduos podem apresentar diferentes "racionalidades" para concebê-lo. O salto entre as duas estruturas (atributo a critério) não precisa ser "lógico" no sentido convencional, a intuição ou mecanismos derivados de experiência e aprendizado tornam-se então mais plausíveis para explicar a avaliação dos indivíduos.

Porém, uma avaliação não sendo formalmente capaz de prescrição é mais dificilmente aceita intersubjetivamente.³⁸ Uma estrutura não prescritiva de um atributo é apenas, no máximo, uma estrutura descritiva. Isso porque o avaliador pode interpretar e descrever o atributo, mas não pode medi-lo em relação a todos seus estados em uma ordem completa e tampouco organizá-lo em um modelo de causa e efeito testável. O máximo que ele obtém são construções *ad hoc* para explicar eventos particulares já observados (uma vez que induzam algum tipo de relação capaz de associar parte dos elementos presentes no conjunto de estados do atributo).³⁹ O fato de tais descrições ou explicações não terem

³⁸ A não ser que o procedimento adotado pelo sujeito avaliador seja socialmente aceito. Conclusões prescritivas só são intersubjetivamente aceitas porque a dedução lógica é aceita intersubjetivamente, mas não se pode perder de vista que a aceitação da validade lógica é também uma construção social.

³⁹ Apenas nos casos em que essas relações se mostrem válidas repetidamente é possível construir teorias, em geral com a concomitante necessidade de reajustar conceitualmente o atributo observado para que o mesmo possa ser transposto para critério booleano segundo alguma heurística aceitável.

validade lógica não significa que não tenham validade para o indivíduo que a construiu. De fato, é a partir delas que ele executará e procurará legitimar suas ações.

XIV. MÉTODOS MULTICRITÉRIO DE APOIO À DECISÃO

A evolução dos métodos multicritério de apoio à decisão é um exemplo instrutivo para ilustrar o que foi discutido de modo abstrato na seção anterior.

A teoria e os métodos de decisão multicritério (ou multiatributo)⁴⁰ são conseqüências naturais das pesquisas de Borda e Condorcet. O problema abordado por esses métodos é a obtenção de uma decisão a partir de múltiplos pontos de vista sobre um mesmo objeto (multiatributos) ou múltiplos pontos de vista sobre um mesmo atributo (múltiplos indivíduos decidindo em conjunto, ou seja, o caso da escolha social). Do ponto de vista da estrutura matemática utilizada nas formulações originais, essas duas variações do problema são estruturalmente equivalentes. Assim, pode-se dizer que para um único indivíduo, cada diferente atributo de um objeto precisa ser avaliado e se converte em um critério para sua decisão. No caso de muitos indivíduos, o problema é formulado como se cada um avaliasse um atributo distinto (o que explicaria as diferentes opiniões) e cada avaliação individual se converteria em um critério para a decisão coletiva. O objetivo dos métodos se resume em fornecer uma ordenação síntese a partir das ordenações obtidas para cada critério em separado.

A motivação original dos métodos multicritério de decisão era explicitamente prescritiva (KENNEY & RAIFFA, 1976, p. 8), ou seja, pretendia-se fornecer ao tomador de decisão algum algoritmo que indicasse a melhor decisão, dada uma situação complexa que envolvesse múltiplos atributos com ordenações contraditórias, interdependências e interesses conflitantes.

A base conceitual para os métodos multicritério remonta à obra de von Neumann e Morgenstern (1953): Teoria dos Jogos e Comportamento Econômico. A teoria dos jogos fornece modelos para explicar as decisões de indivíduos em diferentes situações (cooperativas, competitivas, com informação completa ou incompleta, sob risco etc.). Os

⁴⁰ A denominação de multicritério ou de multiatributo é a rigor indiferente para os primeiros métodos multicritério, uma vez que atributos e a estrutura de preferências adotadas são ambas modeladas no reticulado [0, 1].

autores rejeitam a noção de que os indivíduos seriam entidades constantemente empenhadas em maximizar a utilidade de todas suas ações. Ao invés disso, argumentam que eles são capazes de comportamento estratégico, deixando de maximizar a utilidade de decisões presentes em troca de ganhos futuros.

Desse modo, os indivíduos não avaliariam simplesmente a partir da utilidade das alternativas disponíveis, eles raciocinariam em termos da *utilidade esperada* de uma ação, incluindo todos seus possíveis desdobramentos.

Há a substituição do atributo considerado no objeto sob avaliação. A noção de utilidade, isto é, a medida de prazer ou felicidade, associada a um elemento é substituída pela probabilidade da escolha desse elemento vir a gerar a maior utilidade.

Portanto, na teoria dos jogos, a medida $[0, 1]$ associada como atributo aos objetos não é uma medida de utilidade, mas sim uma medida v de probabilidade.⁴¹ Essa estrutura implica em avaliar as opções disponíveis como alternativas aditivas (*pay-offs*) sendo a situação de indiferença entre a e b descrita por $v(a) = v(b) = 0,5$. Em qualquer outra situação, por exemplo se a utilidade esperada para a for maior que para b , a desigualdade $v(a) > v(b)$ implica $a \blacktriangleright b$. Note que esse resultado pode valer mesmo que $u(a) \leq u(b)$, o atributo medido pela teoria dos jogos não é a utilidade u presente, é a probabilidade v da utilidade futura.⁴²

Há na definição de utilidade esperada uma ligação com a racionalidade decisória de Simon. Ambas pressupõem a existência de alternativas bem definidas e conhecidas de antemão e a projeção das conseqüências futuras dessas alternativas.

⁴¹ As medidas de probabilidade obedecem aos axiomas de Komolgorov, de 1933. Sendo (Ω, A, P) um espaço de probabilidade, tal que Ω é o conjunto de todos os eventos, A é um conjunto de subconjuntos de Ω e P é uma função do tipo $P: A \rightarrow [0, 1]$, valem os axiomas (DANTAS, 2000, p. 29):

① $P(A) \geq 0$ para todo $A \in \Omega$;

② Se $(A_n)_{n \geq 1}$ é uma sequência de eventos de Ω , que são mutuamente exclusivos, então

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$$

③ $P(\Omega) = 1$.

Uma propriedade importante que emerge desses axiomas é a *aditividade*, ou seja, dados a e b , se $P(a) = x$, $P(b) = (1 - x)$.

⁴² A aditividade $\sum v(x_i) = 1$, válida para a utilidade esperada, não vale necessariamente para a utilidade presente. Por isso, considera-se aceitável comparar as utilidades esperadas por diferentes indivíduos sobre um conjunto de alternativas, o que não é o caso para a utilidade presente.

A idéia básica dos métodos de decisão multicritério é modelar as conseqüências das diferentes escolhas possíveis em seus múltiplos níveis e *pay-offs*. Para isso, é construída uma hierarquia de objetivos a serem maximizados (Figura 2.1). No topo da hierarquia é colocado o objetivo principal **a**. Supõe-se que ele será atingido se objetivos intermediários **b1, b2,..., bn**, que compõem o primeiro, forem atingidos pela decisão. Um objetivo intermediário **bi** pode ser condicionado por outros objetivos intermediários **c1, c2,..., cn**, e assim por diante.

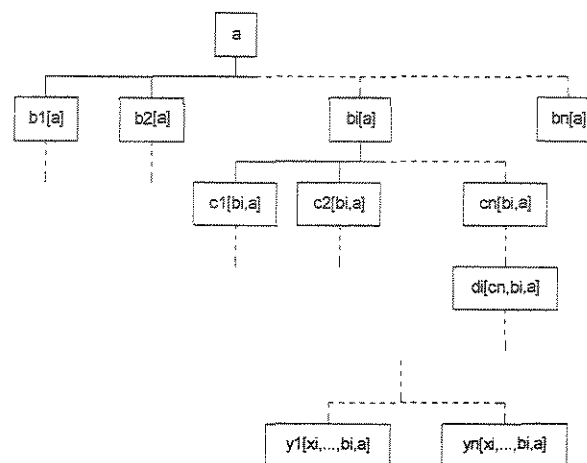


Figura 2.1: Esquema geral de hierarquização de objetivos para decisão multicritério.

Esse conjunto de objetivos modelados como *pay-offs* forma uma estrutura ramificada (também chamada hierarquia de objetivos) na qual cada nó (ponto de ramificação) define um espaço de probabilidades (Ω , A , P) aditivo.

Esse método é conhecido como Método de Utilidades Multiatributo (MAUT).⁴³ O algoritmo para calcular o impacto de uma alternativa a sobre cada objetivo da hierarquia consiste em avaliar qual a utilidade esperada da alternativa para cada objetivo mais desagregado e, em seguida, medir os *pay-offs* entre os objetivos concorrentes em cada ponto de ramificação. Para um nó qualquer x , composto pelos objetivos intermediários $y_1, y_2, \dots, y_n$ a utilidade esperada do objetivo x é dada por:

⁴³ Multi-Attributes UTilities. Cf. KEENEY & RAIFFA (1976), BELL, KEENEY & RAIFFA (1977), ROY & BOUYSSOU (1993), SAATY (1997), BARBA-ROMERO & POMEROL (1997), MARTÍNEZ & ESCUDEY (1998), ZUFFO (1998) e CLEMEN & REILLY (2001) para maiores detalhes sobre o método de utilidades multiatributo, suas variações e aplicações.

$$v(x) = \sum_{i=1}^n k_i v(y_i)$$

Os coeficiente k_i representa a utilidade esperada da comparação entre o i -ésimo objetivo intermediário com seus objetivos alternativos visando maximizar o objetivo imediatamente superior. Note que para esse ponto da estrutura de objetivos, o valor para cada $v(y_i)$ é uma utilidade esperada já calculada em um nível inferior. Por conta da propriedade aditiva, para cada espaço probabilístico definido pelos nós na hierarquia de objetivos, vale a igualdade:

$$\sum_{i=1}^n k_i = 1$$

Na base da hierarquia de objetivos, entram as utilidades esperadas para as alternativas disponíveis para a decisão segundo os objetivos mais desagregados ali presentes. A alternativa que maximizar a utilidade esperada do objetivo no topo da árvore de decisão é a melhor alternativa para a decisão.

Para que o modelo da decisão seja compatível com a realidade é preciso que ele obedeça a pelo menos três propriedades: exaustividade, não-redundância e coesão.⁴⁴

Os coeficientes k são também chamados pesos e o conjunto de todos os pesos pode ser interpretado como uma representação do sistema de preferências do decisor. Os pesos não se referem às alternativas avaliadas, mas sim às diversas composições de *pay-offs* utilizadas para compor o modelo arbóreo das conseqüências de cada escolha. A estrutura aditiva utilizada para a agregação das utilidades esperadas mantém os resultados para qualquer ponto da árvore dentro do intervalo $[0, 1]$ e é fácil verificar que, apesar de complicada, a estrutura concatenada de pesos preserva a todo momento as propriedades matemáticas originalmente definidas para os atributos de utilidade esperada segundo cada objetivo desagregado. A decisão final ocorre por meio de um critério único de síntese (o objetivo no topo da árvore).

Muitas críticas podem ser feitas ao método de utilidades multiatributo, especialmente pelo fato de ser extremamente difícil realizar na prática suas premissas de exaustividade, não-redundância e coesão. Em particular, a premissa da coesão implica em

⁴⁴ Exaustividade significa que o conjunto de objetivos é suficiente para captar todas as conseqüências da decisão. Não-redundância significa que não há sobreposição entre os objetivos descritos. Coesão significa que todas as medidas realizadas são homogêneas, isto é, podem ser comparadas entre si.

dificuldades profundas, relacionadas com as estruturas formais que estão na base do método.

Keeny e Raiffa (1976, p. 255) propuseram variações metodológicas para lidar com o problema da não-redundância, modelando objetivos inter-dependentes por meio de funções de agregação multiplicativas ou paramétricas e sugerindo modificar os objetivos problemáticos.

Esses mesmos autores são cautelosos em afirmar que o método de utilidades multiatributos possa ser utilizado para processos de decisão com mais de um decisor (escolha social), uma vez que conhecem os resultados de Arrow (KEENEY & RAIFFA, 1976, p. 2p. 523). Contudo, assumem que o método pode eventualmente ser o mediador de um procedimento de escolha social. O argumento utilizado é que a comparação interpessoal não se dá entre utilidades, mas entre expectativas probabilísticas de utilidade e que a formalização (mesmo que necessariamente incompleta e imperfeita) traz uma série de vantagens, entre elas o fato de que "a especificação detalhada dos objetivos e dos atributos associados pode levar a uma articulação mais clara sobre os elementos essenciais do problema e a um melhor entendimento por parte dos indivíduos" e de que "pode ajudar a decompor em partes um problema geral e facilitar a comunicação efetiva entre os membros de um grupo". Além disso, os autores adiantam a interpretação de que o método pode ser aplicado a um grupo para que o tomador de decisão conheça em profundidade as estruturas de preferência dos integrantes do grupo, em que termos divergem ou concordam, e possa a partir disso construir processos de resolução de conflitos e tomar decisões informadas sobre temas que afetam o grupo (p. 544).

Com o objetivo de tentar tratar o problema da coesão e, por extensão, da comparabilidade entre utilidades esperadas para diferentes objetivos ou por diferentes indivíduos, novas axiomatizações para as estruturas de preferência surgiram a partir dos anos 1970. A chamada escola francesa de métodos multicritério ocupa papel de destaque nesse contexto. Em 1968, Bernard Roy afirmava a necessidade de se "desotimizar os métodos multicritério".⁴⁵ Em um artigo de 1977, esse mesmo autor introduz o conceito de *superação* para modelar as preferências.

⁴⁵ Cf. BARBA-ROMERO & POMEROL, 1997 (p. 10).

Os métodos baseados na utilidade esperada usam uma estrutura de preferências que pode ser descrita por apenas dois tipos de relação entre qualquer par de alternativas a e b : indiferença (I) ou preferência (P).⁴⁶ Já os métodos fundamentados no conceito de superação lançam mão de um outro esquema de relações para modelar as preferências.

Roy (1977, p. 42) define uma estrutura de *comparabilidade parcial*, na qual as preferências são modeladas a partir de quatro relações binárias I, P, Q e R segundo as seguintes propriedades:

I (indiferença): reflexiva e simétrica

P (preferência estrita): irreflexiva e antissimétrica

Q (preferência fraca): irreflexiva e antissimétrica

R (incomparabilidade): irreflexiva e simétrica

Por definição, diz-se que a supera b (escreve-se aSb) quando vale a relação de preferência estrita, de preferência fraca ou de indiferença. Essa relação é inspirada numa afirmação de Condorcet de 1785: “Quando uma alternativa A é ao menos tão boa quanto outra B em uma maioria de critérios, e não há nenhum critério em que A seja notoriamente inferior a B, podemos afirmar sem risco que A supera B” (BARBA-ROMERO & POMEROL, 1997, p. 206).

Nenhuma dessas relações é transitiva. A ausência de transitividade traz uma série de conseqüências importantes. Por exemplo, sem a transitividade não se pode afirmar que entre duas alternativas a e b ou vale aPb ou vale $\neg(aPb)$, como se podia afirmar na axiomatização convencional.⁴⁷ No caso da comparabilidade parcial, podem ocorrer situações em que mesmo que não se possa afirmar que a é preferível a b , também não se pode afirmar que não é preferível. É justamente nessa “zona cinzenta” que se aplica a relação Q de preferência fraca.

⁴⁶ As relações P e I são derivadas da relação $\blacktriangleright$. A relação P é irreflexiva, anti-simétrica e transitiva. A relação I é reflexiva, simétrica e transitiva. A relação $\blacktriangleright$ é reflexiva, anti-simétrica e transitiva. Há isomorfismo entre as relações de preferência ($\blacktriangleright$, P, I) e as relações ($\geq$, $>$, $=$) quando aplicadas ao intervalo $[0, 1]$ dos números reais.

⁴⁷ Para se demonstrar que ou aPb ou $\neg(aPb)$ valem necessariamente é preciso a propriedade transitiva. Segue a demonstração:

- | | |
|-------------------------------|------------------------------------|
| 1. $\neg(aPa)$ | (assimetria) |
| 2. $\neg(aPb \wedge bPa)$ | (transitividade) |
| 3. $\neg(aPb) \vee \neg(bPa)$ | (decomposição) |
| 4. $\neg(aPb) \vee aPb$ | (antissimetria) |
| 5. $aPb \vee \neg(aPb)$ | (propriedade do conectivo $\vee$) |

Dados um atributo j descrito por uma escala qualquer (totalmente ordenada) e duas alternativas a e b , as definições necessárias para modelar a transição entre a indiferença e preferência forte, passando pela preferência fraca, são representadas por funções especiais que definem certos intervalos sobre a escala do atributo. Elas são o limiar de preferência ($\mathbf{p}$) e o limiar de indiferença ($\mathbf{q}$).

Na comparabilidade parcial, a escala que mede os atributos não é necessariamente a utilidade esperada. Qualquer escala E completamente ordenada pode ser usada. Desse modo, para cada atributo, ocorre uma avaliação independente que induz uma ordem completa em um critério diretamente associado ao atributo. Se a diferença entre o valor medido para b (representado por $g(b)$) e a ($g(a)$) for maior que um certo intervalo definido por uma função $\mathbf{p}: g(a) \rightarrow E$, então vale a relação bPa . Para o atributo j , isso significa que:

$$g_j(b) - g_j(a) > \mathbf{p}_j[g_j(a)] \Rightarrow (bP_ja)$$

Se, por outro lado, a diferença for menor que um outro intervalo definido por $\mathbf{q}: g(a) \rightarrow E$, tal que $\mathbf{q} \subseteq \mathbf{p}$, vale a relação de indiferença.

$$0 \leq g_j(b) - g_j(a) \leq \mathbf{q}_j[g_j(a)] \Rightarrow (bI_ja)$$

O intervalo entre os intervalos definidos por $\mathbf{p}$ e $\mathbf{q}$ define a região de preferência fraca.

$$\mathbf{q}_j[g_j(a)] < g_j(b) - g_j(a) \leq \mathbf{p}_j[g_j(a)] \Rightarrow (bQ_ja)$$

Como não há transitividade, comparar a com b pode ser diferente de comparar b com a . Para comparar a com b os limiares de preferência e indiferença precisam ser definidos em função de b .

Também é fácil notar que se os intervalos definidos por $\mathbf{p}$ e $\mathbf{q}$ forem vazios (nulos) apenas as relações I e P permanecem, com comportamento similar à situação convencional de comparabilidade total.

Roy chamou as estruturas de preferência que contém os limiares de indiferença e de preferência não-nulos de *pseudocritérios*. Em contraste com os critérios verdadeiros

induzidos pelas utilidades multiatributo que, devido à premissa de coesão, obedecem homogeneamente todas as propriedades matemáticas do intervalo real $[0, 1]$ e podem ser agregados diretamente, nos pseudocritérios, cada atributo induz uma estrutura de preferência própria, com mais ou menos "zonas de sombra" (dependendo de como o decisor o interprete) e, portanto, não podem ser agregadas diretamente e tampouco arranjadas em estruturas de árvore.

Para a agregação de n critérios, quantifica-se em quantos critérios a supera b (aSb) em relação a todos os critérios. Como, para o decisor, os critérios podem ter importância relativa diferenciada, eles podem ser ponderados por constantes $k_1, k_2, \dots, k_n$ arbitrárias. De modo geral, a avaliação global da hipótese aSb é dada por uma medida de credibilidade $c(a, b)$ do tipo:

$$c(a, b) = \frac{\sum_{j: aSb} k_j}{\sum_{j=1}^n k_j}$$

Não é o caso de detalhar as possibilidades de variação do algoritmo geral, pois podem ser bastante complicadas. No Anexo 1, o algoritmo de um método (ELECTRE III) que utiliza essas estruturas de preferência e de agregação multicritério é detalhado e comentado em todos seus pormenores.

Para os objetivos desta seção, o ponto a ser enfatizado é outro. O que precisa ser comentado são as consequências práticas que a mudança operada nas premissas sobre as propriedades matemáticas obedecidas provoca.

Do ponto de vista cognitivo, a criação de uma zona cinzenta entre as noções de indiferença e preferência significa para o decisor algo como o juízo "eu não sei". Do ponto de vista lógico, uma valoração do tipo "não sei" fere o princípio do terceiro excluído. Entretanto, do ponto de vista algébrico, ainda é possível falar de uma ordem total entre as valorações "falso", "não sei" e "verdadeiro" e, portanto, é possível formular estruturas lógicas (e de decisão) com esse tipo de conjunto de chegada. Lógicas com mais de 2 valores de chegada são chamadas de *modais* (existem, portanto, infinitas lógicas modais). Em

especial, a lógica modal que utiliza como conjunto de chegada o intervalo real $[0, 1]$ é chamada lógica nebulosa (*fuzzy*).⁴⁸

Um pseudocritério é isomórfico a um conjunto nebuloso. Conjunto nebuloso é todo subconjunto A do conjunto U que pode ser representado por uma função $\mu_A: U \rightarrow [0, 1]$, chamada função de pertinência. Em contraste, num conjunto nítido (*crisp*) uma função de pertinência é do tipo booleana $\mu_A: U \rightarrow \{0, 1\}$, ou seja, o conjunto A pertence ou não pertence a U , sem possibilidade de situações intermediárias.

Por definição, para qualquer conjunto nebuloso, o comportamento da função μ_A é definido subjetivamente (NGUYEN & WALKER, 2000). Não há um método ótimo (ou objetivamente correto) para modelar uma função de pertinência.

Nos métodos multicritério baseados em comparabilidade parcial isso fica evidente pela necessidade do decisor precisar escolher subjetivamente os intervalos de indiferença e preferência segundo cada atributo, sua escala e os coeficientes k .

Em contraste com os métodos que se baseiam em utilidades esperadas, que procuram objetivar ao máximo a avaliação dos atributos e dos objetivos intermediários, nos métodos de comparabilidade parcial muita coisa é deliberadamente definida de modo subjetivo. Além dos parâmetros dos pseudocritérios, seus pesos relativos também são arbitrários. Nos algoritmos para agregação dos critérios outros parâmetros arbitrários são ainda encontrados (vide Anexo 1) e a própria escolha do método de agregação pode ser arbitrária, privilegiando uma ou outra característica.⁴⁹ Enquanto no método de utilidade multiatributo o resultado final é uma ordenação completa e cardinal (isto é, com a indicação das distâncias entre cada par de alternativas segundo a quantidade de preferência associada), nos métodos de comparabilidade parcial o resultado agregado é um conjunto parcialmente ordenado, sem quaisquer medidas cardinais e contendo relações de incomparabilidade (R).

⁴⁸ A primeira formalização algébrica de uma lógica modal foi a Lógica $(0, u, 1)$ ou Lógica de Lukasiewicz, devida a esse mesmo autor em 1920 (CIGNOLI *et al.*, 1995). A lógica nebulosa foi formalizada por Zadeh, em 1965, inicialmente por meio do tratamento de conjuntos nebulosos (ZADEH *et al.*, 1975). As propriedades dedutivas associadas a ambas (cálculo proposicional) são as mesmas (NGUYEN & WALKER, 2000, p. 68).

⁴⁹ Na lógica nebulosa, os conectivos lógicos convencionais ($\neg$, $\wedge$, $\vee$, $\Rightarrow$) podem ser redefinidos arbitrariamente em novas famílias e outros ainda podem ser criados, dependendo das características que o sujeito deseje ressaltar. A noção clássica de “e” é substituída por t-normas (Δ) e a de “ou” por t-conormas (∇). Em cada família de conectivos nebulosos, apesar de ser definida a partir de propriedades subjetivamente orientadas, t-norma, t-conorma e negação são interdependentes e precisam ser coerentes entre si.

Toda essa discussão pode parecer demasiado técnica, mas é necessária para realçar os fatos estruturais que estão por trás de uma mudança fundamental de perspectiva. Os métodos de comparabilidade parcial não se pretendem prescritivos e não se espera que eles objetivem o processo que leva à decisão. Pelo contrário, são chamados de métodos de *apoio* à decisão porque procuram uma descrição acurada que auxilie o decisor a compreender em maior profundidade o problema a ele colocado, mas não retiram do indivíduo a prerrogativa da decisão.

A avaliação de objetos sociais pode perfeitamente percorrer esse caminho. É preciso, para isso, abrandar premissas formais. Isso, como foi visto, leva necessariamente à inclusão do subjetivo e permite muitas variações de método, em contraste com a abordagem prescritiva da racionalidade científica objetivada.

XV. SOBRE O SIGNIFICADO DE CIENTIFICIDADE EM AVALIAÇÕES

O exemplo trabalhado na seção anterior é apenas um entre outros possíveis. Por meio dele, procurou-se chamar a atenção para a importância que as estruturas formais básicas possuem na construção de métodos para avaliação (entendendo que o apoio à decisão é na realidade um processo de avaliação).

Vale a generalização de que toda explicação científica objetivada é ancorada em estruturas matemáticas da lógica convencional, isto é, no reticulado booleano. No contexto do que foi discutido no início deste capítulo, dizer que uma avaliação é objetivada é dizer que ela está construída sobre certas estruturas matemáticas que permitem trabalhar atributos e critérios como estruturas similares e operacionalmente fortes, isto é, gozando de propriedades suficientemente sofisticadas e auto reproduzíveis para que se possa operá-las sem intervenção subjetiva.

Entretanto, se por um lado essas estruturas criam facilidades e permitem avançar conclusões sobre objetos reais seguindo os passos da abstração, por outro também impõem regras e limites que são incapazes de acomodar importantes características de certos objetos. Como foi discutido no decorrer deste capítulo, isso vale particularmente para os objetos sociais.

A indeterminação subjetiva, a capacidade de aprendizado e mudança auto determinada dos indivíduos, a dualidade da condição concomitante de avaliador e objeto e a possibilidade de múltiplas racionalidades (ou sistemas de preferência) são características que escapam da capacidade de operação dos instrumentos matemáticos convencionais.

Entretanto, como mostra o exemplo dos métodos multicritério de comparabilidade parcial, o fato dos instrumentos matemáticos convencionais criarem problemas não configura um limite intransponível. É sempre possível a criação de outros instrumentos matemáticos. "A essência da matemática radica na sua completa liberdade".⁵⁰ A solução encontrada por Roy e colaboradores foi alterar as propriedades das relações aceitas como constituintes das estruturas de preferência dos indivíduos. Como resultado, alcançaram uma modelagem muito mais realista para situações práticas que eram mal modeladas pela abordagem de utilidade esperada (por exemplo, a possibilidade de hesitação por parte do avaliador entre as situações de indiferença e de preferência). Por outro lado, a estrutura formal é mais geral, necessita de intervenção subjetiva e pode assumir diferentes características de inferência. Em outras palavras, muda a lógica subjacente da interpretação. As propriedades formais que decorrem das premissas são, para esse caso, similares à lógica nebulosa.

Outros exemplos poderiam ainda ser citados. Em primeiro lugar, e ainda na linha de investigação sobre estruturas de preferência, pode ser mencionada a técnica de Conjuntos Aproximativos (Rough Sets) introduzida por Greco e colaboradores (2001, 2002) para construir regras do tipo se..., então... que representem a estrutura de preferência de um decisor a partir de um conjunto de exemplos de referência é um exemplo de abordagem especificamente desenhada para lidar com conhecimento parcial e situações de ambigüidade.

Outras abordagens alternativas hoje se desenvolvem na tentativa de modelar a incerteza por outros caminhos. A forma convencional de representar a incerteza é por meio de funções de probabilidade. A noção de probabilidade, originalmente concebida com característica de crença subjetiva por Bayes (1764), passou ao final do século XIX a ser rigidamente formalizada e objetivada, principalmente por meio dos trabalhos Gauss e

⁵⁰ Frase de Cantor, matemático alemão do final do século XIX e início do XX e principal sistematizador da teoria dos conjuntos. A época de Cantor marcou o início da tradição lógico-matemática fundamentada em construções puramente abstratas e abertas a infinitas variações (DA COSTA, 1992, p. 62).

Kolmogorov. A interpretação objetiva de distribuições de probabilidades se tornou dominante e influenciou, por exemplo, quase toda a tradição econométrica do século XX.

Apenas mais recentemente, novos desenvolvimentos metodológicos explicitamente fundamentados na abordagem bayesiana têm ganhado espaço.⁵¹ Entretanto, os métodos bayesianos, apesar de incorporarem diretamente avaliações subjetivas de probabilidades, ainda consideram aditivo o espaço dos eventos, ou seja, o postulado $P(\Omega) = 1$ vale.

Por sua vez, a teoria da *possibilidade* abre mão dessa premissa (NGUYEN & WALKER, 2000). Isso significa que nela é válido aceitar (e novamente a lógica nebulosa fornece a estrutura necessária), por exemplo, situações em que dois eventos mutuamente excludentes a e b tenham, por avaliação subjetiva, possibilidades $\pi(a) = 0,8$ e $\pi(b) = 0,4$.

As noções de aditividade e de eventos aleatórios não são as únicas possíveis para modelar a incerteza. Elas podem ser bastante eficientes para lidar com experimentos, mas sua estrutura matemática é insatisfatória para certas peculiaridades dos objetos sociais. O que os novos desenvolvimentos metodológicos apontam é que é possível distinguir diferentes tipos de incerteza. A ambigüidade, a imprecisão semântica, a hesitação e a contradição⁵² são características humanas e não podem ser tratadas apenas como erros aleatórios.

Explorar as possíveis conseqüências conceituais e metodológicas suscitadas por esses exemplos consumiria mais algumas dezenas de páginas e fugiria do objetivo deste capítulo. Além disso, esses desenvolvimentos são relativamente recentes e seus resultados são ainda em geral preliminares. O importante a destacar com essa rápida listagem de exemplos é que há de fato um movimento de mudança epistemológica nos fundamentos dos métodos, com mudanças profundas de perspectiva, e que precisam ser compreendidas.

E a questão importante nesse sentido é até que ponto os instrumentos construídos sobre essas bases não convencionais são ainda ciência.

Se, por um lado, os métodos concebidos sobre tais bases carregam alto grau de abstração e formalismo (vide o exemplo do Anexo 1), por outro, eles são freqüentemente incapazes de previsão e prescrição. Se o critério para ser ciência significar rigor e escrutínio

⁵¹ Modelos de aprendizado autônomo (*machine learning*) que empregam premissas bayesianas e aplicações a jogos e a modelos de otimização são alguns dos exemplos mais comuns.

⁵² Especificamente para modelar situações contraditórias simultaneamente válidas, portanto que violam o princípio da não-contradição da lógica convencional, existem as lógicas paraconsistentes (DA COSTA *et al.*, 1998).

a resposta é sim, são ciência e de alta qualidade. Se significar capacidade de previsão e repetibilidade a resposta é não, definitivamente essa característica se perde.

Além disso, a tensão entre objetivo e subjetivo se recoloca. Nesse ponto, é preciso ter clareza para refutar algumas conclusões precipitadas. A perda da capacidade de previsão e repetibilidade não é decorrência da displicência subjetiva. Essas propriedades estão intimamente relacionadas com a estrutura lógica booleana. O problema real é que a lógica booleana é que é insuficiente para o objeto social. Não se pode abstrair certas características subjetivas a partir de sua estrutura. Ora, ao se mudar a estrutura lógica é efeito simultâneo perder capacidade de previsão e incorporar o subjetivo. A causa da perda da capacidade prescritiva está na eliminação da separação artificial entre sujeito e objeto provocada pela objetivação da função avaliação, com atributos e critérios vinculados ao reticulado booleano. A flexibilização na modelagem dessa função – ou apenas o fato de tratar os critérios como reticulados diferentes do booleano – altera a essência dedutiva da capacidade prescritiva, ao mesmo tempo que recupera a necessidade da intervenção subjetiva explícita na relação com o objeto avaliado.

Outro cuidado que é preciso ter é não utilizar o princípio do terceiro excluído para interpretar essa tensão. Ou seja, concluir que pelo fato dos métodos não convencionais serem, por definição, incapazes de previsão e repetibilidade então eles seriam triviais, isto é, levariam necessariamente a "qualquer coisa" e, portanto não seriam úteis. No caso, pode ser vantajoso substituir conceitualmente a característica desejável de repetibilidade pela característica mais branda de estabilidade procedural e capacidade de reaplicação. Um método precisa ser bem definido para que se possa aplicá-lo mais de uma vez e ocorra aprendizado sistemático.

A chave para interpretar o papel dos métodos (convencionais ou não) numa avaliação é sua característica de construção social, com a função de legitimar socialmente as decisões dos indivíduos.

Para Newton da Costa, o conhecimento científico é "crença verdadeira e justificada" (DA COSTA, 1999, p. 27). Essa é uma boa definição porque concatena duas características decisivas para a cientificidade de um método. Ele precisa ser socialmente aceito, além de conceitualmente e formalmente consistente. A "verdade" que emerge de uma atividade científica precisa, no mínimo, dessas características.

Historicamente, muitas variações de método foram concebidas pela ciência, refletindo o aprendizado social provocado direta ou indiretamente pela própria ciência. A diferença para as novas construções metodológicas não convencionais é que elas se prestam justamente para as atividades de avaliação de objetos sociais e têm como consequência direta o aprendizado (tornando-se essas suas reais condições de validade).

Além disso, e em parte também por isso, o tipo de conhecimento que emerge de uma avaliação mediada por um método não prescritivo é intrinsecamente dependente das pessoas envolvidas e, portanto, indissociável do processo de aprendizado e legitimação social que pode engendrar.

Sendo assim, o caminho para escapar das limitações impostas pelo tipo de racionalidade científica que procura a objetivação no tratamento de objetos sociais é justamente dotar os métodos de avaliação de intenção explícita de provocar aprendizado.

Esse caminho leva a pelo menos duas importantes consequências. Primeiro, qualquer método assim concebido é uma extensão que generaliza métodos convencionais, porque usa estruturas matemáticas mais gerais. Segundo, a avaliação coletiva ganha um novo sentido e importância destacada, porque o aprendizado social se torna o efetivo critério de validade para os métodos.

Essas duas consequências abrem frentes importantes de investigação. No primeiro caso, novos procedimentos podem ser estabelecidos para concatenar diversos métodos convencionais que operam sobre bases conceituais diferenciadas e que são incompatíveis entre si em relação a suas próprias estruturas, mas que podem ser acomodados por uma estrutura não convencional, porém mais geral. Por exemplo, a incompatibilidade de métodos econométricos que transformam todas as variáveis de um problema a unidades monetárias com qualquer outro que trabalhe com variáveis não monetárias. Os resultados prescritivos desses métodos podem ser internalizados como atributos de uma avaliação mais geral que não opere de modo prescritivo.

A segunda consequência decorre do fato do aprendizado assumir lugar de destaque, se tornando objetivo e não apenas consequência do método. A legitimação não seria algo dado *a priori* por uma estrutura de validade lógica, cabendo ao método apenas demonstrar quais hipóteses seriam verdadeiras e quais seriam falsas em relação a um conjunto de fatos dados ou esperados. Mais que isso, a legitimação passaria pela capacidade de construir

socialmente uma interpretação válida para um conjunto de fatos dados ou esperados. Desta feita, a participação coletiva passaria a ser fundamental, quanto mais participação mais validade adquiriria uma avaliação. Nesse contexto, os problemas já discutidos sobre mecanismos de participação e de estruturas metodológicas para viabilizar o aprendizado social se tornam temas de renovada importância para investigação.

Esta seção encerra a discussão sobre os aspectos formais subjacentes à avaliações e suas relações com o método científico. Concluiu-se aqui que é possível e necessária uma mudança de perspectiva quanto ao uso de bases formais não convencionais para a construção de métodos e metodologias de avaliação.

No próximo capítulo, a argumentação muda de enfoque. Como assumimos que todo método de avaliação é uma construção social, é de se esperar que mudanças também ocorreram e estão ocorrendo na sociedade, de modo que esses novos procedimentos sejam aceitáveis ou mesmo necessários socialmente.

O Capítulo 3 examina algumas recentes mudanças sociais úteis para justificar a pertinência da renovação metodológica preconizada neste capítulo. A discussão é conduzida a partir do estudo da evolução recente das práticas de avaliação em relação a um objeto social específico: as atividades de produção de ciência e tecnologia. Nos últimos anos, houve uma importante renovação conceitual em relação ao significado e à pertinência de se avaliar ciência e tecnologia. Essa percepção inaugurou uma série de novos desafios metodológicos, ainda apenas precariamente solucionados.

Parte II – Avaliação na C&T em Rede e na Coordenação da Inovação

Avaliação da Ciência e da Tecnologia

Capítulo 3 – O contexto de legitimação dos principais métodos de avaliação das atividades de ciência e tecnologia

A clash of doctrines is not a disaster – it is an opportunity

Alfred North Whitehead

O objetivo deste capítulo é discutir as trajetórias metodológicas da avaliação orientada para as atividades de ciência e tecnologia (C&T). Essas trajetórias são inicialmente decorrentes de avanços conceituais acerca da interpretação do papel da C&T na dinâmica geral de desenvolvimento da sociedade, mas também são frutos de mudanças organizacionais sofridas pelos atores ligados à P&D e pela difusão de valores culturais associados a expectativas sociais de participação, aprendizado compartilhado e controle sobre o desenvolvimento tecnológico.

Para orientar a discussão das características dos diferentes tipos de avaliação de C&T existentes, até chegar no contexto das demandas atuais, optou-se por dividir o universo da C&T em três tipos institucionais. A idéia é realçar as características institucionais, no sentido amplo adotado no Capítulo 1 de fundamentos culturais, cognitivos, normativos e regulatórios, que regem diferentes configurações das atividades de C&T. Para cada um dos três tipos, diferentes métodos de avaliação foram desenvolvidos e ali encontram legitimidade. Esse eixo de discussão vai também evidenciar que partes substantivas das críticas e limitações dos diferentes métodos de avaliação são antes conflitos entre diferentes formas institucionais da C&T do que reais problemas dos métodos em si. Os problemas aparecem quando se espera resultados de um determinado método que ele não pode dar, simplesmente porque ele foi concebido em um outro contexto e para cumprir uma outra função.

Os três diferentes tipos institucionais funcionam como estereótipos que dão o pano de fundo para reforçar a idéia de que as formas das práticas de avaliação das atividades de ciência e tecnologia e os debates associados ao alcance e a legitimidade de seus métodos estão ligados a valores e objetivos bastante específicos. As mudanças institucionais que as

atividades de C&T sofreram na história ajudam na compreensão das motivações subjacentes aos desenvolvimentos metodológicos em avaliação de cada fase.

O capítulo se inicia com a caracterização dessas três formas institucionais das atividades de C&T. Elas constituem os moldes gerais sobre os quais a C&T se desenvolveu até o presente; são formas que *grosso modo* caracterizam épocas distintas, mas que de fato se acumularam durante o processo de evolução da C&T, produzindo também formas híbridas. A Pequena Ciência e Tecnologia, a Grande Ciência e Tecnologia e a Ciência e Tecnologia em Rede, são tomados como eixos condutores para a descrição de diferentes métodos de avaliação, concebidos para seus respectivos contextos. Em meio às breves introduções aos métodos, são apresentadas também algumas interpretações acerca das tendências da organização social em função do aprofundamento progressivo das interações entre sociedade, conhecimento e tecnologia. Esse panorama é a base para o delineamento do contexto institucional contemporâneo e das demandas que são atualmente colocadas para a C&T. Os desafios e as implicações para o desenvolvimento metodológico da avaliação de C&T nesse contexto apresentam particularidades e necessidades diferenciadas, não contempladas anteriormente.

I. TRÊS DIFERENTES FORMAS INSTITUCIONAIS DAS ATIVIDADES DE C&T

Esta seção é dedicada à descrição das três formas institucionais gerais escolhidas para caracterizar neste capítulo as atividades de ciência e tecnologia, em diferentes momentos de sua história. Apesar de cada uma delas ser típica de um determinado momento histórico, não há uma correspondência biunívoca entre forma institucional e desenvolvimento cronológico das formas de organizar a produção de conhecimentos científicos e tecnológicos. Dada uma forma institucional predominante, ela não exclui a coexistência das demais. Ou seja, as atividades de C&T de hoje preservam elementos institucionais das formas anteriores. Do mesmo modo, na releitura de épocas passadas é possível identificar de modo embrionário formas precursoras daquelas que depois predominaram. Há, sem dúvida, certa correspondência entre o aparecimento de novos métodos de investigação e a consolidação de novos arranjos institucionais de legitimação, entre a ênfase na produção teórica ou aplicada e as formas organizacionais que subjazem à

pesquisa; e entre os critérios e métodos de avaliação empregados para julgar os resultados obtidos e os interesses dos grupos sociais que financiam essas atividades. A produção social de conhecimentos científicos e tecnológicos se reproduz e se aperfeiçoa por meio do detalhamento das linhas de investigação socialmente aceitas (ciência normal), mas também por meio do embate provocado por interpretações conflitantes, que mesmo que se apresentem a si próprias como mais "gerais" (portanto preservadoras dos resultados válidos até então), acabam por solapar as bases mais profundas de legitimação das explicações anteriores (KUHN, 2001). Isso vale também na evolução das formas institucionais e organizacionais da ciência e da tecnologia. Cada uma representa uma espécie de novo paradigma que "generaliza" o anterior. Efetivamente, após cada salto, elementos anteriormente válidos permanecem, mas adquirem novo significado devido a convivência (ou sobrevivência) com um novo conjunto de regras.

A intenção de examinar as atividades de C&T a partir de uma óptica institucional é evidenciar os elementos que constituem sua justificação social em diferentes contextos históricos e organizacionais. A definição daquilo que se entende por produção de conhecimento científico e tecnológico é uma tarefa socialmente relativa, isto é, depende do entendimento de quais são as regras procedurais e de quais são os critérios de validade socialmente aceitos para enquadrar determinados trabalhos como científicos ou tecnológicos e legitimar o grupo social que os realiza. Disso decorre que as próprias definições sobre quais seriam produtos científicos e quais seriam produtos tecnológicos não podem ser definições absolutas. É preciso ter em conta o contexto histórico e cultural ao se tentar isolar as atividades de C&T entre si e das demais atividades de uma sociedade. Na prática, entretanto, é preciso ter clareza de que as atividades de C&T são em grande medida indissociáveis de processos sociais mais amplos (MAYR, 1982) e que os recortes conceituais para isola-las como objeto de estudo são necessariamente artificiais. A divisão proposta neste capítulo em três grandes grupos de métodos, cada um ligado a um contexto institucional não se pretende, portanto, absoluta. Trata-se antes de um apoio conceitual, uma tipologia construída para facilitar a localização dos diferentes métodos segundo seus objetivos e limitações e, especialmente, localizar claramente a proposta metodológica que será detalhadamente apresentada no próximo capítulo.

Um marco fundamental na história da organização social da ciência e da tecnologia – e que cumpriu importante papel na percepção de que a ciência pode ser produzida segundo distintas lógicas institucionais – foi o advento da chamada Grande Ciência (Big Science).

O termo Grande Ciência foi cunhado por Weinberg (1961 e 1962) e popularizado pelo estudo quantitativo a respeito do crescimento da ciência realizado por Price em 1963 (PRICE, 1976). A noção de Grande Ciência marcaria a distinção entre a pesquisa como empreitada coletiva, fartamente financiada por governos e empresas e explicitamente organizada para atingir metas e cumprir objetivos desses financiadores, da Pequena Ciência, caracterizada como a pesquisa típica do século XIX, individual, desinteressada e em regime de mecenato. A Grande Ciência representaria um novo estágio de maturidade da organização da ciência, e das atividades tecnológicas baseadas em ciência, mas também refletiria uma mudança institucional fundamental.

Mais do que uma distinção entre épocas – apesar desta valer no plano geral⁵³ – a diferença entre a Pequena e a Grande Ciência está em sua organização e nos valores e motivações a elas subjacentes.

O que se convencionou aqui por Pequena C&T inclui toda a produção de ciência e tecnologia que contrasta com a Grande C&T em termos de mobilização de recursos financeiros e humanos e, especialmente, em termos de objetivos. A inclusão deliberada do T, de tecnologia, amplia o escopo da terminologia original de Weinberg e Price. Essa extensão encontra relativamente poucos problemas para o caso da Grande C&T, uma vez que os grandes projetos da Grande Ciência quase sempre envolviam também objetivos tecnológicos. No entanto, no caso da Pequena C&T é preciso cautela. Antes do final do século XIX, e do advento da Segunda Revolução Industrial, ciência e tecnologia eram universos desconectados que operavam segundo lógicas independentes e particulares. Ambas, entretanto, possuíam o traço comum de serem atividades promovidas pelo talento e

⁵³ Desde muito tempo os Estados já investiam em pesquisa aplicada e grandes projetos de engenharia organizados para solucionar problemas tecnológicos específicos. Por exemplo, a Escola de Sagres em Portugal do século XV que viabilizou as grandes navegações, as pesquisas em vinicultura coordenadas por Pasteur na França do século XIX e a criação de institutos de pesquisa como a Imperial Estação Agronômica de Campinas, em 1887, destinado a fornecer tecnologia para os produtos agrícolas exportados pelo Brasil. No entanto, é a partir da II Guerra Mundial que o investimento estatal em P&D se intensifica como fenômeno mundial.

esforço de indivíduos ou pequenos grupos, sem o aporte de recursos em escala industrial característico do período posterior.

No período entre a Segunda Revolução Industrial e a Segunda Guerra Mundial, a pesquisa em tecnologia, a partir desse ponto baseada em ciência, passou a ser promovida em escalas cada vez mais ambiciosas por empresas químicas e eletromecânicas principalmente.

A pesquisa científica em grande escala financiada pelo Estado é característica do período subsequente à Segunda Guerra Mundial. Grande C&T, portanto, já existia muito antes da terminologia. Além disso, é importante deixar claro que esses tipos gerais não significam que não existam situações híbridas. Laboratórios que podem ser muito bem caracterizados como Pequena C&T existem até hoje em universidades e até em empresas, muitas vezes financiados pelo Estado e caracterizados por um comportamento ofertista. O custo de incorrer em alguma imprecisão pela generalidade das classificações de Pequena C&T e Grande C&T é compensado pela possibilidade de se usar um quadro de referência simples e suficiente para orientar o exame dos métodos de avaliação que interessa a este capítulo.

A Pequena Ciência pode ser muito bem caracterizada em seus valores segundo o esquema formulado por Merton em 1942. Para Merton, compondo o *ethos* da ciência há um conjunto de valores que se constitui em verdadeira obrigação moral para o cientista; mesmo que não codificado, ele permeia os usos e costumes dos envolvidos na prática científica e a indignação suscitada quando o mesmo é desafiado ou contrariado (MERTON, 1979). A pesquisa em tecnologia num contexto puramente mertoniano é em geral lida como ciência degenerada, de má qualidade ou a serviço de interesses alheios à ciência. A Pequena Tecnologia precisaria então se aproximar da ciência em termos de método, rigor e desinteresse para se legitimar nesse contexto. Desse modo, a lógica de sua produção torna-se ofertista.

Para Merton, quatro imperativos institucionais compõem o *ethos* da ciência: universalismo, comunismo, desinteresse e ceticismo organizado (MERTON, 1979, p. 41). O primeiro deles, universalismo, assume que "as pretensões à verdade, quaisquer que sejam suas origens, têm que ser submetidas a critérios impessoais preestabelecidos: devem estar em consonância com a observação e com o conhecimento já previamente confirmado" (p.

41). Portanto, não importaria nenhuma característica ou valores pessoais de um pretendente a participar da comunidade científica ou de quaisquer de seus membros, vale apenas o compromisso com a objetividade dos resultados obtidos e apresentados.

O valor de comunismo remete ao imperativo de propriedade coletiva dos bens produzidos pela ciência e de sua comunicação imediata. Desse modo, "o direito do cientista à sua propriedade intelectual limita-se à gratidão e à estima que, se a instituição funciona com um mínimo de eficácia, são mais ou menos proporcionais aos aumentos trazidos ao fundo de conhecimentos" (p. 46). Nem o segredo, nem a apropriação privada de conhecimentos científicos e tecnológicos, por meio, por exemplo, de patentes, são vistos com bons olhos. Pelo contrário, geram conflitos por se oporem à noção de bem público a favor de toda a humanidade característica do *ethos* científico.

O desinteresse traduz o valor que impede comportamentos oportunistas por parte dos praticantes de ciência. "A fraude e as pretensões irresponsáveis (charlatanismo)" (p. 50) são paulatinamente desencorajadas pela "necessidade que os cientistas têm, mais cedo ou mais tarde, de prestar contas perante o controle exigente de seus colegas peritos" (p. 50).

Por fim, o ceticismo organizado, que se configura em padrão institucional e metodológico da ciência. A ciência pressupõe "o exame imparcial das crenças, de acordo com critérios empíricos e lógicos" (p. 51). Desse modo, "a ciência, que coloca questões de fato, incluídas as potencialidades, concernentes a todos os aspectos da natureza e da sociedade, pode entrar em conflito com outras atitudes em relação a esses mesmos dados que foram cristalizados e, amiúde, ritualizados por outras instituições" (p. 51). É o caráter de ceticismo organizado que impele a ciência a penetrar e influenciar virtualmente todas as atividades humanas e esferas da sociedade a partir do século XX, inclusive a criação tecnológica.

Veblen, em 1906, já afirmara que a ciência, em sua acepção ampla de produção social de conhecimento e institucionalização de critérios de validade e relevância, se configura em função das características predominantes em uma dada sociedade. Para Veblen, "os hábitos de pensamento que comandam o funcionamento de um sistema de conhecimento são válidos enquanto incentivados pelas atividades sociais mais destacadas, pela estrutura institucional sob a qual a comunidade vive" (VEBLEN, 1964, p. 14). Isso significa que não só a racionalidade científica penetrou os afazeres sociais dos séculos XIX

e XX, mas implica que a própria racionalidade científica concebida como estruturas de causa e efeito e no critério booleano de verdade objetiva são produtos da época. Segundo Veblen, historicamente essas formas de pensamento adquirem legitimidade a partir do momento que o artesão se torna figura proeminente na vida social (início da Idade Moderna) e se intensifica após o advento da sociedade industrial (século XVIII).

Desse modo, os valores de raízes aristocráticas que caracterizam o *ethos* mertoniano da Pequena Ciência criaram uma situação de crescente conflito entre a lógica mecanicista das regras de pensamento e a lógica comunitária das regras de comportamento. Conforme se desenvolvia a indústria, mais esse conflito se caracterizava. A lógica da produção de conhecimento estava implacavelmente em sintonia com a lógica do desenvolvimento industrial das sociedades capitalistas, porém a lógica da organização dessa produção de conhecimento não era a lógica industrial, isto é, não era fundamentada em direitos de propriedade privada, hierarquia e metas pragmáticas.

A sintonia entre a racionalidade científica e o avanço industrial torna-se evidente a partir do final do século XIX. Braveman (1987, p. 138) atribui à ciência, ou melhor, à internalização do método científico pelas indústrias para gerar novas tecnologias, a chave que viabilizou o salto de produtividade e o surgimento das inovações fundamentais que caracterizam a chamada Segunda Revolução Industrial. A indústria química e a indústria eletro-mecânica estabelecidas ao final do século XIX são os exemplos mais marcantes. Os métodos de organização e gestão da produção introduzidos na indústria por Taylor no início do século XX, a chamada administração científica, devem também ser lembrados como exemplos importantes do espírito da época. Os padrões tayloristas iriam depois influenciar também a produção científica de conhecimento nos grandes laboratórios e centros de pesquisa públicos e privados da Grande C&T.

As formas embrionárias da Grande Ciência podem ser encontradas na pesquisa organizada por indústrias inovadoras do final do século XIX. Nye (1996, p. 3) conta que na década de 1870, a BASF na Alemanha foi pioneira no estabelecimento de grandes laboratórios de pesquisa e na contratação de químicos com sólida formação teórica e doutoramentos obtidos nas jovens e dinâmicas universidades alemãs da época. No início do século XX, várias empresas privadas já contavam com laboratórios de pesquisa e desenvolvimento (P&D) que reuniam cientistas de renome, trabalhando segundo códigos de

conduta que aliavam método científico com segredo industrial. Braveman (1987, p. 144) conta que em 1920 havia já em torno de 300 laboratórios de pesquisa em empresas e em 1940 mais de 2.200.

Em 1951, historiadores da ciência como Conant já tinham clareza que uma mudança importante havia ocorrido na organização da pesquisa. Para esse autor, "uma das características do inventor do período clássico da invenção científica, digamos de 1825 a 1925, era que ele podia trabalhar quase que solitário... hoje não mais, um grupo de pessoas altamente treinadas em ciência e tecnologia em grandes laboratórios tomou o lugar do inventor solitário". Nos EUA, o número de pessoas empregadas em pesquisa industrial cresceu de menos de 10.000 ao final da I Guerra Mundial para algo em torno de 50.000 no início da II Guerra, e para mais de 130.000 em 1949. Conant acrescenta que "isso é o resultado de uma fusão de avanços na ciência e progressos na tecnologia que forçou o inventor a cooperar com outros inventores com diferentes habilidades e conhecimentos; as necessidades resultantes em termos de equipamentos e insumos requereram um novo e ampliado método para financiar as invenções. Esses dois fenômenos sociais são interrelacionados tal como o são boas estradas e carros modernos" (CONANT, 1963, p. 303 e 304).

A mudança se tornava cada vez mais nítida. Schumpeter, economista austríaco que em 1912 apontou a importância do indivíduo empreendedor para a saúde do capitalismo, do indivíduo capaz de transformar uma nova idéia (tipicamente um novo resultado científico ou uma nova realização técnica) em um novo produto ou processo e em um novo mercado, em 1950 constatou a perda da relevância social desses indivíduos arrojados e altamente criativos. A inovação se tornara prerrogativa da pesquisa coletiva, institucionalizada sob os auspícios das grandes organizações empresariais. Para Schumpeter, essa mudança institucional em relação aos modos pelos quais a inovação se produz na sociedade seria a causa de mudanças profundas que o próprio capitalismo estaria ainda por sofrer (SCHUMPETER, 1976 e 1982).

No século XX, as duas guerras mundiais foram fartamente subsidiadas pela ciência e pela tecnologia. Na primeira delas, a química (com sua indústria) exerceu papel destacado. Na Segunda Guerra Mundial foram logrados avanços em muitas áreas. As bases das modernas tecnologias de informação e comunicação e das tecnologias espaciais foram

estabelecidas. Técnicas de planejamento e estratégia sofreram um enorme aperfeiçoamento e a tecnologia nuclear foi dominada e utilizada.

Foi a partir do término da II Guerra Mundial, em 1945, que o espírito da Grande C&T se tornou amplamente dominante. A Guerra Fria que se seguiu e o vigoroso crescimento econômico que floresceu até o início da década de 1970 forneceram o ambiente propício para a replicação das estratégias de P&D que se mostraram extremamente eficazes durante a guerra.

A ciência havia provado sua importância e o homem havia por meio dela adquirido poderes sequer imaginados 50 anos antes. Em decorrência, mais e mais recursos financeiros e humanos foram postos à disposição. A nova organização da pesquisa incluiu grandes projetos de P&D militares e civis bancados por recursos estatais e um investimento maciço em formação superior, especialmente nas duas potências que polarizaram o poder global na segunda metade do século XX: EUA e União Soviética.

Uma nova institucionalidade para a C&T, característica da Grande Ciência, ganhava relevo. Há uma multiplicação de institutos de pesquisa orientados a dar respostas científicas e tecnológicas em campos como agricultura, saúde, energia, espaço e defesa (NYE, 1996, p. 225). São organizações com mandatos específicos e crescentes consumidoras de recursos públicos. Do lado privado, os laboratórios industriais de P&D se tornam figura comum aos mais diversos setores de atividade produtiva e surgem indústrias intensivas em conhecimento, especialmente aquelas ligadas às tecnologias de informação, cujos investimentos em P&D consomem fatias expressivas de seus faturamentos.

Em seu estudo de 1963, Price relatou um crescimento vertiginoso na ciência. A partir da análise do crescimento do número de Periódicos Científicos, Price encontrou um ciclo exponencial que dobrava a cada período de 10 a 15 anos. Um ciclo assim era tão curto e impressionante que permitia ao autor concluir que naquele momento (1963) a maior parte de todos os cientistas que já existiram em todos os tempos estava viva e atuando, muitos deles experimentando graus de competição e de divisão do trabalho totalmente diferenciados daqueles vivenciados há apenas algumas décadas.

Price, ao constatar tal crescimento exponencial, rapidamente deduz seu caráter de crescimento logístico (curva em S). O ciclo de duplicação de 10 anos não poderia continuar indefinidamente, ou o número de cientistas ultrapassaria a própria população. Um limite de

saturação seria naturalmente encontrado. Entretanto, o autor desenvolve uma interessante argumentação acerca de limites de saturação no desenvolvimento da ciência. Price atenta para o fato de que, em certos casos, um limite de saturação pode ser suplantado, como de fato o foi no crescimento e nos resultados de algumas disciplinas (PRICE, 1976, p. 15–20). O que ocorre, invariavelmente, é uma mudança de escopo ou de paradigma na atividade de produção de conhecimento na disciplina. Ora, ao estender esse raciocínio para a própria Grande Ciência é evidente que seu crescimento para além de um limite de saturação, dado pela suas próprias estruturas organizacionais e institucionais, só pode acontecer com alguma mudança profunda nessas estruturas.

Tornando-se a C&T uma atividade socialmente financiada, o crescimento da Grande Ciência fez com que a mesma competisse seriamente por recursos e se tornasse objeto de políticas deliberadas por parte de diversos países. Por um lado, C&T entrou definitivamente na agenda das estratégias de poder militar e econômico. Por outro, seu financiamento passou a ser justificado também em termos de eficiência alocativa e em relação a outros gastos estatais e empresariais.

Os esforços de política científica e tecnológica empreendidos a partir da Grande C&T evidenciaram o conflito de valores e contradições fundamentais que subsistem entre Pequena e Grande C&T. Van der Meulen (1998) faz uma ilustrativa análise desses conflitos com o auxílio da teoria dos jogos, mais precisamente considerando a política científica como um jogo agentes-principal.

Uma relação agentes-principal é um modelo para uma situação de interação entre atores tal que um deles, chamado principal, transfere recursos para outros, os agentes, para que estes realizem objetivos estipulados pelo principal, mas que ele próprio é incapaz de realizar. Em troca do fornecimento de recursos, o principal adquire o direito de monitorar as ações dos agentes. Quatro características são típicas de um jogo agentes-principal: 1) os agentes possuem seus próprios objetivos e interesses, que em certa medida estão em conflito com os do principal, e podem usar os recursos adquiridos para atingi-los; 2) o principal não tem conhecimento, capacidade ou os instrumentos para julgar o que é preciso fazer para realizar seus próprios objetivos e depende dos agentes para isso; 3) o principal tem o direito de monitorar e avaliar o desempenho dos agentes; 4) o principal confia nos agentes (VAN DER MEULEN, 1998, p. 399).

As relações entre governos e comunidades científicas se encaixam bastante bem nesse esquema. A política científica pode ser interpretada como um jogo agentes-principal instável, para o qual é necessário institucionalizar formas de estabilização.

Por muito tempo, os governos delegaram o monitoramento e a avaliação da ciência aos próprios cientistas, que possuíam sua própria lógica de controle de qualidade: a revisão pelos pares. A confiança nesse mecanismo era sustentada pela crença de que o progresso da ciência – e automaticamente o de toda a humanidade⁵⁴ – pressupunha a autonomia da ciência.

Entretanto, numa relação duradoura, a repetição do jogo agentes-principal pode levar à modificação das preferências dos atores em função dos resultados das estratégias adotadas no decorrer do tempo. Apesar do efeito de dependência do caminho (*path dependence*), cristalizado pela existência de contratos de estabilização entre os atores que se institucionalizam em rotinas racionais e viáveis para as quais o custo de mudança é elevado, da repetição do jogo podem surgir desconfianças mútuas, novas normas tácitas ou novas regras que paulatinamente alteram os pilares institucionais (cultural-cognitivo, normativo e regulatório) que definiam o arranjo institucional inicial.

Os programas governamentais de pesquisa aplicada típicos da Grande C&T, os processos de avaliação de resultados e de impactos são exemplos que indicam a deterioração da confiança dos governos em um processo de auto organização na ciência, atingindo em cheio a premissa de autonomia do *ethos* da Pequena Ciência.

O interessante do modelo de agentes-principal para interpretar a relação entre Estado e Ciência é que ele explicita a dimensão estratégica das atitudes dos atores na defesa de suas próprias lógicas institucionais. No que tange especificamente à avaliação, no decorrer do jogo, a própria comunidade científica é em grande parte responsável pelo aperfeiçoamento dos métodos de avaliação adotados pelos governos. O Estado legitimaria então as avaliações de base científica como seu mecanismo de monitoramento, em detrimento do mecanismo interno da própria ciência que, por sua vez, era acusado de não ser científico, e sim corporativo.

A evolução desse jogo trouxe no decorrer de suas sucessivas interações novos elementos. Por um lado, devido em grande parte aos custos crescentes das pesquisas, os

⁵⁴ Clássico argumento de legitimação da ciência popularizado a partir do relatório "The Endless Frontier" de Vannevar Bush (1945).

governos sofreram uma relativa perda na sua capacidade de financiamento da pesquisa e as empresas, principalmente as transnacionais, ampliaram sua parcela como financiadoras também para fora de seus próprios laboratórios, por meio de contratos de parceria com universidades e institutos de pesquisa. Por outro, a comunidade científica desenvolveu novos conceitos e teorias para explicar a mudança técnica e as relações entre progresso científico e social. Isso significa mudanças importantes nos mecanismos de legitimação, auto preservação e critérios de qualidade para as atividades de produção de conhecimento.

Considerando ainda outros elementos da macro estrutura técnica e social que se tornaram importantes mais recentemente (discutidos na Seção IV deste capítulo), conclui-se que as atividades de C&T não podem mais ser exclusivamente interpretadas apenas a partir dos arquétipos da Pequena e da Grande C&T. A leitura atual de sistemas de inovação se refere a um novo arranjo institucional em formação. O conceito de *inovação* reforça o caráter de processo social das atividades de C&T⁵⁵ e a noção de *sistemas de inovação* destaca a sintaxe de interdependência na organização dessas atividades, dados a multiplicidade possível de arranjos e atores participantes e os diversos níveis e instâncias de tomada de decisão.

Nessa nova institucionalidade, a multiplicidade de formatos e valores organizacionais se tornou uma das regras (SALLES-FILHO *et. al*, 2000, FERREIRA, 2001). Valores como trabalho em rede, especialização diferenciada, convergência de objetivos entre atores distintos, negociação de prioridades, responsabilidade compartilhada, dentre outros na mesma linha, passaram a ganhar importância.

Trata-se de uma situação ainda um tanto indefinida, porém qualitativamente distinta dos *ethtes* da Pequena e da Grande C&T. A C&T em Rede é ainda um arquétipo em formação e que, ao mesmo tempo, inclui e transforma as formas anteriores.

Conforme previu Price na década de 1960, o crescimento quantitativo da ciência levaria necessariamente também a uma mudança qualitativa. O novo arranjo institucional das atividades de C&T e suas novas formas de legitimação as aproximaram de outros atores sociais, da mesma forma que diferentes atores sociais adotaram a produção de conhecimento e a capacidade de aprendizado constante como valores centrais. O que

⁵⁵ A noção de *inovação* é, portanto, diferente da de *nova tecnologia*. Uma nova tecnologia só é inovação quando apropriada socialmente, quando sai do laboratório e passa a fazer parte do cotidiano da sociedade.

diversos autores chamam de *sociedade do conhecimento*⁵⁶ nada mais é do que uma manifestação dos impactos ao mesmo tempo profundos e difusos das novas instituições que regulam a produção de conhecimento na sociedade.

II. AVALIAÇÃO NA PEQUENA C&T

Esta seção e outras duas que a sucedem são dedicadas ao exame dos principais métodos utilizados para avaliação das atividades de ciência e tecnologia. A separação em três seções obedece ao recorte de tipos institucionais acima discutido. Não é objetivo destas seções detalhar minuciosamente e tampouco exaustivamente o arsenal de métodos existente. A intenção é mais modesta: trata-se de analisar os principais métodos segundo o marco analítico formulado no Capítulo 1 e contextualizá-los segundo seus respectivos estereótipos institucionais apresentados acima. Com isso, espera-se contribuir com o esclarecimento das características atuais da evolução da avaliação em ciência e tecnologia e do contexto em que surgem os principais desafios metodológicos do presente.

O primeiro método examinado é também o mais antigo e consagrado na comunidade científica. Trata-se da avaliação pelos pares.

A avaliação pelos pares reúne, na realidade, uma ampla variação de procedimentos de avaliação que não obedecem a um único método formalizado. Sua fórmula geral, presente em todas as variações, corresponde a um mecanismo corporativo de aceitação e controle por parte dos membros mais antigos e detentores de posições mais elevadas na hierarquia acadêmica. As fórmulas específicas, no entanto, variam desde exames públicos realizados com a presença de uma banca constituída por especialistas reconhecidos na comunidade até o *double blinded review*, no qual um ou mais especialistas fornecem em anonimato seus pareceres sobre uma produção acadêmica também anônima, tendo como intermediário um editor ou um conselho científico.

Diferentes organizações de ensino e pesquisa, agências de financiamento da pesquisa ou revistas científicas desenvolveram e utilizam variações próprias da avaliação pelos pares, como exemplifica o Quadro 3.1. Entretanto, todas elas compartilham das

⁵⁶ Cf. DRUCKER (1994), NOWOTNY *et al.* (2001) e MILES *et al.* (2002).

características comuns de serem avaliações com fundamentos meritocráticos e corporativos, mesmo em caso de serem aplicadas a objetos típicos da Grande C&T e da C&T em Rede.

Quadro 3. 1: Mecanismos típicos de avaliação para auxílio financeiro em universidades (adaptado de GARRETT-JONES, 2000)⁵⁷.

Tipo de auxílio financeiro para pesquisa	Mecanismos típicos de avaliação	
	Ex ante	Ex post
Financiamento institucional	Bibliometria	Bibliometria
	Avaliação institucional	Avaliação pelos pares modificada Painel de usuários Auto avaliação Comissões de avaliação Comparações referenciadas (benchmarks)
Projetos	Avaliação pelos pares	Avaliação pelos pares Painel Análise de relatórios Publicações
	Avaliação pelos pares modificada	Avaliação pelos pares modificada
Manutenção de centros de pesquisa e grandes programas	Bibliometria	Auto avaliação Bibliometria Comparações referenciadas (benchmarks) Painel de usuários Medidas de Cooperação
	Avaliação pelos pares	Número de Pós Graduados Formados Acompanhamento de Carreira Publicações

O Quadro 3.1 inclui também métodos de avaliação diferentes da avaliação pelos pares, como é o caso dos métodos quantitativos, que incluem a bibliometria, a contagem do número de pós-graduados formados e de publicações, medidas de cooperação e comparações referenciadas a *benchmarks*. Há também métodos que modificam drasticamente o conceito de avaliação pelos pares e envolvem na avaliação os usuários da pesquisa ou valorizam atributos irrelevantes ou mesmo antagônicos à lógica acadêmica (retornos econômicos, eficiência, eficácia etc.). Alguns métodos desse quadro se distinguem por, em tese, introduzir na avaliação atributos e critérios externos ao universo acadêmico, embora não necessariamente o façam sempre de modo incisivo, como no caso da avaliação institucional, das comissões de avaliação e da análise de relatórios. Por fim, o acompanhamento de carreiras é um procedimento de cunho mais descritivo e que utiliza em geral abordagens da sociologia ou da antropologia.

⁵⁷ A pesquisa de Garrett-Jones foi feita em agências dos seguintes países: Austrália (4 agências), Canadá (5 agências), Holanda (4 agências), Suécia (1 agência), Reino Unido (2 agências) e EUA (3 agências), além da União Européia (1 agência).

A avaliação pelos pares é, entretanto, uma espécie de matriz para todas essas variações. Isso porque ela é a avaliação acadêmica por excelência. Os imperativos institucionais de universalismo, comunismo, desinteresse e ceticismo organizado atuam como normas tácitas para o reconhecimento acadêmico e são reproduzidos por meio do poder dos pareceres formulados pelos membros mais ativos e reconhecidos da comunidade. É esse controle que delimita e legitima a produção de ciência e tecnologia pelos cientistas num contexto de Pequena C&T.⁵⁸

O desenvolvimento das universidades, das sociedades científicas e das revistas científicas e tecnológicas foram baseados nesse mecanismo de controle. Entretanto, o principal argumento a favor da avaliação ou revisão pelos pares é o "controle de qualidade" exercido pela crítica qualificada a que as novas idéias são submetidas (HARNAD, 2000), apesar de ser claro que isso é apenas parte da função institucional da avaliação pelos pares.

Segundo o esquema analítico introduzido no Capítulo 1, a avaliação pelos pares pode ser entendida como a avaliação de um indivíduo, o cientista escolhido para o papel de revisor, que utiliza um conjunto de critérios interpretados por ele como "critérios de qualidade" para um trabalho científico. Feita a avaliação, o especialista aceita, aceita condicionalmente ou rejeita o trabalho submetido e seu leque de ações concretas inclui a redação das justificativas para sua decisão e a explicação das correções geralmente exigidas.

As formas exatas dos critérios empregados e dos atributos que são observados no trabalho avaliado quase nunca são explicitadas. Eles são subjetivamente construídos a partir das normas gerais dadas pelos imperativos institucionais da ciência, às quais sempre se agrega elementos oriundos de cânones disciplinares ou paradigmáticos, de interesses pessoais e da tradição cultural a que o cientista pertence. Enquanto essa construção é previsível e compartilhada por todos os membros da comunidade, pode-se dizer que existe um método relativamente estável de avaliação que é socialmente legítimo para aquela comunidade.

Porém, os adendos presentes na construção dos critérios e atributos e a inerente dificuldade de objetivação dos critérios para a revisão pelos pares são justamente as

⁵⁸ Essa discussão vale aproximadamente também para a produção de tecnologia num contexto similar ao de Pequena Ciência. Assume-se, nesse contexto, que a produção de conhecimento em geral (seja científico seja tecnológico) é organizada segundo uma lógica acadêmica, ditada pelo *ethos* científico mertoniano.

principais fontes de crítica desse tipo de avaliação. Para os críticos, há muito espaço para os fatores subjetivos e para que o avaliador use seu poder para prejudicar o avaliado ou favorecer os grupos mais fortes e firmemente estabelecidos. Fazendo uma analogia à economia, os cientistas são indivíduos competindo pelos mesmos recursos escassos (fundos para pesquisa, páginas em revistas prestigiadas, fama e poder de controle) e, da mesma forma que a avaliação pode ser empregada para evitar a fraude e controlar a qualidade do conhecimento produzido, ela pode se tornar um importante instrumento para o acesso a esses recursos. Obviamente, aqueles que se sentem prejudicados em algum momento por uma revisão negativa dos pares podem questionar a legitimidade dessa avaliação, num primeiro momento para o caso particular, questionando a capacidade do(s) avaliador(es); mas caso a situação persista (por exemplo, ao se tentar mudar de paradigma no "momento" errado), o questionamento passa naturalmente a valer para o processo como um todo. A partir desse ponto, ocorre a descrença no método de avaliação antes tacitamente compreendido e aceito e a denúncia de que, de fato, não há um método claro.

Por seu turno, as evidências obtidas em pesquisas realizadas sobre o tema da revisão pelos pares concluem que, apesar de tudo, ainda não se demonstrou a viabilidade de nenhuma alternativa ao julgamento especialista para avaliar a produção de outros especialistas e manter a qualidade da literatura especializada (COLE *et al*, 1977; ROY, 1985; HARNAD, 1982 e 2000). Além disso, mesmo recentemente, a maioria dos cientistas aprovava o método da avaliação pelos pares como o mais adequado para suas atividades (ROWLAND, 2003). Esse mesmo autor cita trabalhos que indicam que são relativamente raros os casos de abusos de poder associados à avaliação dos pares, tais como o preconceito frente a mulheres e frente a organizações ou países sem tradição, ou mesmo o plágio e o roubo de idéias por parte dos revisores.

De todo modo, é quase evidente a constatação de que boa parte das modificações introduzidas nos procedimentos da revisão pelos pares ao longo do tempo visou de alguma forma superar suas fragilidades inerentes. Assim, a prática de manter o anonimato do avaliador e, em certos casos, também do avaliado é uma tentativa de evitar a intromissão no julgamento de critérios ligados ao universo pessoal e relacional. Algumas das chamadas "revisões pelos pares modificadas" são caracterizadas pela inclusão de roteiros de avaliação que explicitam os atributos que devem ser verificados, as escalas de avaliação que devem

ser usadas para os critérios e até o algoritmo para agregar esses critérios em um índice geral que vai definir a nota final para o objeto avaliado. Isso é claramente uma tentativa de aumentar a objetividade desses processos de avaliação.

Outros métodos oferecem alternativas ainda mais objetivadas à avaliação dos pares para medir e controlar a qualidade das atividades de ciência e tecnologia.

A bibliometria, inaugurada como atividade sistemática por Price em 1963, é o exemplo mais difundido de metodologia quantitativa para a avaliação da produção acadêmica (PRICE, 1963, COZZENS, 1989e 2000; VELHO, 1989; KOSTOFF et al, 1994; SO, 1998, VAN RANN, 1996).

A bibliometria compreende a quantificação dos produtos (*outputs*) da pesquisa em ciência e tecnologia, isto é, a medida de atributos como o número absoluto (ou seu crescimento) de artigos ou patentes produzidos por um autor, organização ou país, o surgimento de novas revistas, de novas especialidades e disciplinas e assim por diante. Especialmente, a medida do impacto de um autor (ou de um artigo) é dada pela quantidade de vezes em que ele aparece citado por outros autores. Essa é a fórmula adotada pelos estudos bibliométricos para medir a qualidade da produção de um autor e a importância de sua contribuição para a produção de conhecimento. Trata-se, efetivamente, de uma *proxi* para o atributo qualidade.

A lógica desses métodos quantitativos de avaliação da ciência e da tecnologia consiste em identificar e escolher atributos que possam ser mensurados por escalas numéricas e que funcionem diretamente como indicadores de características associadas à quantidade e à qualidade da pesquisa. O atributo qualidade, sendo socialmente relativo, não pode ser diretamente medido em termos numéricos; é necessário então construir algum arcabouço conceitual, uma ligação lógica, entre as variações mensuráveis nos indicadores e o respectivo nível de qualidade associado.

Essa ligação é em geral bastante simples e corresponde à estratégia típica de objetivação de avaliações na qual os atributos são diretamente lidos como critérios. Assim, quanto maiores os números, deduz-se uma maior qualidade e atribui-se uma melhor colocação do autor, da organização ou do país frente à competição pelos recursos do mundo acadêmico.

As técnicas de bibliometria são aplicáveis também a aspectos que estão além das publicações científicas e tecnológicas, como por exemplo, certas medidas de desempenho como o número de doutores ou de estudantes formados por organizações de ensino, o número de pesquisadores em um centro de P&D, indicadores de infra estrutura e organização institucional etc. Esses indicadores, em conjunto com os indicadores bibliométricos, compõem o instrumental da *cientometria* usado para a avaliação de C&T.

O uso de indicadores de resultado para avaliar ciência e tecnologia é objeto de inúmeras críticas (FREEMAN, 1969; VELHO, 1992, 1998 e 2001; MARTÍNEZ, 1998). A principal delas se resume à conhecida noção de que a produção de conhecimento não é um processo de incrementos lineares, ou seja, não há garantia *a priori* que mais artigos, mais patentes, mais doutores e mais instalações signifique proporcionalmente mais conhecimentos. Isso é claramente uma distorção do salto conceitual que trata atributos como critérios. Na realidade, muitas coisas podem ocorrer entre a variação nos atributos medidos (e nos não medidos) e a consubstanciação de conhecimentos socialmente apreendidos, seja em forma de aceitação de novas teorias, seja na forma de inovações tecnológicas. O desafio estaria, portanto, em conceber indicadores relacionáveis a critérios válidos para cada realidade social.

As diferentes áreas do conhecimento evoluem em taxas diferentes e geram impactos diferentes na sociedade, não necessariamente proporcionais ao número de citações que seus autores recebem (um exemplo importante de não proporcionalidade corresponde aos artigos de revisão, metodológicos ou que descrevem técnicas analíticas; eles são geralmente muito citados, sem no geral significarem algum impacto relevante para a evolução do conhecimento). O tempo de maturação de novas teorias ou de inovações tecnológicas também não é o mesmo e tampouco obedecem às mesmas etapas de amadurecimento (assim, sabe-se que nem toda nova tecnologia vem de pesquisas científicas básicas e que desenvolvimentos puramente tecnológicos podem ser a base para novas teorias científicas).

Desse modo, a avaliação quantitativa produzida pelos indicadores da *cientometria* é sempre insuficiente e, apesar de se mostrar coerente para uma quantidade expressiva de casos, não pode ser tomada ao pé da letra, sob o risco de se cristalizar perigosas distorções e apoiar decisões de modo equivocado; isto é, em desacordo com os objetivos perseguidos de controle de qualidade.

Isso significa que a objetivação pretendida na prática apresenta limites que precisam ser subjetivamente considerados e relativizados. O mesmo tipo de problema denunciado na avaliação pelos pares, isto é, a presença de elementos imponderáveis suscitados pelo sujeito, é novamente introduzido, agora em outro nível.

Portanto, o embate entre críticos da subjetividade da revisão pelos pares e os críticos do reducionismo quantitativo da cientometria não se esgota. A conclusão mais razoável e aceita correntemente é que ambas abordagens se complementam, mas são ambas imperfeitas. O uso de uma ou outra depende mais dos objetivos da avaliação e do perfil da audiência dessa avaliação do que de qualquer suposta superioridade metodológica.

Mas um fato chama a atenção. Por que, a comunidade científica buscou objetivar a avaliação de C&T em indicadores quantitativos? A revisão pelos pares sempre gozou de legitimidade dentro da comunidade científica e, apesar dos descontentamentos pontuais, parece ainda ser o mais legítimo mecanismo de avaliação dessa comunidade (ROWLAND, 2003). Como explicar a motivação para a avaliação quantitativa?

A necessidade da linguagem quantitativa, objetivada em indicadores, surgiu a partir do momento que a legitimação da ciência e dos cientistas precisou ir além do universo corporativo, quando a ciência passou a ter que se legitimar socialmente frente a outros atores que, por sua vez, não compartilhavam da mesma lógica de reprodução institucional.

Desse modo, quando a Grande C&T introduziu uma nova lógica de avaliação, a Pequena C&T reagiu. Os métodos da cientometria, concebidos para avaliar mais objetivamente a ciência, são uma resposta às tentativas externas de avaliação, são fruto da necessidade de participar de um diálogo pautado por uma outra racionalidade, a racionalidade econômica e administrativa empregada pelas empresas e pelo Estado. Ironicamente, essa outra racionalidade estava baseada em métodos objetivos, fundamentados em pressupostos "científicos", e se opunha ao sistema corporativo, quase medieval, da revisão pelos pares. A cientometria ilustra uma tentativa de avaliar a ciência e a tecnologia de modo objetivo, mas ainda a partir de seus próprios valores.

III. AVALIAÇÃO NA GRANDE C&T

A pesquisa científica e tecnológica organizada em grandes laboratórios para realizar projetos ambiciosos é um empreendimento dispendioso e de alto risco. Apesar disso, pode trazer enormes retornos econômicos, políticos e de bem estar social. E foram justamente esses retornos que justificaram os grandes investimentos realizados em pesquisa e desenvolvimento. Inicialmente, no período que vai do final do século XIX até antes da Segunda Guerra, esse investimento era majoritariamente privado. Entre a Segunda Guerra e a década de 1970, o investimento público se destaca e, desde a década de 1980, novamente o esforço privado ganha importância.

Desde logo, o caráter desses investimentos é revestido de intenções e objetivos para os quais o avanço do conhecimento era apenas um meio e não o fim. Assim, a avaliação conforme concebida pela Pequena C&T cede lugar a esforços orientados a justificar os investimentos e a avaliar em que medida a C&T seria capaz de transformá-los em ganhos econômicos, políticos ou sociais.

Esse tipo de abordagem desenvolveu seus próprios métodos de avaliação, em geral a partir de conceitos de economia e das teorias de administração. Os objetos preferenciais dessa linhagem de métodos foram e são os grandes programas tecnológicos, sejam privados ou estatais. Os institutos de pesquisa, mesmo se autônomos, também são alvos comuns. As universidades, em sua maior parte, são ainda as organizações que melhor se protegem da avaliação segundo critérios externos à lógica científica, conforme o jogo agentes-principal comentado anteriormente ilustra.

Dois grandes objetivos sustentam as avaliações na Grande C&T: o *accountability* e o *assessment*. Os dois termos são de difícil tradução para o português. A noção de *accountability* vai além da contabilidade, ela inclui também a justificação. Fazer *accountability* significa avaliar a eficiência dos gastos e a eficácia na realização dos resultados planejados, de modo a saber se o investimento em C&T foi vantajoso frente a outras opções de investimento. Pode-se ter *accountability* para justificar gastos *ex ante* ou *ex post*.

O *assessment*, por sua vez, é tributário à difusão da percepção pública de que certas inovações tecnológicas poderiam levar a dramáticos impactos negativos sobre a sociedade e

sobre o meio ambiente. Os mecanismos acadêmicos de controle de qualidade não eram desenhados para lidar com esses tipos de impactos e nem suficientes, uma vez que muitas dessas inovações eram gestadas em instituições pautadas não pelo *ethos* acadêmico mas pela lógica da Grande C&T. Era necessária e oportuna a intervenção de um poder maior (o Estado) para evitar que tecnologias prejudiciais fossem difundidas na sociedade. A palavra mais próxima em português para o verbo *to assess* é *assessorar*,⁵⁹ a avaliação do tipo *assessment* carrega o sentido de um processo no qual especialistas ajudam os governos a identificar, avaliar e mitigar os potenciais impactos negativos de uma nova tecnologia na sociedade e no meio ambiente, de modo que uma inovação possa ser introduzida na sociedade da melhor maneira possível. O *assessment* é normalmente *ex ante*, porém pode também ser realizado durante a difusão da inovação ou *ex post*.

De volta ao primeiro objetivo, de *accountability*, do ponto de vista metodológico, ele se divide em dois grandes ramos de desenvolvimento: os métodos de mensuração econômica e os procedimentos de aferição burocrática.

A mensuração econômica, *grosso modo*, obedece à tradição dessa disciplina em interpretar a sociedade a partir de suas relações macro ou microeconômicas. Hertzfeld (1992, *Apud*. GEORGHIOU e ROESSNER, 2000, p. 659) classifica as estratégias para medir o retorno econômico dos investimentos em pesquisa em três grupos:

1. Adaptação de modelos de funções de produção macroeconômicas para estimar os efeitos das mudanças tecnológicas ou do conhecimento tecnológico, que podem ser atribuídos ao P&D realizado, no PIB ou em outra medida agregada de impacto econômico;
2. Modelos microeconômicos que avaliam os retornos de determinada tecnologia para a economia estimando os ganhos gerados para consumidores e produtores;
3. Medidas de patentes, licenças, contratos, *royalties* pagos, valor em vendas etc., que se relacionam com diferentes modelos macro e microeconômicos.

"A noção de avaliação econômica tem por base não somente uma idéia difusa de impacto sobre as atividades econômicas, mas a concepção de que a pesquisa interfere de alguma forma no sistema econômico; e que o uso de alguma metodologia de avaliação

⁵⁹ *Assessment* vem do verbo latino *assidere*, também raiz do vocábulo *assentar* em português (*asseoir* em francês). A idéia de assentamento tecnológico pode parecer estranha num primeiro momento, mas traz à tona a idéia de resolução de conflitos contida na terminologia anglo-saxã e que *assessorar* não capta satisfatoriamente. A principal função do *assessment* é adequar a tecnologia ao uso prático em larga escala junto ao tecido social, minimizando os riscos e maximizando os benefícios (SALOMON, 1991).

econômica deve captar sua relevância. Os métodos associados à otimização de uma função utilidade de consumo permitem, ao menos em teoria, associar ao processo inovativo um impacto no bem-estar. Como a passagem é feita, é alvo de ampla controvérsia" (FURTADO & SILVEIRA, 2000, p. 14).

É preciso também distinguir na literatura os casos nos quais a teoria econômica é usada para justificar investimentos públicos ou privados em P&D daqueles em que as atividades de inovação são empregadas apenas como objeto empírico de discussão acerca das próprias controvérsias teóricas da economia. Esse segundo caso, apesar de interessante, não cabe diretamente nesta seção.⁶⁰ Evidentemente, as diferentes abordagens econômicas levam a diferentes metodologias de avaliação, fornecendo, portanto, pontos de partida úteis para discorrer sobre os diferentes métodos.

A noção de custos *versus* benefícios constitui a base da maior parte das avaliações econômicas. Essa é uma idéia força muito difundida e aceita socialmente. A métrica monetária, base dos cálculos de custos e benefícios, é universalmente aceita e comparável em diferentes tempos e lugares. O atributo *valor monetário* de um objeto talvez seja o caso mais evidente de transposição direta de atributo a critério. O problema, como apontaram Furtado e Silveira, é como estimar esse valor monetário e como dele deduzir o bem estar.

Dentro da abordagem macroeconômica, os métodos mais utilizados são os estudos de taxas de retorno e os modelos econométricos para estimativa do crescimento do produto agregado de um setor ou de um país. Salter e Martin (2000) discutem o caso mais complicado, qual seja, medir os ganhos da ciência financiada por recursos públicos, com uma ampla revisão dos métodos empregados. O principal ponto destacado pelos autores é a constatação de que os atributos captados por valores monetários não são suficientes para captar todos os benefícios gerados pela ciência. Os modelos econométricos não dão conta daqueles atributos que não são normalmente transacionados nos mercados e as aproximações realizadas, como a precificação de ativos ambientais, de qualidade de vida ou de conhecimentos, no geral deixam muito a desejar. Por outro lado, justamente nesses

⁶⁰ Um bom resumo dos principais elementos dessa discussão, para três importantes linhas de pensamento da economia da mudança tecnológica (ciclos longos/paradigmas tecnológicos, evolucionistas e escola da regulação) pode ser encontrado em POSSAS (1987). Uma interessante discussão mais recente em economia recorre ao caso das atividades de C&T financiadas publicamente para estudar o chamado efeito de deslocamento (*crowding out*), segundo o qual o investimento estatal provocaria sub investimentos privados e prejudicaria a configuração de mercados e a apropriação de bens e serviços (DAVID & HALL, 1999, DAVID *et al.*, 1999).

atributos é que muitos impactos positivos (ou negativos) da ciência produzida com investimento público se manifestam. A revisão, nesta seção, dos detalhes desses métodos e de seus mecanismos de estimativa monetária, extrapola o necessário para a breve caracterização pretendida. A título de exemplo, dois campos de pesquisa são típicos do investimento público em C&T e, por abrangerem desde ciência básica até o desenvolvimento e aplicação de tecnologias sofisticadas, prestam-se bastante bem para diferentes formatos de avaliação: a área espacial e a agricultura. Para ambos, a literatura de avaliação é bastante rica e é possível encontrar toda a variedade metodológica mencionada aqui.⁶¹

Para os programas tecnológicos, os métodos de estimativa de ganhos de produtividade pelo deslocamento da fronteira tecnológica, devido à introdução de uma nova tecnologia que altera a combinação dos fatores de produção (capital, trabalho e matérias-primas), deixam escapar atributos como o aprendizado e os chamados efeitos de transbordamento (*spill overs*), isto é, as novas indústrias ou os incrementos sistêmicos de competitividade que a nova tecnologia eventualmente engendre.

Os impactos de uma nova tecnologia em geral assumem contornos distintos para diferentes setores da economia e da sociedade, o que gera mais uma dificuldade para os métodos de corte macroeconômico e demanda o emprego de abordagens centradas em agentes e atores específicos.

As metodologias microeconômicas também utilizam métodos de estimativa de taxas de retorno ou de ganhos de produtividade. Comparativamente, conseguem resultados mais precisos, no sentido de maior proximidade com a realidade e melhor definição do significado de bem estar caso a caso, mas a comparação intersetorial ou mesmo entre casos é mais difícil. Algumas abordagens mais recentes, como o caso da metodologia BETA,⁶² incorporam elementos da abordagem econômica evolucionista, que enfatiza os processos

⁶¹ Para o caso da agricultura cf. CRUZ et al. (1982), EVENSON (1987), AVILA e EVENSON (1995), ALSTON et al. (1995) e MUNDLAK (2000). Para as tecnologias espaciais HERTZFELD (1998), BACH e LAMBERT (1992), WILLIAMS e RANK (1998). Um trabalho pioneiro, e importante, sobre avaliação econômica da pesquisa industrial é o de MANSFIELD (1968). Revisões críticas das principais metodologias econômicas de avaliação da pesquisa podem ser encontradas em FURTADO e SILVEIRA (2000) e FURTADO e SOUZA (1999).

⁶² Método para avaliação de grandes programas tecnológicos, desenvolvido pelo Bureau d'Economie Théorique et Appliquée (Universidade Louis Pasteur de Estrasburgo, França), que alia medidas econômicas com indicadores de aprendizado e outros efeitos de segunda ordem importantes para captar os impactos das inovações na sociedade. Cf. BACH e LAMBERT (1992 e 1998), BACH et al. (1998) e BACH (2001).

microeconômicos de geração e adoção de inovações, e dão importância a atributos como o aprendizado, a circulação do conhecimento e os efeitos de transbordamento. Trata-se de um avanço, no sentido de extrapolar os atributos monetários e incorporar medidas não precificadas.

Há certamente um descompasso fundamental entre os objetivos econômicos de eficiência alocativa e os objetivos de justificação de investimentos em ciência e tecnologia. De fato, é possível escolher ou ajustar metodologias que enfatizem cada caso e destaquem o que for de interesse destacar. A escolha dos modelos empregados é necessariamente subjetiva e reflete os valores e interesses daqueles que avaliam. Fazendo novamente alusão ao jogo agentes-principal, esses valores são diferentes para os diferentes participantes do jogo da Grande C&T e, especialmente, essas diferenças se amplificam se a Pequena C&T for avaliada por métodos de *accountability* e *assessment*. É natural que a academia critique a eficiência alocativa e procure introduzir no debate, em novos termos, elementos que valorizem a produção de conhecimentos e o aprendizado.

O terceiro grupo de metodologias identificado por Hertzfeld ilustra diretamente esse fenômeno. Ele inclui metodologias nas quais certas medidas tradicionais da ciëntometria são lidas como indicadores econômicos. A contabilidade de patentes registradas e dos fundos arrecadados com as respectivas licenças seria uma forma de medir simultaneamente os retornos econômicos (assumindo que as tecnologias licenciadas se tornarão inovações e proporcionarão ganhos de produtividade), a eficiência do investimento em P&D e a qualidade desse P&D.

De fato, esse terceiro grupo já se encontra na transição entre método econômico e método de aferição burocrática. O objetivo de eficiência alocativa quando transposto para uma organização gera rotinas específicas, materializadas em controles burocráticos desenhados para atingir os indicadores de desempenho esperados.

Desse modo, havendo a suposição de que basta o investimento em C&T para se ter inovação (modelo linear), medir o volume desse investimento traduziria a capacidade de inovação de uma empresa ou de um país. Como esses indicadores de investimento (*inputs*) foram e ainda são muito utilizados para comparar a intensidade de P&D de diferentes indústrias e países, deduzir daí também a capacidade de inovação se torna tentador, mas não

é verossímil. Senão, a conclusão de qualquer avaliação de C&T seria sempre a de que falta investimento.

Para as organizações de P&D, é a relação *input/output* que interessa como medida de eficiência. Há necessidade de minimizar essa relação para que uma organização se mantenha competitiva na arena em que são disputados os recursos, sejam públicos ou privados. Aumentar o número de patentes depositadas e licenciadas, o número de artigos publicados, o número de estudantes formados e o número de cursos ou treinamentos oferecidos por unidade monetária é uma forma de *accountability*. E a qualidade? A qualidade não é diretamente aferida pelo investidor, como é típico num jogo agentes-principal. Para o investidor, a qualidade já é aferida pelos mecanismos internos da C&T, a revisão pelos pares e a análise das patentes (realizada por técnicos especialistas nos temas) ou pelo desempenho dos profissionais no mercado de trabalho. Para avançar além, torna-se necessário realizar avaliações dos impactos dos produtos gerados pela C&T em aspectos econômicos, sociais, ambientais, dentre outros, que, como foi visto, não são simples.

Ainda no âmbito da aferição burocrática, indicadores como o número de pesquisadores doutores empregados e o número total de funcionários, o número de projetos, o número de grupos de pesquisa, as diferentes fontes de financiamentos conseguidas etc. fazem parte de todo um arsenal de medidas (que efetivamente se tornam objetivos da organização) para justificar e legitimar sua existência. Após o advento da Grande C&T, a necessidade de *accountability* se tornou progressivamente universal, atingindo inclusive as organizações que nitidamente funcionam como Pequena C&T, como universidades e alguns institutos de pesquisa básica, os quais precisaram aprender a conviver com uma situação de competição por recursos baseada em indicadores burocráticos que, pela óptica dos valores da Pequena C&T, seria contraditória e equivocada.

Passando para o outro objetivo da avaliação na Grande C&T, o *assessment* é fruto de demandas sociais por controle sobre os resultados da C&T. Os estudos de *assessment* proliferaram assim que a percepção pública dos impactos negativos gerados pela pesquisa tornou-se um fator de ameaça à continuidade das atividades de P&D, especialmente após o surgimento das bombas nucleares em 1945 e dos impactos ambientais provocados pela agricultura tecnificada sentidos a partir do final da década de 1960.

Assim, os procedimentos de *assessment* surgiram na década de 1960 como "uma resposta do governo, indústria e academia a um clima hostil crescente por parte da opinião pública em geral em relação ao 'progresso' tecnológico" (WYNNE, 1975, p. 116).

Na prática, o *technological assessment* opera com o uso de muitas técnicas emprestadas da Pesquisa Operacional, técnicas de *forecasting* e técnicas de análise econômica (OCDE, 1992; VAN DEN ENDE *et al.*, 1998. BERLOZNIK, 1998). Os métodos com fundamentação econômica já foram comentados. A Pesquisa Operacional é uma disciplina tributária dos desenvolvimentos da organização militar e da organização científica da produção industrial (taylorismo), sendo pautada por métodos de otimização e controle de procedimentos e processos. A avaliação no contexto da Pesquisa Operacional está ligada à análise das rotinas e dos procedimentos constituintes de uma atividade produtiva (ou de uma ação militar) com o objetivo explícito de maximizar sua eficiência. A análise científica de táticas, estratégias e ações, seja voltada a conflitos bélicos e políticos, seja para competição em mercados teve sua origem na Pesquisa Operacional.

Nos estudos de *assessment* as ferramentas da Pesquisa Operacional são empregadas para modelar diversas situações, tais como os ciclos de vida e a difusão de novas tecnologias e de seus impactos, além da análise e estimativa probabilística dos riscos associados a determinados impactos potenciais. Modelos de otimização de várias naturezas (lineares, dinâmicos, de rede, estocásticos etc.) podem servir de base para simular o comportamento de determinadas variáveis de *design* desde o início de um projeto tecnológico de modo a minimizar futuros impactos que seriam avaliados negativamente (por exemplo, a otimização da eficiência em termos de consumo energético em determinados dispositivos a fim de reduzir os impactos ambientais).⁶³

Os métodos de apoio à decisão multicritério discutidos no capítulo anterior também se inserem no escopo da Pesquisa Operacional. Gomes e Duarte (1997) descrevem uma aplicação desse instrumental para a avaliação de indicadores de produção tecnológica de organizações de pesquisa. A partir de um modelo multicritério, é possível aproveitar dados gerados por cientometria e conjugá-los com outros tipos de atributos e critérios para apoiar decisões em termos de *assessment* e *accountability*. Em Zackiewicz *et al.* (2005) é descrita

⁶³ Para uma introdução aos métodos da Pesquisa Operacional confira os manuais assinados por WAGNER (1986) e GOLDBARG e LUNA (2000). Para uma visão mais aprofundada do instrumental matemático para modelagem cf. MARLOW (1978). Um texto pioneiro que vale a pena ser conhecido é o de SAATY (1959).

uma abordagem multicritério para estudar prioridades de pesquisa a partir de resultados de enquetes do tipo Delphi. O modelo é capaz de considerar simultaneamente perspectivas conflitantes de diferentes atores para encontrar frentes de consenso ou para permitir análises comparativas e de sensibilidade em relação a diferentes cenários de decisão.

O *forecasting* reúne um grande conjunto de técnicas para antecipação, análise de tendências e previsão de eventos futuros. Dentro do chamado *technology forecasting*, ramo especializado na análise da evolução das tecnologias e seus impactos, o procedimento mais famoso é o método Delphi, mas outras técnicas também se destacam, entre elas a análise morfológica, os modelos de ciclos tecnológicos, a análise de sistemas e a matrizes de impacto cruzado. Os detalhes sobre os diversos métodos existentes e suas variações podem ser encontrados nos manuais assinados por Jantsch (1967), Makridakis, Wheelwright e McGee (1998) e Armstrong (2001).

As técnicas de *forecasting* há muito vêm sendo criticadas por suas limitações. Muitos trabalhos discutiram e sistematizaram as principais limitações dos métodos mais utilizados.⁶⁴ Também nos estudos de *forecasting*, componentes subjetivos, preferências e interesses definem as escolhas dos métodos e a aceitação ou não dos resultados. O uso de métodos formalizados e de análises e deduções matemáticas muitas vezes pleiteiam graus de objetividade e neutralidade que não correspondem à realidade. Mesmo assim, é interessante observar que grande parte das críticas sempre denunciou a incapacidade dos métodos ou de seus usuários, mas quase nunca as premissas de objetividade e neutralidade. Poucos autores, como o caso de Wynne (1975) ainda na década de 1970, tinham clareza acerca da impossibilidade dessas premissas e só mais recentemente algumas novas abordagens passaram a procurar caminhos para superar esse tipo de crítica e a modificar tais premissas em suas avaliações sobre os impactos futuros das tecnologias. O exemplo da abordagem de *technology foresight* discutido adiante na seção IV é emblemático.

Ainda assim, nos últimos quinze anos, importantes autores da escola de *forecasting* têm realizado esforços para contornar os problemas de seus métodos sem abrir mão de suas premissas originais. É notável a incorporação, a seu modo, de novos elementos, especialmente daqueles que tratam da complexidade dos sistemas. Assim, ouve-se cada vez

⁶⁴ Cf., por exemplo, WHISTON (1979), HOGWOOD e GUNN (1984) e SCHMÉDER (1988).

mais falar de modelos de sistemas dinâmicos não lineares, de caos determinístico,⁶⁵ de atratores e de simuladores de evolução nos modelos de *forecasting*. A noção de sistema há muito está presente nesses modelos, só que limitada a suas formulações mais elementares: a de um sistema estável e tendente ao equilíbrio e a de sistemas de oscilações estáveis. Linstone (1999) reconhece outras três configurações sistêmicas complexas que estariam presentes em sistemas não-lineares: 1. sistemas caóticos com fronteiras previsíveis; 2. sistemas instáveis e divergentes (ou evolutivos) e 3. sistemas auto organizantes. Entretanto, também reconhece que nem mesmo a matemática possui atualmente teorias e modelos adequados para tratar esses casos e que, portanto, os avanços no campo de *forecasting* estariam condicionados por esses limites, que corresponderiam a uma das fronteiras do próprio conhecimento científico atual (ZACKIEWICZ, 2003).

Como todos os métodos baseados nos pressupostos científicos de causa e efeito, que são aqueles que permitem inferência e previsão rigorosos e reprodutíveis sem interferência subjetiva, tanto os procedimentos típicos de análise de risco e de otimização da Pesquisa Operacional quanto os métodos de previsão do *forecasting* encontram limitações ao serem aplicados a objetos sociais ou a sistemas complexos com intercausação. A incerteza associada aos processos de inovação, o aprendizado e as mudanças de estratégia dos atores sociais e agentes econômicos inviabilizam a validade dos modelos construídos sobre essas bases.

A incerteza se revela preponderante. Até mesmo as variáveis que eram importantes em um momento podem se tornar menos importantes; enquanto variáveis muitas vezes sequer imaginadas passam a desempenhar um papel central. Como imaginar o efeito dos clorofluorcarbonetos (CFC) na incidência de câncer de pele na população se no período em que essa inovação se difundia (e gerava trajetórias de investimento praticamente irreversíveis) os gases do tipo CFC eram considerados inertes e inofensivos? A descoberta da propriedade de catalisador da decomposição do ozônio em altas altitudes, permitindo assim o aumento da incidência de raios UV na superfície terrestre, só aconteceu pelo menos

⁶⁵ Um comportamento caótico é aquele que obedece a equações determinísticas não-lineares, mas possui desdobramentos imprevisíveis. Esse aparente paradoxo é explicado matematicamente como dependência extrema das condições iniciais, isto é, qualquer ínfima variação das condições iniciais de um modelo provoca grandes variações no comportamento do sistema e os resultados mudam completamente. Assim, para um sistema caótico ser previsível, seria necessário medir suas variáveis a partir de um referencial com precisão absoluta, o que é impossível na prática. Um sistema caótico não é um sistema aleatório, é um sistema complexo (RUELLE, 1993).

vinte anos depois da tecnologia ter se tornado dominante. O conhecimento científico para prever essa variável em um modelo de *assessment* ou em qualquer outro tipo de avaliação simplesmente não estava disponível no momento adequado.

Finalmente, conforme argumenta Salomon (1991 e 1992), há uma dimensão política importante associada à emergência do *assessment*; dimensão essa que anteriormente não se destacava nitidamente. Segundo o autor, não é por acaso que a sociedade que mais institucionalizou a mudança técnica – os EUA – também foi a primeira a institucionalizar mecanismos de controle dessas mudanças: o *assessment* tecnológico, as associações de consumidores e a intervenção do poder judiciário na regulação das tecnologias.

Entretanto, em sua essência, o *assessment* seria de fato um mecanismo de despolitização (SALOMON, 1991), primeiro porque confinaria o poder de ação sobre o ritmo e a orientação da mudança técnica a um círculo restrito dos especialistas, e segundo porque funcionaria como um mecanismo de integração da dissidência e da contestação ao funcionamento normal das instituições. Os procedimentos técnicos do *assessment* deixariam de fora a participação de outros atores, daqueles que potencialmente sofreriam as consequências se algo der errado. De fato, o *assessment* perdeu progressivamente o espaço conquistado nas décadas de 1960 e 1970, processo simbolicamente representado pelo fechamento da agência encarregada de executar estudos de *assessment* nos EUA, o Office of Technology Assessment (OTA), que funcionou entre 1972 e 1995.⁶⁶ A realidade se mostrava demasiadamente incerta e os mecanismos de avaliação criados dentro da racionalidade da Grande C&T mostravam paulatinamente sérias dificuldades em dar conta do que deles era exigido.

IV. MACRO TENDÊNCIAS NA ORGANIZAÇÃO DA PRODUÇÃO DE CONHECIMENTO

No final da Seção I deste Capítulo, é mencionada uma conexão geral entre a emergência de certos elementos da macro estrutura técnica e social e o fortalecimento de uma terceira forma institucional das atividades de ciência e tecnologia, batizada de C&T em Rede. Esta seção retoma essa linha de raciocínio, naquele ponto interrompida para apresentar os métodos de avaliação da Pequena e da Grande C&T, e avança na

⁶⁶ Os estudos do OTA estão disponíveis em <http://www.wws.princeton.edu/~ota/>

caracterização dos principais elementos de um contexto institucional que embasa e legitima novas formas de organizar a produção de conhecimentos na sociedade.

A partir da década de 1980, diversos desenvolvimentos conceituais progressivamente aprofundaram a compreensão dos fatores e dos mecanismos da dinâmica das inovações. De modo geral, esses desenvolvimentos estão apoiados no modelo sintático de sistema dinâmico aberto e em evolução.⁶⁷ Nem a tecnologia nem tampouco a sociedade seriam sistemas fechados e isolados, ambos estariam interrelacionados e co-evoluíam segundo padrões nem sempre previsíveis de auto organização. Entre os diferentes enfoques desenvolvidos, quatro se destacam e formam o essencial da base conceitual desenvolvida para estudar as formas da C&T em Rede.

a) A crítica do modelo linear de inovação. A seqüência de etapas independentes e subsequentes – pesquisa básica (ou ciência), pesquisa aplicada (tecnologia), desenvolvimento de produto e processo e entrada no mercado – foi substituída por um modelo interativo, que generaliza o caso linear e no qual cada uma dessas etapas exerce influências sobre as outras e é influenciada por elas e outros fatores presentes nos processos de inovação, tais como conhecimentos tácitos e aprendizado (ROSENBERG, 1982). Diversos modelos sofisticados surgiram a partir de então. Um exemplo recente seria o modelo *triple helix*,⁶⁸ segundo o qual governos, empresas e universidades evoluem adquirindo características uns dos outros a partir de interações sucessivas que produzem inovações como resultado. Nesse modelo, a universidade deteria papel preponderante na geração de conhecimentos que são as bases para a inovação.

b) A interpretação evolucionista da teoria econômica e os sistemas de inovação. A concepção do sistema econômico como um conjunto de relações maximizadoras das satisfações individuais com fluxos de oferta e demanda tendendo para o equilíbrio é reinterpretada por um sistema constantemente desequilibrado pela introdução de inovações tecnológicas e organizacionais que alteram estes fluxos. O comportamento dos agentes passou a ser percebido como rotinas de busca por inovações, cujas tentativas são selecionadas pelo mercado (ambiente) em um processo interativo (NELSON e WINTER, 1982). Diversos autores desenvolveram outras interpretações sistêmicas para descrever e explicar a indissociabilidade entre os processos de inovação e suas instituições portadoras.

⁶⁷ BERTALANFFY (1973) e MOIGNE (1994).

⁶⁸ ETZKOWITZ e LEYDESDORFF (2000).

Porter (1990) utilizou as empresas multinacionais como elemento de referência e estudou os determinantes sistêmicos de sua competitividade; Nelson (1987) procurou na unidade dos Sistemas Nacionais de Inovação a explicação para as diferenças de dinamismo econômico entre países; Lundvall (1992) enfatizou os processos de aprendizado proporcionado pelos fluxos compartilhados de conhecimento nos sistemas de inovação.

c) A emergência de uma nova organização da produção de conhecimento. Gibbons e colaboradores (1994) indentificaram um novo paradigma da produção de conhecimento (Modo 2), com as características de ser socialmente distribuído, orientado a aplicações, transdisciplinar e sujeito a múltiplas formas de *accountability*. O Modo 2 estaria paulatinamente substituindo o paradigma anterior, ou Modo 1, caracterizado pela hegemonia das disciplinas científicas, hierárquica e desenvolvida segundo a agenda autônoma de cientistas e universidades. O Modo 2 evoluiria em paralelo com a sociedade do conhecimento, num processo de auto organização conjunta. "O tipo ascendente é o *pesquisador*, que precisa agora mostrar talento empreendedor; enquanto isso o *cientista* está eclipsado. Em outras palavras, a profissionalização da ciência trouxe seus praticantes para perto de outros profissionais e grupos altamente educados da sociedade, ao mesmo tempo em que ocorreu uma bem sucedida difusão das características do *ethos* científico para esses grupos" (NOWOTNY *et al.*, 2001, p. 46).

d) Os conceitos de redes técnico-econômicas, redes sócio-técnicas e ator-rede. Michel Callon (1992) propôs o conceito de redes sócio-técnicas para apreender o papel de atores heterogêneos na produção de inovações. Para Callon, as redes técnico-econômicas seriam um conjunto coordenado de atores heterogêneos (laboratórios públicos, centros de pesquisa, empresas privadas, organizações financeiras, usuários e governos) que participam coletivamente na concepção, desenvolvimento, produção e difusão de tecnologias para a produção de novos bens ou serviços. Estes atores estariam organizados em torno de três pólos de coordenação: científico, técnico e de mercado. Através desses "ambientes" é que ocorreriam as transações que caracterizam a rede. Elementos intermediários, como textos, artefatos técnicos, comportamentos, habilidades humanas e dinheiro, indicariam a materialização das interações entre os componentes da rede. O conhecimento compartilhado estaria em parte codificado, enquanto importante parcela permaneceria tácita, embutida em artefatos, procedimentos e habilidades cuja transferência não é direta

nem fácil. A noção de rede sócio-técnica é uma generalização desse conceito que enfatiza a importância das competências existentes em um determinado contexto social para que haja capacidade sistêmica de produzir e compartilhar novas tecnologias, especialmente levando em consideração a relação entre os elementos tácitos do conhecimento e os novos atores socialmente construídos. Nesse sentido, a teoria dos atores-rede estabelece parâmetros para uma sociologia na qual inclusive artefatos técnicos passam a compor tipos relevantes para explicar as relações sociais contemporâneas.⁶⁹

Todas essas linhas de investigação teórica e empírica têm em comum a indeterminação causal entre fatores tecnológicos e sociais. A abstração de *sistema* é empregada justamente para reforçar o caráter de determinação múltipla, de complexidade e incerteza que acompanha a evolução conjunta das técnicas e das relações sociais. Afinal, "além da mudança na composição e organização dos fatores de produção, ocorrem transformações de caráter psicológico, social, cultural e político que acompanham a mudança técnica" (SALOMON, 1992, p. 189). E, "quando uma transformação ocorre no universo cultural ... então o universo do conhecimento sofre uma mudança análoga" (VEBLEN, 1964, p.14).

A rigor, ambas as direções da implicação são válidas e podem ser empregadas como hipóteses para orientar a investigação e esclarecer as diferentes interrelações entre mudança técnica e sociedade. É difícil, portanto, dizer se as mudanças institucionais vêm antes ou depois das mudanças técnicas ou até que ponto os métodos de avaliação empregados por uma comunidade são definidos por ela ou a definem. As duas implicações ocorrem simultaneamente e se reforçam.

Na linha desse último argumento, Inglehart já identificava em 1977 a ocorrência de uma "revolução silenciosa" nas sociedades desenvolvidas, um movimento transformador de valores, competências e formas de expressão política. O crescimento econômico, a expansão do ensino secundário e superior, a importância e a diversidade dos meios de comunicação em massa e a ampliação na descontinuidade das experiências a que cada vez mais pessoas estavam submetidas seriam as principais causas desse movimento. Do ponto de vista da organização das sociedades, estaria em declínio a legitimidade da autoridade

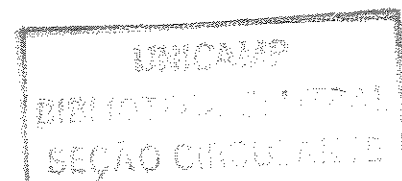
⁶⁹ Michel Callon (1990) considera o estudo da mudança técnica e a caracterização das tecnologias um ótimo pretexto para se estudar a sociedade e sua evolução. Cf. ainda CALLON (1999) e LATOUR (1999) para uma discussão crítica sobre a teoria do ator-rede (actor-network theory).

hierárquica e os valores tradicionais da sociedade industrial, num processo impulsionado principalmente pelas classes médias (INGLEHART, 1990; 1997).

Michel Crozier, já em 1970 chamava a atenção para como as novas as formas de convívio social, estabelecidas a partir da rotina de trabalho das grandes organizações empresariais, estariam favorecendo a inovação, a criatividade e o não conformismo como fenômenos coletivos. Apesar de primeiro identificadas nos ambiente empresariais, essas capacidades estariam se difundindo para o conjunto da sociedade e influenciando seus valores e instituições. Desse modo, o desafio estaria em "desenvolver conjuntamente o sistema de grandes organizações e as formas de governo e de relações humanas democráticas que permitem a manutenção da criatividade dos indivíduos no caminho da inovação coletiva indispensável às sociedades pós-industriais" (CROZIER, 1983, p. 41). Leituras mais recentes dessas tendências apontam que efetivamente, a partir dos anos 1990, uma nova configuração empresarial ganhou espaço. Boltanski e Chiapello (1999) identificam a flexibilidade como o atributo essencial dessa nova configuração. A flexibilidade se manifestaria internamente em decorrência de mudanças expressivas na organização e nas técnicas do trabalho e, externamente, pelo aumento da subcontratação de serviços (inclusive de pesquisa tecnológica) e pela maior rotatividade da mão-de-obra (inclusive daquela altamente qualificada). À noção de flexibilidade estariam também relacionadas a tendência de diminuição de estoques, a organização da produção em grupos autônomos e a diminuição dos níveis hierárquicos.

Associado a essa tendência de flexibilização ganharia terreno na sociedade um discurso crítico contra a burocracia, a planificação rígida distanciada da realidade e a autoridade. Tratar-se-ia de uma oposição ao modelo taylorista de organização dominante no período anterior, impulsionada socialmente pelo imperativo da competitividade e pelos ciclos cada vez mais rápidos de mudança tecnológica. Entram em cena o trabalho em rede, por projetos, voltado à satisfação do cliente, organizado a partir das *competências essenciais*⁷⁰ da organização e compartilhado com fornecedores e diferentes tipos de prestadores de serviço (*ad hoc*, temporários ou parceiros). A empresa não é mais administrada por um "chefe", e sim por um "líder" que possui uma *visão* e coordena tal empresa para que ela se auto organize em torno dessa visão (BOLTANSKI e CHIAPELLO,

⁷⁰ Cf. PRAHALAD, HAMEL (1994).



1999, p. 119). O valor econômico de uma empresa depende cada vez mais se seus conhecimentos e da capacidade de articula-los com *ativos complementares*⁷¹ para criar inovações.

Há também um nova geopolítica, que com o fim das economias planificadas e o significativo aumento no predomínio das representações políticas democráticas, promoveu uma maior homogeneidade institucional para o desenvolvimento dessa nova organização econômica. Os EUA e a União Européia são ainda as grandes forças econômicas, seguidas por diversos países asiáticos (Japão, China e Coréia do Sul são os mais evidentes) e outros emergentes.

Do ponto de vista da organização do espaço global, o aperfeiçoamento das tecnologias de comunicação e de transporte agiram encurtando as distâncias e aumentando a integração física entre os países. A comunicação instantânea e a custos baixos para qualquer ponto do globo é um dos fatores que viabilizaram negócios globalmente e estabeleceram novos canais de diálogo interculturais. A internacionalização das empresas e a globalização cultural seriam também importantes motores do processo de modernização social (DICKEN, 1998).

Quanto às implicações para os Estados, segundo Castells (1998), os Estados estariam também assumindo uma configuração em rede, apoiado nas tecnologias de informação como mediadoras das tensões entre o global e o local. Na prática, o fortalecimento de oito características de funcionamento administrativo pressupõem o surgimento desse Estado-Rede: 1) a tendência à descentralização, delegando a níveis administrativos inferiores as atividades que possuam especificidades locais; 2) a flexibilidade organizacional e de atuação, reduzindo a rigidez dos procedimentos e abrindo espaço para soluções inovadoras; 3) a coordenação, através de regras de subordinação democraticamente estabelecidas; 4) a participação dos cidadãos, introduzindo mecanismos que permitam que estes emitam suas opiniões e possam participar de atividades do Estado; 5) a transparência administrativa, facilitada pelos recursos de informação; 6) a modernização tecnológica da administração; 7) a profissionalização da burocracia, reduzindo quadros e pagando salários acima do mercado e 8) uma forma de gestão retroativa, capaz de aprender com seus erros e de corrigi-los no mínimo tempo.

⁷¹ Cf. TEECE (1986).

Entretanto, todas essas mudanças institucionais seriam mudanças que ocorreriam em regime de adaptação, sem grandes saltos. Offe (1998) procura desfazer mitos tais como o de que o Estado nacional tenderia para a incapacidade de ação autônoma no mundo globalizado, ou de que, por outro lado, seria ele ainda a única instituição capaz de resolver os problemas sociais e garantir o bem estar. Da mesma forma, as forças de mercado não seriam o único mecanismo viável de desenvolvimento, sem, no entanto, se tornarem irrelevantes em algum horizonte previsível. Finalmente, o processo de aproximação entre as diferentes culturas não significaria nem uniformização cultural nem radicalização dos antagonismos, mas a oportunidade para um processo dialético de assimilação mútua.

Todo esse conjunto de tendências de mudanças culturais e de organização institucional foi acompanhado por um impressionante desenvolvimento das tecnologias de informação. As tecnologias de informação reúnem um amplo conjunto convergente de tecnologias em microeletrônica, computação (*software* e *hardware*), telecomunicações, radiodifusão, optoeletrônica e genética (CASTELLS, 2002, p. 67).

Essas tecnologias, ao mesmo tempo em que constituem as bases físicas das redes, também são em grande parte concebidas a partir dos conceitos analíticos de sistema e demandam capacitação e capacidade de aprendizado contínuo de seus usuários.

Na década de 1960, as pressões sociais que levaram ao surgimento dos processos de *assessment* eram a expressão de uma nova leitura do significado de progresso. E a partir do momento que se instaurou a nova leitura, uma tecnocracia formada por avaliadores profissionais, treinados em métodos específicos surgiu com novos leitores. A transposição desse raciocínio para o período atual também pode ser feita. Há espaço para uma nova leitura acerca das relações entre mudança social e tecnológica, como atestam os diversos desenvolvimentos conceituais e teóricos destacados no início desta seção.

Mas, como defende Salomon (1991; 1992), essa leitura não pode ficar restrita a esses mesmos profissionais, haveria necessidade de mecanismos de participação mais democráticos na avaliação das tecnologias e o respeito à dissidência (mesmo àquela desinformada tecnicamente). Salomon aposta na ampliação da capacidade dos atores externos ao universo técnico-científico tradicional em opinar sobre mudança técnica, uma vez que a sociedade caminharia para padrões educacionais cada vez mais elevados e os arranjos para produção de conhecimento são cada vez mais heterogêneos.

Como disse Bruno Latour "Ciência é certeza; pesquisa é incerteza. Supõe-se que a ciência é fria, rígida e determinada. A pesquisa é quente, envolvente e arriscada. A ciência é capaz de por fim às divagações das disputas humanas; a pesquisa cria novas controvérsias. A ciência produz objetividade por escapar o máximo possível das algemas da ideologia, das paixões e das emoções; a pesquisa se alimenta de tudo isso para tornar familiares seus objetos de investigação" (LATOUR, 1998, p. 208-9). Na Pequena e na Grande C&T, a ciência e a pesquisa passaram a andar juntas, eram parte de uma mesma atividade, organizadas em Modo 1 – para usar a terminologia de Gibbons. Na C&T em Rede, em contraste, o impulso e a legitimidade para pesquisar ganham novamente as ruas, passam a fazer parte da agenda e do dia a dia de outros atores – numa configuração Modo 2 de produção de conhecimentos. O casamento de mais de um século entre ciência e tecnologia (celebrado na Segunda Revolução Industrial) se tornou, no mínimo, mais liberal, afeito a estabelecer conexões muitas vezes inusitadas.

Numa sociedade que se torna cada vez mais definida pelo *ethos* da pesquisa, da inovação, pela capacidade continuada de aprendizado, pela ampliação dos horizontes espaciais e culturais, pelos arranjos flexíveis de produção, mas também pelo aumento da incerteza, dos riscos e da rapidez com que desigualdades econômicas e de inserção sócio cultural são construídas, a produção e a circulação de conhecimentos se reconfiguram, se impregnam ao tecido social de um modo essencialmente novo, mais democrático mas também mais imprevisível e descoordenado.

V. AVALIAÇÃO NA C&T EM REDE

As características institucionais da C&T em Rede são produto do entrelaçamento de novas tecnologias, mudanças institucionais da sociedade como um todo e da evolução dos arranjos organizacionais pelos quais se realiza pesquisa e se produz inovação.

É evidente que os elementos institucionais da Pequena e da Grande C&T não desapareceram e que em muitas organizações de C&T são ainda dominantes. Eles permanecem, talvez em alguma medida sofram mutações, tenham alguns de seus aspectos reconfigurados. Por outro lado, algumas características, como a disposição para a pesquisa e para compartilhar conhecimentos se difunde para além de suas antigas fronteiras. Mas há

também aqueles que representam rotinas e valores enraizados e que entram em conflito com o novo *ethos* nascente. De todo modo, as tentativas para definir precisamente as características desse *ethos* em rede da produção de conhecimento são ainda insuficientes, pelo fato incontornável dele ainda estar em formação.

Por exemplo, retomando o esquema do jogo agentes-principal, na C&T em Rede não haveria mais apenas um jogo, entre os institutos de pesquisa e universidades e o Estado financiador. A pesquisa em rede supõe múltiplos jogos, nos quais determinados atores podem ser ora agentes ora principais. As opções de formas e fontes de financiamento se multiplicam e também as opções de parceria e de arranjos envolvendo empresas, organizações não governamentais, fundações etc.

A crescente heterogeneidade das formas de produção de conhecimento, das formas de financiamento, dos modos de apropriação dos conhecimentos produzidos, das profissões envolvidas em pesquisa, das novas disciplinas e cruzamentos interdisciplinares implica em dificuldades descritivas e normativas também crescentes para o campo da C&T. E, sem dúvida, as atividades de avaliação também esbarram em impasses.

Esta seção procura, a partir de exemplos de novas abordagens metodológicas para a avaliação da C&T em Rede, explicitar quais seriam os principais desafios para avaliar a produção de conhecimento sob esse tipo institucional.

Em um primeiro momento, a avaliação da pesquisa expandiu sua perspectiva. De seu propósito tradicional de verificar apenas em que extensão os objetivos eram cumpridos e a qualidade era mantida, ganhou importância o entendimento de como os resultados obtidos condicionavam e eram condicionados pelo entorno social ao qual se relacionavam (CALLON *et al.*, 1995). Essa fase é simultânea aos desenvolvimentos teóricos comentados na seção anterior.

Entretanto, a percepção de que uma nova institucionalidade estaria em gestação, implicando e demandando novos métodos de avaliação, provocou a emergência de esforços metodológicos mais ousados. Neles, a maior novidade era a introdução de um caráter de pró atividade. Os métodos de avaliação se propunham a ir além da mera constatação dos fatos. Por que não aproveitá-los segundo uma disposição concreta em produzir efeitos de governança e coordenação para a inovação?

Em ambas as formulações, os esforços de avaliação ganharam maior abrangência metodológica e incorporaram aspectos voltados ao entendimento das estratégias e da articulação entre os diferentes atores relacionados aos processos de inovação. Georgiou & Roessner (2000) identificaram três tendências, correntes a partir da década de 1980, no desenvolvimento das abordagens de avaliação de programas tecnológicos, advindas das mudanças nas condições institucionais e da concepção do processo de inovação. Para esses autores, as seguintes tendências seriam preponderantes:

1. Ocorre convergência entre as tradições de avaliação interna (do tipo revisão pelos pares e cientometria) e os preceitos oriundos das avaliações adotadas para as políticas públicas em geral (*accountability* e *assessment*);
2. Aumenta a requisição, por parte dos gestores públicos, de indicadores de desempenho e de programação para as instituições de C&T;
3. Difunde-se, a partir do plano conceitual, a correlação entre produção científica e desempenho competitivo, provocando a busca de meios efetivos para estabelecê-la na prática.

A linha mais descritiva, que marca os primeiros esforços metodológicos de avaliação da C&T em Rede, se fundamenta em estudos de caracterização, com descrições sistemáticas das instituições, campos de conhecimento e das estruturas em rede que perpassam e articulam diferentes atores (GEORGHIOU, 1995). De fato, essas abordagens inovam em seus objetos de avaliação. Já que o processo de inovação é incerto, complexo e pode derivar para trajetórias diferentes das planejadas, seus objetivos podem se tornar alvos móveis e, portanto, caracterizar a organização e a evolução do sistema passa a ser mais importante do que simplesmente medir a qualidade e eficiência de seus *outputs*.

Diferentes abordagens ilustram essa orientação metodológica.

Em certos casos, os desenvolvimentos teóricos para estudar a C&T em Rede são simultâneos a desenvolvimentos metodológicos que permitem avaliá-la. Isso ocorre, por exemplo, com as redes sócio-técnicas de Callon. Embora a abordagem tenha sido inicialmente elaborada para interpretar as relações sociais de atores heterogêneos no processo de inovação, ela se presta também para avaliar as situações de sucesso ou fracasso da C&T em Rede, para nelas identificar elementos relevantes que expliquem os desdobramentos do caso e que possam se tornar "lições" a serem replicadas ou evitadas em

experiências posteriores. Na avaliação de redes sócio-técnicas entram atributos como estrutura, localização e extensão geográfica, densidade, dispersão, conectividade e outros. A partir de medidas realizadas sobre esses e outros atributos as redes podem ser caracterizadas segundo cinco categorias: redes incompletas ou encadeadas, curtas ou longas, dispersas ou convergentes, emergentes ou estabilizadas, e polarizadas ou sem dominância.⁷² De acordo com a combinação obtida dessas categorias, diferentes ações práticas para promover a inovação se justificam.

Bozeman *et al.* (1999) propuseram um modelo para avaliação da pesquisa baseado em medidas de capital humano da C&T. Esses autores são críticos aos métodos de corte microeconômico geralmente empregados para avaliar a pesquisa nos EUA (uma vez que a partir de 1993 a avaliação de eficiência institucional se tornou mandatória nesse país para todas as organizações financiadas pelo Estado). Para eles quatro fatores pesariam contra essas abordagens: 1. os métodos microeconômicos apenas podem ser aplicados com sucesso quando as fronteiras do objeto avaliado são perfeitamente definidas; 2. a existência de produtos da ciência e da tecnologia que se difundiram e se desvincularam das pessoas ou dos contextos em que foram gerados geram um escopo irrealista para as avaliações; 3. as avaliações tendem a ser estáticas e são incapazes de captar as mutações dos produtos avaliados e de seus impactos; 4. as avaliações orientadas a produtos prestam geralmente pouca atenção à geração de capacitação de habilidades para manter sustentável o fluxo de inovações (BOZEMAN, 1999, p. 5).

O que esses autores propõem como alternativa metodológica é o acompanhamento das trajetórias individuais dos pesquisadores, da evolução de suas realizações e das interações de que participa. As redes poderiam ser reconstruídas e estudadas no tempo, no espaço e do ponto de vista qualitativo.

Esses breves exemplos mostram que as tentativas em curso procuram levar a prática da avaliação para além das medidas da relação *input/output*. Outros elementos, que procuram captar a complexidade do sistema “pesquisa científica e tecnológica” e suas relações com outros sistemas, seja o sistema econômico, os sistemas sociais e culturais, o sistema ambiental ou outros, vêm sendo incorporados.

⁷² Outros exemplos de indicadores, ou seja, atributos que já carregam sentido de critério, para redes incluem o Grau de Centralidade de Freeman, a Centralidade Emparelhada de Freeman (*betweenness centrality*) e a Centralidade da Proximidade (*closeness centrality*) (GUEDES, 1999).

Há também um segundo grupo de abordagens de avaliação que surge no contexto da C&T em Rede. Mais ligado a abordagens originalmente desenvolvidas nas escolas de planejamento e estratégia, esse grupo reúne procedimentos de avaliação que podem ser caracterizados também como instrumentos de fomento à inovação. As avaliações realizadas auxiliam na explicação dos casos de sucesso, a dar coerência para as formas organizacionais, a sinalizar necessidades de competências específicas, a indicar o grau de aderência das atividades de pesquisa com os anseios internos e externos à comunidade ou instituição de pesquisa avaliada.

Para Gaffard (1991), a produção e troca sistemática de informações entre um mesmo nível de hierarquia ou de competências gera coordenação horizontal e um processo de aprendizado que fortalece a coordenação vertical entre os diversos níveis que possibilitam, viabilizam e orientam uma inovação. A avaliação pode auxiliar na importante tarefa, especialmente ao envolver diversos atores – dos que produzem o conhecimento aos que por ele são afetados – de fazer circular informações estratégicas para a consecução de impactos desejáveis. Uma tendência geral observada é a de expandir o alcance das metodologias e incorporar elementos úteis à própria organização dos sistemas avaliados.

A ênfase atual das metodologias que incorporam essa perspectiva é nitidamente colocada sobre procedimentos participativos. É a partir das percepções de diferentes atores ligados à C&T que se espera despertar a capacidade criativa coletiva necessária à inovação. O planejamento, portanto, não é explícito, ele é incorporado no processo e se manifesta pelo aumento da compreensão e do comprometimento – institucionalizados – sobre os caminhos a seguir. Nesse sentido, a coerência das decisões acompanharia o fortalecimento de processos de reflexão coletiva sobre futuro, experiências passadas e contexto atual.

Um exemplo importante desse movimento é dado pelos exercícios prospectivos chamados *technology foresight*.

Não por acaso, no termo *foresight* a referência está no presente e não no futuro. O procedimento de *foresight* remete à idéia de construção intencional do futuro. O ponto de partida é assumido como não neutro e no geral os resultados do pensar o futuro em termos de *foresight* são normativos, levando à definição de prioridades e/ou outras políticas visando impactos pré-definidos. Para tanto, procura-se mobilizar proativamente a inteligência coletiva dos atores.

O *technology foresight* surgiu na Inglaterra, principalmente com trabalhos de pesquisadores do SPRU (Science Policy Research Unit – Sussex) e do PREST (Policy Research in Engineering, Science and Technology – Manchester). A grande contribuição dessa linha de desenvolvimento conceitual e metodológico é a percepção de que, se as tecnologias são produtos sociais, o planejamento quanto a seus rumos deve ser negociado socialmente para ser efetivo. Assim, a abordagem de *foresight* enfatiza a legitimação das prioridades “pré-vistas” para o desenvolvimento futuro por meio de processos coletivos de busca, com os quais se procura também comprometimento e consenso entre os pesquisadores de instituições públicas e privadas. Os exercícios de *foresight* encontraram vasto campo e difusão a partir do final dos anos 1980 como ferramenta para fortalecer e coordenar Sistemas Nacionais de Inovação, sendo institucionalizados junto a agências governamentais em diversos países.⁷³

Mais recentemente, Miles *et al.* (2002) deliberadamente aumentaram o escopo dessa abordagem para o que eles chamaram de *knowledge society foresight*. Desse modo, ficou explícita a intenção de envolver outros atores além daqueles convencionalmente percebidos pelos gestores de C&T e da inovação. A União Européia têm desenvolvido inúmeros estudos com base em abordagens de *foresight* para avaliar suas potencialidades e problemáticas, especialmente na escala micro regional, e planejar seu desenvolvimento, não apenas do ponto de vista tecnológico, mas integrando o elemento inovação às mais diversas dimensões da vida social.

Não é o caso de detalhar os métodos empregados pelo *foresight*, pois são muitos. A abordagem se caracteriza mais por seus princípios que por seus métodos. Em geral, para um exercício de *foresight* são combinados diversos métodos, alguns mais formais, outros mais informais, alguns buscando objetivação de atributos, outros simplesmente aceitando o conflito de opiniões e a ambigüidade dos resultados. Tudo depende dos objetivos e do

⁷³ Entre eles destacam-se, pela ordem de implementação, o Japão, a Alemanha, o Reino Unido, a Austrália, Nova Zelândia, Itália, França, Espanha, Holanda e Áustria. Existem, entretanto, significativas diferenças metodológicas entre eles. Países como Reino Unido, Austrália e Nova Zelândia adotaram uma orientação mais pragmática e voltada para a competitividade da indústria. Nos países latinos – Itália, França e Espanha – o processo é mais centralizado pelas agências governamentais, com um viés menos afeito à participação plural de distintos atores e mais a uma linha de planejamento com a participação de poucos especialistas. Países pequenos como Holanda e Áustria inovaram e procuraram realizar exercícios de *foresight* em busca de nichos tecnológicos alternativos e em resposta a necessidades sociais específicas. No Brasil, exemplos dessa linha metodológica são encontrados no estudo conduzido pelo Programa PROSPECTAR (realizado entre 2001 a 2003) e pela abordagem de prospecção de oportunidades de ciência, tecnologia e inovação do Centro de Gestão e Estudos Estratégicos (CGEE).

objeto da avaliação e do delicado equilíbrio que é preciso construir entre aprendizado, participação e convergência de resultados. A combinação metodológica, a condução do processo e a capacidade de envolvimento dos atores (e das pessoas certas para os representarem) se conjugam para criar um ambiente propício para a avaliação coletiva.

No próximo capítulo, uma abordagem de avaliação mais específica que os exercícios de *foresight* é apresentada. Trata-se de um método para avaliar impactos da pesquisa, concebido para funcionar de modo participativo e para gerar aprendizado coletivo. Como o *foresight*, esse método é também aberto à combinação de outros métodos e depende da adequada condução e de seus participantes para ser efetivo. Diferentemente do *foresight*, seus procedimentos são mais formalizados, há passos a seguir e modos pré-definidos para lidar com a ambigüidade e com as diferentes perspectivas que possam surgir acerca dos impactos do projeto, programa, instituto ou outro objeto da C&T em Rede avaliado.

O Método de Avaliação em Múltiplas Dimensões (MDM)

Capítulo 4 – Concepção e propriedades de um método para avaliação na C&T em Rede

In the modern world the intelligence of public opinion is the one
indispensable condition for social progress

Charles W. Eliot

A C&T em Rede é uma forma de organização da produção social de conhecimento que traz novos desafios para a avaliação. No Capítulo 3, foram arroladas as evidências da progressiva constituição e difusão dessa nova forma institucional da C&T, que generaliza as anteriores ao mesmo tempo em que atinge novos atores sociais e amplia as dificuldades de planejamento.

As alternativas metodológicas para avaliar a C&T em Rede foram, em um primeiro momento, derivadas dos avanços conceituais e teóricos alcançados a partir da apreensão da dinâmica do desenvolvimento da ciência, da tecnologia e da sociedade como fenômenos necessariamente entrelaçados. Em seguida, essas abordagens passaram a ser base para processos deliberados de envolvimento dos atores ligados à produção de conhecimentos e à viabilização de inovações para gerar coordenação entre eles. Há uma aposta implícita de fomento da auto organização dos sistemas de relações existentes nas Redes, em direção a maior eficiência, maior efetividade e a trajetórias socialmente desejadas de inovação.

Essa linha de desenvolvimento metodológico carece, entretanto, de métodos claramente definidos. Obstáculos importantes são colocados pelas especificidades dos casos de aplicação e pela impossibilidade de objetivar a participação de atores com expectativas contraditórias em um método racional de avaliação e decisão (conforme foi discutido no Capítulo 1).

Neste capítulo, descreve-se uma tentativa de formalizar os procedimentos de um método para avaliação no contexto da C&T em Rede. Sua primeira característica, como não poderia deixar de ser, é também ser capaz de generalizar e acomodar os métodos característicos das formas institucionais anteriores da C&T. Nesse sentido, o método

fornece uma heurística geral para tratar de diferentes objetos em diferentes situações, e precisa ser modelado em seus componentes internos de acordo com cada caso de aplicação.

A segunda característica é ser um método participativo. Trata-se de uma estrutura para uma avaliação coletiva, segundo os termos definidos no Capítulo 1. Nela, há espaço para diferentes atores expressarem suas concordâncias e suas dissidências. O algoritmo de agregação para as diferentes visões é essencialmente subjetivo e evita uma racionalização objetivadora. A agregação é, nesses termos, realizada com o auxílio de uma heurística não fundamentada na lógica convencional (devido às implicações das propriedades discutidas no Capítulo 2).

Por fim, como terceira característica, o método proposto aproxima-se de um instrumento tecnológico, desenhado para responder da melhor maneira possível as múltiplas, e muitas vezes difusas, demandas que a C&T em Rede impõe à avaliação e ao planejamento. Por ser um método, isto é, ser uma construção social que pode ser perfeitamente descrita em seus procedimentos, pode também ser transposto para um *software*. Efetivamente, isso foi feito, demonstrando o caráter de método da abordagem desenvolvida.

Este capítulo contém nove seções, todas elas dedicadas à explanação dos elementos e dos detalhes da aplicação prática de um método para avaliação da C&T em Rede, aqui intitulado MDM – Método de Avaliação em Múltiplas Dimensões.

VI. ANTECEDENTES E PRINCÍPIOS GERAIS

A trajetória percorrida para se chegar a um método de avaliação adequado para responder às necessidades da C&T em Rede foi em grande medida estimulada pelo programa de investigação e construção metodológica formulado por um projeto de pesquisa conduzido entre os anos 1999 e 2003 no Departamento de Política Científica e Tecnológica da Unicamp, situado em seu Instituto de Geociências. O projeto, denominado "Políticas Públicas para a Inovação Tecnológica na Agricultura do Estado de São Paulo: Métodos para Avaliação de Impactos de Pesquisa", foi financiado inicialmente pelo primeiro Edital de Políticas Públicas da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).⁷⁴

⁷⁴ FURTADO (2003).

A motivação central desse projeto estava no desafio de conceber um método de avaliação que superasse as deficiências em geral presentes nos métodos de avaliação empregados para medir os impactos da introdução de novas tecnologias na agricultura.

A agricultura se constitui num objeto privilegiado para revelar a insuficiência de certos métodos de *accountability* e de *assessment* na avaliação do desempenho dos esforços de inovação. Grande parte da pesquisa orientada para o desenvolvimento de novas variedades de cultivares e para outros melhoramentos tecnológicos voltados ao aumento da produtividade e da qualidade dos produtos agropecuários é de caráter público e organizada em moldes de Grande C&T ou, mais recentemente, de C&T em Rede (conforme a terminologia discutida no capítulo anterior).

Nesses termos, as avaliações da pesquisa agropecuária focam, em geral, no desempenho do esforço de financiamento da pesquisa *vis a vis* o desempenho das tecnologias no rebaixamento dos custos dos produtos agropecuários na sociedade (ver referências do Capítulo 3). Como o aumento de produtividade dos cultivares impacta diretamente no rebaixamento dos custos, os ganhos econômicos são óbvios e a taxa de benefícios sobre custo é normalmente muito elevada quando se computa os ganhos monetários resultantes da difusão da nova tecnologia frente ao montante gasto para manter as estruturas de pesquisa.

Entretanto, os modelos microeconômicos desenhados para medir esses ganhos são baseados em variáveis econômicas, técnicas e geográficas, como a variação de preços, as variações nos coeficientes técnicos de produtividade e as áreas ocupadas pelos diferentes cultivares. Os impactos sobre o meio ambiente e sobre indicadores sociais como emprego, renda e qualidade de vida não são computados.

Esses impactos são mais difíceis de apreender, por diversos motivos. Os principais estão ligados à dificuldade de isolar a influência exclusiva de uma nova tecnologia sobre o meio ambiente ou sobre diferentes indicadores sociais. Essa influência é chamada atribuição, ou seja, a pergunta é até que ponto se pode atribuir a uma nova tecnologia os impactos não econômicos sentidos no meio ambiente e na sociedade. Os modelos econômicos de avaliação de impactos não são capazes de captar isso diretamente. Quando tentam considerar esses impactos não econômicos (considerados externalidades do modelo)

o fazem de modo indireto, por meio da precificação dos impactos ambientais e sociais ou da expectativa futura sobre a ocorrência de impactos.

Outras limitações se verificam em relação à intertemporalidade dos impactos e à percepção dos impactos (ganhos ou perdas econômicas) por parte dos diferentes atores envolvidos. A intertemporalidade é fenômeno difícil de apreender em qualquer método de avaliação (de resultados ou de impactos), exatamente porque dificulta ainda mais as questões de atribuição de causa aos efeitos verificados. Já a heterogeneidade da percepção dos atores envolvidos se refere às diferentes formas de apropriação dos benefícios (ou malefícios) gerados pela introdução de uma tecnologia.

Outras variáveis econômicas importantes como os impactos sobre a organização dos mercados e sobre os custos de transação e os efeitos de transbordamento ocasionados pela nova tecnologia também são em geral ignorados. O aprendizado tecnológico e a capacitação adquirida e acumulada para gerar novas inovações também não são levados em consideração.

O estudo original definido pelo projeto de pesquisa acima citado consistia em considerar os impactos de programas tecnológicos voltados à agricultura em pelo menos quatro dimensões: econômica, social, ambiental e na geração de competências. Dois programas tecnológicos reais foram escolhidos como casos para orientar e testar o desenvolvimento metodológico: o Programa de Melhoramento Genético de Cana-de-Açúcar (PROCANA) e o Programa de Produção de Borbulhas e Mudanças Sadias de Citros, ambos conduzidos no Instituto Agrônomo de Campinas (IAC).

Ao longo do projeto, diversas oficinas e reuniões de trabalho foram realizadas, algumas delas com a participação de mais de 80 pesquisadores de várias áreas do conhecimento. O desafio era a proposição de uma metodologia que pudesse contemplar várias dimensões simultaneamente. Em outras palavras, buscou-se uma abordagem metodológica mais condizente com a realidade dos programas de ciência e tecnologia, cada vez mais praticados em rede e cada vez mais voltados para a inovação. Procurou-se então métodos e abordagens que permitissem a combinação de avaliações econômicas com sociais, ambientais e de capacitação. Provisoriamente, essa avaliação multidimensional recebeu o nome de ESAC, iniciais das quatro dimensões em foco.

Já nos primeiros momentos do estudo, ficou claro que a melhor abordagem para enfrentar esse desafio seria a dos métodos de multicritério de apoio à decisão, conforme experiências anteriores localizadas na literatura (EDWARDS e GUTTENTAG, 1975). Entretanto, diferentemente dos casos revisados na literatura, a abordagem desenvolvida não se limitou a aplicar um método multicritério convencional. Havia a percepção de que outros conceitos desenvolvidos para interpretar e descrever a C&T em Rede de alguma forma precisavam ser incorporados.

Desse modo, o Projeto havia se colocado um desafio instigante, cuja extensão não era de todo conhecida. A metodologia almejava conceber procedimentos que alcançassem o máximo de objetividade e reprodutibilidade, mas que, ao mesmo tempo, dessem conta de aspectos importantes para apreender impactos sob o marco conceitual dos sistemas de inovação, o que envolveria adotar procedimentos participativos, integradores de múltiplas perspectivas conceituais e úteis do ponto de vista do aprendizado e da inovação. Pela discussão já realizada nos Capítulos 1 e 2, pode-se dimensionar as dificuldades desse plano.

O método científico em sua forma clássica apreende a realidade a partir da análise (decomposição) do todo em partes. Cada parte, uma vez simplificada o suficiente, pode ser entendida em suas leis gerais a partir da formulação de hipóteses cuja validação empírica na explicação causal (tese) do fenômeno analisado lhes atribuem valor de verdade. O método prossegue com a tarefa de novamente agregar as partes por um processo de síntese. Nesse processo, as explicações das diversas partes são confrontadas e concatenadas e a verificação de estruturas isomórficas entre diferentes conjuntos de hipóteses e teses verdadeiras leva a generalizações teóricas progressivas. Dessa forma, torna-se válido deduzir o comportamento do geral pela observação do particular e explicar os fenômenos por meio de estruturas matemáticas que relacionem as diferentes naturezas de suas partes relevantes, captadas a partir das escalas e unidades de medida apropriadas. O todo, no caso, é a síntese das partes.

Em outras palavras, o plano de se obter medidas dos impactos, no sentido numérico que se associa à idéia de *medida*, e de rotinas de avaliação, no sentido da implementação de algoritmos computáveis e estáveis, pressupõe a existência de um método de avaliação do tipo clássico. Por outro lado, apreender impactos sob o marco conceitual dos sistemas de

inovação pressupõe um modelo dinâmico e evolutivo, de acordo com a complexidade existente no contexto do objeto avaliado.

Para se construir um método de avaliação de impactos que pudesse dar conta dessas necessidades antagônicas foi preciso lançar mão de uma estratégia de modelagem que pode ser ilustrada a partir de dois níveis. Um nível estruturado, que opera no universo numérico e computável e segue os princípios clássicos de análise e síntese para modelar os impactos e suas medidas. E um segundo nível, contextual, que situa o nível estruturado e incorpora premissas como: a) a coexistência de dimensões de impacto não redutíveis umas às outras; b) a coexistência de múltiplas racionalidades identificadas com a coexistência de múltiplos juízos de valor acerca dos impactos; c) a noção de que os impactos não são dissociáveis de um contexto objetivo complexo – que inclui espaço, tempo, estado inicial, sinergias multicausais etc. – e do contexto subjetivo dos atores impactados – incluindo seus valores, seus próprios processos de adaptação e o conhecimento tácito existente.

Assim, dada a geração e difusão de novas tecnologias em um determinado contexto da sociedade, a medida dos impactos causados (ou previstos) seria dada a partir de uma combinação indissociável entre intensidade objetiva e importância subjetiva das transformações que essas tecnologias provocariam em diferentes dimensões da realidade.

A denominação *dimensão*, adotada aqui, é dada a atributos que generalizam outros atributos mais específicos, mas que não podem ser generalizados sob qualquer outro atributo ainda mais geral. É evidente, que a escolha das dimensões faz parte da estratégia metodológica adotada em uma avaliação. Por exemplo, para o caso do Projeto de Avaliação de Impactos da Agricultura, acima mencionado, as dimensões escolhidas foram a econômica, a social, a ambiental e a de capacitação. Isso constitui um ponto de partida metodológico, uma vez que se pretendia deliberadamente evitar a redução de todas as dimensões a uma só, como fazem algumas abordagens econométricas.

O próximo passo foi incorporar ao processo de modelagem do nível estruturado os elementos característicos do nível contextual de modo que as medidas obtidas fizessem sentido. Desse modo, o nível estruturado não mede os impactos a partir de um referencial de "fora" do sistema, como faz, por exemplo, as diferentes métricas utilizadas nas avaliações do tipo *input* ou *output*. Pelo contrário, ele procura o significado de suas medidas partindo de "dentro" do sistema, ou seja, de sua composição multidimensional (e

portanto, multidisciplinar do ponto de vista de teorias explicativas) e multi atores. Porém, a medida fica restrita aos limites definidos para esse sistema. Por esse motivo, a medida de impacto obtida é uma medida relativa ao próprio sistema avaliado, dependente do espaço e do tempo e intimamente ligada ao universo de atores considerado. A comparação entre diferentes medidas de impacto depende do uso de modelos de avaliação similares e, eventualmente, de se estabelecer referenciais externos aos sistemas avaliados.

Na prática, a modelagem adotada para a avaliação de impactos segue um processo de simplificação da realidade como se segue:

a) Ponto de partida: a realidade completa. As relações entre os atributos relevantes contidos na realidade são complexas, e delas emergem os impactos percebidos nas diferentes dimensões. As fronteiras entre a pertinência de atributos a uma ou outra dimensão é indefinida. Muitas são as sobreposições, os graus e direções das causalidades possíveis, assim como as diferentes importâncias para distintos atores.

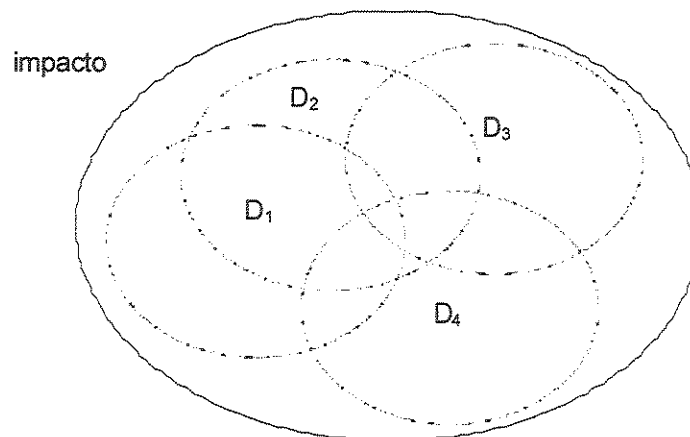


Figura 4.1: Na realidade completa os impactos são complexos.

b) Análise: hipótese de independência e delimitação das dimensões. A análise (divisão) das dimensões e seus componentes de impacto reduz a quantidade e isola os atributos da realidade a um conjunto manejável de componentes do impacto. A análise se faz a partir do estado-da-arte do conhecimento das diferentes teorias explicativas em cada dimensão, da colaboração de especialistas e da delimitação do contexto de aplicação

pretendido para a avaliação de impactos que, tomados em conjunto, apontam para os aspectos mais relevantes para se monitorar a transformação em cada dimensão.

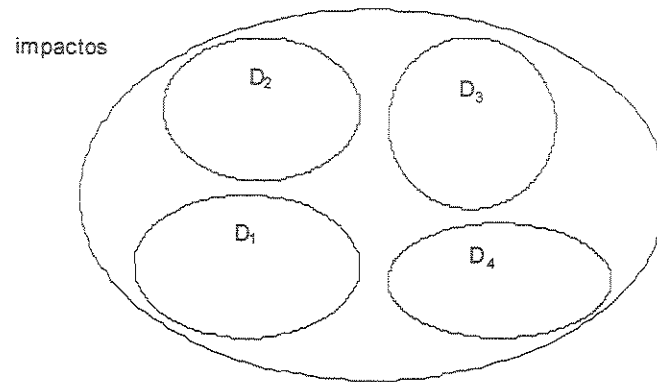


Figura 4.2: As dimensões de impactos analisadas são independentes e delimitadas.

c) Síntese: aplicação de um algoritmo de cálculo dos impactos. As variações observadas nos aspectos escolhidos da realidade são projetadas em uma mesma unidade de impacto definida sob um conjunto reticulado com função de critério. Esses são agregados de acordo com a importância de cada componente para a definição dos impactos para fornecer um valor total para o impacto em cada dimensão.

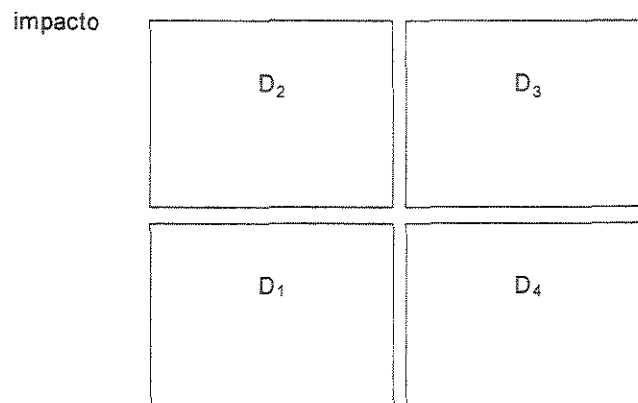


Figura 4.3: A síntese das medidas de impactos as tornam reticulados homogêneos.

Desse processo de simplificação fica evidente que a medida final obtida para os impactos não contém todas informações que caracterizam os fenômenos de impacto no sistema avaliado. Entretanto, com ela se ganha poder comparativo e de comunicação, organização cognitiva e explicativa e um modelo estruturado com detalhamento de medidas e valorações explicitamente consideradas. Os números obtidos ao final são decorrentes tanto dos efeitos provocados pela pesquisa e pela introdução de uma nova tecnologia quanto pela capacidade de apreensão e os juízos de valor sobre esses efeitos. Assim, as medidas de impactos são totalmente relativas ao contexto avaliado e precisam ser associadas à caracterização desse contexto para poderem ser entendidas fora dele.

A escala definida para medir impactos é uma escala artificial e dependente do sistema avaliado, trata-se de uma estrutura matemática de reticulado que adquire significados qualitativos e quantitativos próprios para cada caso real de avaliação. A metodologia de avaliação de impactos foi concebida para ser executada sobre um modelo formal que permitisse a mensuração dos impactos a partir de "dentro" de um sistema sócio-técnico dado, ao mesmo tempo que incorporasse e negociasse na prática, durante o seu processo de execução, os diversos elementos característicos dos esforços de apreensão e organização da inovação: a participação de atores heterogêneos, a circulação de informação, o aprendizado, a multi-causalidade, a incerteza e a ambigüidade.

VII. ASPECTOS FORMAIS DO MÉTODO DE AVALIAÇÃO EM MÚLTIPLAS DIMENSÕES

A principal estrutura matemática subjacente ao Método de Avaliação em Múltiplas Dimensões é intencionalmente simples: uma estrutura hierárquica de reticulados completos, simples ou duplos com sobreposição espelhada, doravante chamada apenas de estrutura de avaliação. Essa estrutura se constitui no elemento organizador e mediador das medidas e juízos necessários para a avaliação. Seu formato simples traz vantagens do ponto de vista computacional, facilita o tratamento dos resultados, a comunicação e o uso da metodologia.

Uma das referências mais usuais de utilização de estruturas hierárquicas, adotadas como orientadoras da investigação inicial e das primeiras tentativas de modelagem para a medição de impactos, é a dos métodos multicritério de apoio à decisão. Entretanto, a

aplicação desses métodos ao problema da avaliação de impactos revelou algumas dificuldades que precisaram ser contornadas.

O principal limite se referia ao fato de que os métodos de apoio à decisão multicritério são desenhados para problemas de escolha entre alternativas, de modo a seleccionar e hierarquizar opções alternativas para solução de um problema. Na avaliação, especialmente na de impactos *ex-post*, não se trata de comparar alternativas, mas sim de medir estados distintos (ou a variação) de certos aspectos de um certo subsistema da realidade em relação ao surgimento e incorporação de uma inovação. Do ponto de vista formal, nos principais métodos multicritério hierárquicos aditivos, o reticulado empregado, $[0,1]$, assume o significado de medida de utilidade esperada. A medida de impacto não significa o mesmo que a de utilidade esperada porque não está ancorada na estrutura de preferências de um decisor racional e maximizador, a medida de impacto depende de um efeito objetivo de mudança em determinado aspecto da realidade e de um juízo de valor social sobre esse efeito.

A medida impacto é diferente da medida da utilidade e é mais abrangente que a medida objetiva de um efeito de mudança. Assim como a própria inovação, a avaliação dos impactos provocados por uma inovação é um fenómeno eminentemente social, sujeito à história e às contradições típicas das relações sociais. Surge daí a sugestão de se estabelecer distinção conceitual entre impacto e efeito de um programa de pesquisa ou nova tecnologia. Efeito da introdução de uma tecnologia seria uma mudança objetivamente mensurável a partir de uma variável quantitativa, enquanto impacto seria uma noção mais ampla que efeito, por incorporar também juízos socialmente construídos e manifestos. Formalmente falando, os impactos conteriam efeitos, mas não necessariamente o inverso.

$$\text{impactos} \subseteq \text{efeitos}$$

$$\text{impactos} \subseteq \text{juízos}$$

A estrutura de avaliação adotada pela metodologia de avaliação de impactos é uma base formal que funciona como uma "rede" cognitiva que indica sobre quais aspectos se deve examinar a extensão dos efeitos de uma determinada nova tecnologia/programa. A

estrutura de avaliação organiza os elementos que permitem a interpretação e o cálculo dos impactos.

Cada dimensão é analisada em uma estrutura própria. A ligação com os aspectos da realidade a que se refere depende do nível contextual da modelagem – o contexto de avaliação, de acordo com os objetivos da avaliação. A construção de uma estrutura de avaliação está condicionada à existência de referenciais conceituais e teóricos que permitam identificar elementos constituintes que tenham *aderência*⁷⁵ ao contexto de avaliação, satisfaçam os *axiomas de funcionalidade*⁷⁶ para a estrutura de avaliação e explicitem as causalidades, correlações ou contingenciamentos existentes entre tais elementos, de modo que se possa definir fórmulas adequadas para a agregação posterior das medidas de impacto obtidas.

As propriedades formais de uma estrutura de avaliação são similares às propriedades das hierarquias de atributos dos métodos multiatributos de decisão. Para sua construção, realiza-se a decomposição dos impactos obedecendo a uma morfologia hierárquica a exemplo dos métodos de utilidade multiatributos. Muda, entretanto, o significado atribuído a seus elementos, como foi comentado acima. As principais diferenças entre as duas situações ocorrem na maneira de executar e interpretar os resultados das metodologias, já que os significados associados à base formal são distintos.

Para a construção das estruturas de atributos de avaliação, o processo consiste no detalhamento das diferentes dimensões iniciais, via decomposição de seus atributos mais gerais em componentes relevantes para o contexto de avaliação. No caso da avaliação de impactos, a partir de cada dimensão D admite-se que ela possa ser analisada em um conjunto de n impactos mais detalhados $I_{D1}, I_{D2}, \dots, I_{Dn}$, que se supõe como um conjunto de componentes capazes de substituir o impacto I_D . Por sua vez, cada i -ésimo elemento I_{Di} pode ser novamente descrito em n elementos $I_{Di1}, I_{Di2}, \dots, I_{Din}$ e assim por diante até um

⁷⁵ A noção de aderência de uma estrutura de impactos é depois definida formalmente como um indicador *ex post* da qualidade da adequação de uma determinada estrutura de impactos a um determinado contexto de avaliação.

⁷⁶ Esses axiomas são similares aos axiomas dos métodos multicritério e apresentados mais abaixo. Esses axiomas definem as propriedades matemáticas básicas da estrutura de impactos e não tem nenhuma ligação com o conceito de utilidade. O conceito de utilidade é fundamentado a partir de um determinado axioma que modela as preferências em funções de valor de tal modo que elas possam ser decompostas em uma hierarquia aditiva de critérios (Cf. ZACKIEWICZ, 2001). Como tanto o conceito de utilidade como o de medida de impacto são modelados sobre a mesma estrutura matemática (o reticulado) as propriedades matemáticas necessárias para relacionar a base formal com os conceitos são similares, mas apenas essas. As formas de interpretar os números obtidos dependem essencialmente dos significados atribuídos a esses reticulados.

limite finito de modo que se forme toda a hierarquia ramificada da estrutura de avaliação. O nível mais desagregado da hierarquia corresponde a seus componentes básicos, identificados, no caso, com impactos a serem mensurados no campo por meio de variáveis apropriadas. A profundidade da desagregação não é pré-definida nem precisa ser a mesma em diferentes ramos da hierarquia.

Uma representação geral de uma estrutura de avaliação que decorre desse procedimento pode ser vista na Figura 4.4. Os índices mostrados em cada elemento permitem reconstruir o caminho de filiação de cada componente em relação à noção de impacto total I_D .

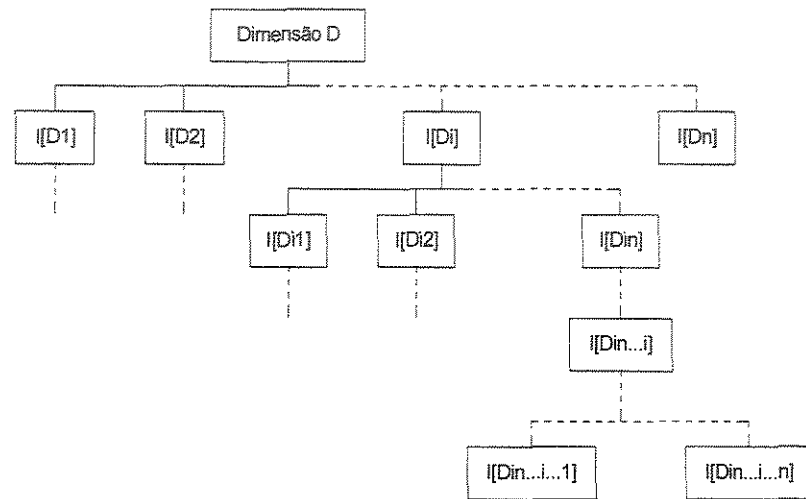


Figura 4.4: Esquema geral de uma estrutura para avaliar impactos.

Numa estrutura de impactos, cada elemento está associado a um determinado aspecto da realidade e possibilita uma interpretação do impacto correspondente a seu respectivo nível de detalhamento.

Uma notação genérica é mostrada a seguir. O nível na hierarquia de impactos (profundidade da desagregação) corresponde à quantidade de índices grafada.

I_{tD} : impacto do programa de pesquisa ou tecnologia t sob o ponto de vista da dimensão D (nível 1);

I_{tDi} : componente do impacto I_{tD} (nível 2);

I_{tDii} : componente do impacto I_{tDi} (nível 3);

...

$I_{D_{L...i}}$: impacto a ser medido (último nível da hierarquia).

A funcionalidade de uma estrutura de avaliação para permitir a *mensuração* e a *síntese* de impactos está ligada à obediência a três propriedades fundamentais, similares às exigências impostas a uma *família coerente* de critérios nos métodos multicritério. Estas exigências são axiomas para a operação matemática da estrutura. Os resultados numéricos a ela associados sempre pressupõem essas propriedades. Portanto, se no modelo contextual construído elas valem por definição, então é preciso ter claras as limitações que disso decorrem quando se substitui a realidade pelo modelo e tomar os cuidados necessários para contorná-las do melhor modo possível.

Partindo das definições de Roy & Bouyssou (1993) e adaptando-as para as estruturas de impactos, tem-se os seguintes enunciados para os axiomas de funcionalidade:

Axioma de exaustividade: *Não existe qualquer novo aspecto da realidade que possa ser considerado, além dos contidos na estrutura de avaliação, tal que, para duas situações a e b, se modifique uma avaliação $I(a)=I(b)$ inicialmente obtida.*

Esse axioma garante que a estrutura de avaliação é suficiente para captar os impactos, uma vez que não existiriam novos impactos capazes de distinguir duas situações de impactos idênticos.

Na prática, a decisão sobre a exaustividade efetiva, ou seja, sobre a quantidade de aspectos cobertos por uma estrutura de avaliação depende do contexto avaliado e dos objetivos da avaliação. Como essa decisão é externa ao modelo, ela é invisível a ele e não há como medir se uma estrutura é exaustiva a partir dela própria. Da mesma forma, não faz sentido atribuir ao modelo sua incapacidade de medir impactos não previstos inicialmente. A inclusão de componentes que garantam a exaustividade faz parte da tarefa de construção da estrutura de avaliação e depende do processo de aprendizado subjacente e pressuposto na aplicação da metodologia de avaliação.

Axioma da não redundância: *A estrutura de avaliação não contém atributos que possam ser retirados sem que ela contrarie o axioma de exaustividade.*

Esse axioma completa o anterior ao impor limites para a quantidade de atributos pertinentes ao modelo. Esse limite é dado pela necessidade de eliminar todos impactos além daqueles que garantam a exaustividade.

Para se fazer a agregação dos impactos por um procedimento aditivo, esse axioma precisa ser mais restritivo e incluir a independência entre impactos. Isso porque, do contrário, a soma dos impactos passa a sobrepô-los. Entretanto, nada impede que uma mesma medida seja contabilizada mais de uma vez, desde que esteja se referindo a impactos distintos, isto é, estabeleça mais de um nexos causal – conceitualmente independentes entre si – no modelo.

Axioma de coesão: Existe uma única função que relaciona uma mudança 'x' observada em determinado aspecto da realidade a uma medida de impacto I.

O axioma da coesão possibilita a ponte matemática entre uma medida empírica de variação observada no campo e seu respectivo componente na estrutura de avaliação. A unicidade dessa relação possibilita coerência comparativa entre diferentes tecnologias (desde que avaliadas pela mesma estrutura de impactos) e sua existência permite a realização de operações de síntese com os números encontrados para os componentes, desde que todos sejam obtidos a partir de funções que levem as diferentes medidas de campo à mesma escala.

Essas funções, por serem únicas, não conseguem lidar com a possibilidade de existência de ambigüidades na medida de impacto. Do ponto de vista do modelo hierárquico, todos os valores nele introduzidos são considerados como únicos e representativos. A questão da qualidade e representatividade dessas medidas precisa ser resolvida externamente, como será visto adiante.

VIII. AS VARIÁVEIS ASSOCIADAS A UMA ESTRUTURA DE AVALIAÇÃO

A estrutura de atributos definida para o método MDM, e utilizada para a avaliação de impactos, se torna uma estrutura de avaliação pela associação de três variáveis: a medida de impacto (I), a variação observada no componente básico (x) e a participação da pesquisa/tecnologia nesse efeito observado (α). Além delas, existem as funções ϕ e ρ responsáveis por levar x e α à escala da medida de impacto I. Por fim, coeficientes auxiliares podem ser necessários para realizar a síntese dessas variáveis na estrutura de avaliação. As funções e coeficientes auxiliares são constantes para uma determinada

estrutura de avaliação, ajudando a caracterizar matematicamente os significados conceituais e critérios nela incorporados.

A medida de impacto

Para realizar operações de síntese de impactos todos os componentes de uma estrutura de avaliação são medidos sobre uma escala similar. A variável I é definida sobre os intervalos contínuos dos números reais $[-1,0]$ e $[0,1]$, de modo que os extremos signifiquem:

- $I = 1 \Rightarrow$ máximo impacto positivo (desejável);
- $I = -1 \Rightarrow$ máximo impacto negativo (indesejável);
- $I = 0 \Rightarrow$ inexistência de impacto.

Os dois intervalos que compõem a escala são dois reticulados espelhados e coincidentes no ponto 0, isto é, na situação de inexistência de impacto. Na prática, essa estrutura pode ser simplificada para um reticulado $[-1, 1]$, de modo a viabilizar um modelo aditivo, por exemplo. Entretanto, a separação em dois reticulados independentes é necessária para a identificação de ambigüidades entre os participantes da avaliação e para a determinação do índice de coesão, como será discutido adiante.

Por definição, essa escala de impactos incorpora duas características importantes. A primeira é que o significado da escala inclui a uma só vez a medida de efeitos decorrentes da pesquisa e os juízos sobre esses efeitos e a segunda é que são definidos os limites de máximo positivo e máximo negativo para esse juízo, passando por um estado intermediário de indiferença.

Isso permite que se projete sobre a estrutura de avaliação qualquer variável x , medida em qualquer escala ou unidade desde que se defina uma função $\phi: x \rightarrow [-1,1]$ que leve as medidas dessa variável adequadamente para a escala de impacto. A variável x , ou variação observada no componente básico, é a porta de entrada da estrutura de avaliação por onde o universo empírico é captado. A natureza da escala de impactos pressupõe que a medida de x mescle uma medida da variação em um determinado aspecto da realidade – um

feito – com um juízo sobre a importância dessa variação em termos de impactos, dada pela definição de correspondência estabelecida entre a escala utilizada para medir x e os extremos da escala de impactos. As dificuldades que emergem dessa característica são tratadas em sessão dedicada à medida dos impactos no campo, uma vez que essas não interferem nas propriedades formais da estrutura de avaliação.

Uma vez finalizada a etapa de decomposição dos impactos e verificada a obediência aos axiomas, a fórmula mais simples e geral para novamente agregar os valores das medidas das variações x dos aspectos da realidade avaliada – ou seja, do contexto de aplicação do programa/tecnologia t – é a adição ponderada a partir dos componentes básicos da estrutura de impactos até a medida de síntese do impacto para a dimensão D .

$$I_t = \sum_{i=1}^{Dn} k_{Di} \dots \left(\sum_{i=1}^{Dn \dots n} k_{Di \dots i} \cdot \varphi_{Di \dots i}(x_{tDi \dots i}) \right) \quad \text{Equação 4.1}$$

Os pesos k da Equação 1 são parâmetros constantes que definem a importância relativa dos impactos no contexto para o qual a estrutura de avaliação for definida. Nesse caso, é necessário distinguir a importância do impacto, dada pelo peso, e a medida do impacto, dada pela variação captada no campo. O peso faz parte da definição da "malha" cognitiva que referencia a interpretação dos impactos. A magnitude final do impacto é resultante da combinação desses dois fatores: importância e variação. Para um mesmo contexto de avaliação, a diferença entre as medidas de impacto para dois diferentes programas/tecnologias será decorrente apenas de diferentes medidas da variável x , uma vez que os pesos dos impactos componentes da estrutura de impactos serão os mesmos para os dois casos.

Os pesos k são definidos dentro do intervalo contínuo $[0,1]$, tal que para cada ramo de uma estrutura de impactos com n componentes:

$$\sum_{i=1}^n k_i = 1 \quad \text{Equação 4.2}$$

A atribuição dos impactos

Roche (2000, p. 43) interpreta a noção de impacto como o final de uma cadeia causal iniciada com determinados insumos que viabilizam determinadas atividades, que levam a determinados resultados e só então causam (ou não) impactos. Uma vez que tais impactos se referem a mudanças mensuráveis para além do programa ou projeto que gerou os resultados, eles também sofreriam grande influência dos contextos social, econômico, ambiental e político adjacentes.



Figura 4.5: Cadeia causal da geração de impactos.

Disso decorre que nem sempre é possível ter certeza que dado um contexto de avaliação a variação medida para um determinado componente da estrutura de impactos possa ser integralmente atribuída à existência da nova tecnologia/programa, podem sempre existir outras causas gerando mudanças nos mesmos aspectos que interessam do ponto de vista do programa/tecnologia avaliado, caracterizando uma situação multicausal tratada indistintamente como impacto pelo modelo (Figura 4.6).

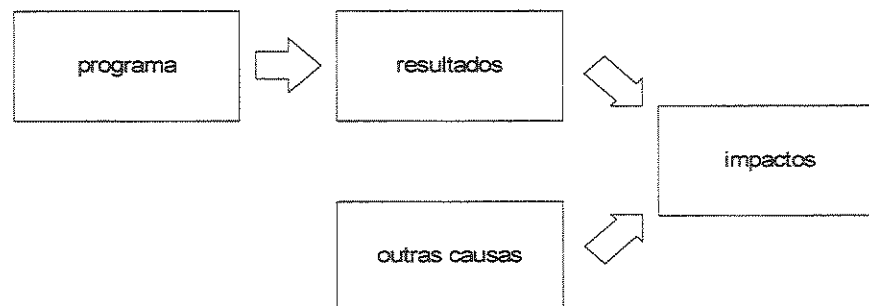


Figura 4.6: A múltipla causalidade dos impactos.

A variável de participação da pesquisa/tecnologia (α) procura separar a fração da medida de impacto I atribuível efetivamente à pesquisa/tecnologia da fração remanescente atribuível a outras causas. Essa variável é medida em complemento a cada x e introduzida no modelo por uma função $\rho: \alpha \rightarrow [0,1]$, de modo que:

- $\rho(\alpha) = 0$ significa que a variação x observada no componente de impacto não pode ser atribuída à tecnologia avaliada, ou seja, $\neg(x \Rightarrow I)$;
- $\rho(\alpha) = 1$ significa que a causalidade $x \Rightarrow I$ é plenamente aceitável;
- $0 < \rho(\alpha) < 1$ significa que ocorre um certo grau de incerteza na atribuição devido à relação apenas parcial entre a introdução da tecnologia e a variação x observada.

A escala de valores de $\rho: \alpha \rightarrow [0,1]$ modela a incerteza da causalidade $x \Rightarrow I$ segundo o princípio da lógica *fuzzy*. A função ρ faz o papel de uma função de pertinência definida no intervalo dos números reais $[0,1]$, ampliando a noção de verdade da lógica clássica que é definida pelo conjunto de elementos discretos $\{0,1\}$. Assim, o modelo pode incluir a relação de causalidade entre a variação observada para um determinado aspecto da realidade e a medida de impacto como uma variável que regula a variável I e não simplesmente a contabiliza ou ignora. Não se trata de uma medida probabilística da relação causa-efeito, trata-se efetivamente de um julgamento subjetivo da plausibilidade dessa relação ser verdadeira para cada componente de impacto.⁷⁷

Do ponto de vista da lógica *fuzzy*, a medida do impacto I_p – efetivamente atribuído ao programa/tecnologia – torna-se então decorrente de duas premissas, a existência e a importância da mudança x em determinado aspecto da realidade e o grau de participação α da pesquisa/tecnologia nessa mudança.

A t-norma⁷⁸ $\Delta: [-1,1] \times [0,1] \rightarrow [-1,1]$ permite relacionar essas duas premissas, tal que:

$$\varphi(x) \Delta \rho(\alpha) = \varphi(x) \cdot \rho(\alpha) = I_p \quad \text{Equação 4.3}$$

⁷⁷ Os impactos não causados pela tecnologia são descontados mas não são explicados pelo modelo. Para explicá-los, quando for o caso, é necessário recorrer a informações complementares obtidas no contexto de avaliação, porém externas ao modelo de cálculo.

⁷⁸ As t-normas (normas triangulares) são relações binárias que realizam um papel similar ao conectivo "e" na lógica clássica (NGUYEN & WALKER, 2000, pg. 83).

Na Figura 4.7, é possível observar graficamente como a t-norma Δ relaciona as três variáveis associadas à estrutura de avaliação de impactos.

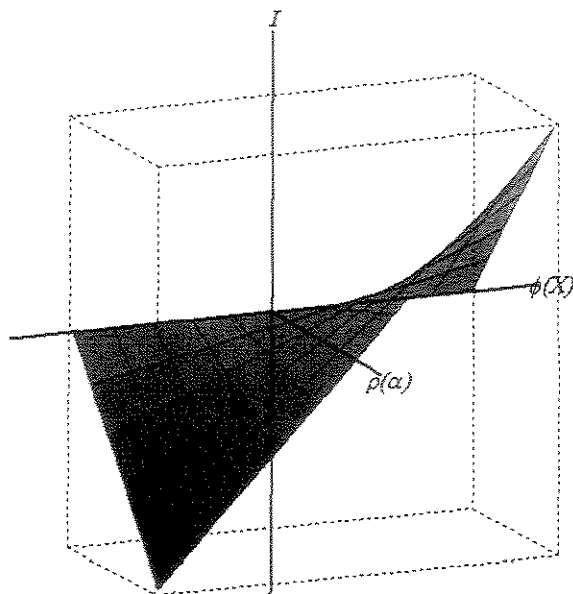


Figura 4.7: Representação gráfica da Equação 4.3.

A incorporação desse princípio foi inspirada na noção de pseudocritério da escola francesa de métodos multicritério de apoio à decisão (ver Capítulo 2 e Anexo 1). Recorreu-se a uma definição de t-norma que permitisse o desconto progressivo da importância do impacto I a partir da medida de plausibilidade da relação $x \Rightarrow I$.⁷⁹ Esse procedimento ajuda a tornar o cálculo de impactos mais realista e a evitar sub ou superestimativas do impacto.

A expressão geral para o cálculo dos impactos em um contexto de aplicação de uma tecnologia t , considerando agora a fração efetivamente atribuída à existência da pesquisa/tecnologia, se redefine em:

⁷⁹ Em alguns dos métodos multicritérios de apoio à decisão da família ELECTRE há um enfraquecimento progressivo da discernibilidade na comparação entre dois objetos segundo um critério, que se torna, então, um pseudocritério (ou critério difuso). Cf. VALLÉ & ZIELNIEWICZ (1994).

$$I_{IP} = \sum_{i=1}^{Dn} k_{Di} \dots \left(\sum_{i=1}^{Dn \dots n} k_{Di \dots i} \cdot \varphi_{Di \dots i}(x_{tDi \dots i}) \cdot \rho_{tDi \dots i}(\alpha_{tDi \dots i}) \right) \quad \text{Equação 4.4}$$

A fim de evitar confusão de nomenclatura, o impacto grafado na Equação 4.1, que se refere ao impacto global observado no contexto de t passa a ser chamado também de *impacto geral* e escrito como I_G .

$$I_{IG} = \sum_{i=1}^{Dn} k_{Di} \dots \left(\sum_{i=1}^{Dn \dots n} k_{Di \dots i} \cdot \varphi_{Di \dots i}(x_{tDi \dots i}) \right) \quad \text{Equação 4.1b}$$

A diferença entre I_G e I_{IP} é a fração do impacto geral atribuída a outras causas que não a existência da pesquisa/tecnologia t , denominada de *impactos de outras causas* e representada por I_{IO} . Por definição:

$$I_{IO} = I_{IG} - I_{IP} \quad \text{Equação 4.5}$$

No caso geral de síntese dos impactos via as equações aditivas acima, I_{IO} fica:

$$I_{IO} = \sum_{i=1}^{Dn} k_{Di} \dots \left(\sum_{i=1}^{Dn \dots n} k_{Di \dots i} \cdot \varphi_{Di \dots i}(x_{tDi \dots i}) \cdot (1 - \rho_{tDi \dots i}(\alpha_{tDi \dots i})) \right) \quad \text{Equação 4.6}$$

O cálculo de agregação que leva à síntese dos impactos não precisa ser necessariamente aditivo. Ele depende das definições dos componentes da estrutura de impactos. Quando houver interdependência entre componentes em um determinado ramo da estrutura, a medida de síntese para o componente imediatamente acima na hierarquia precisa levar em conta essa interdependência e a modelagem da equação passa a ser realizada caso a caso.

IX. A OBTENÇÃO DAS MEDIDAS

Do ponto de vista formal, as medidas de campo não trazem maiores problemas para as operações da estrutura de avaliação de impactos e suas variáveis, bastando estarem enquadradas nas propriedades definidas acima. Entretanto, tais medidas são cruciais para o corpo de significados associados às estruturas de impactos e para o resultado da avaliação de impactos em si.

Os modos de medir x e α e suas respectivas funções de introdução na estrutura de impactos não são escolhas independentes. Diferentes modos de medida são possíveis e levam a diferentes concatenações entre efeitos e juízos para a composição de uma medida de impacto.

Dependendo da natureza do aspecto da realidade referente a um componente básico da estrutura de impacto definida, as variáveis x e α associadas a ele podem ser medidas por *escalas qualitativas ordinais* ou por *escalas quantitativas*. Essas medidas são em geral feitas por meio de pesquisa de campo, seja por entrevistas, por medidas instrumentais diretas, por contagens etc. Quando possível podem ser obtidas também de dados secundários.

Escala Qualitativa Ordinal: utiliza medidas categóricas como as das escalas de Likert.⁸⁰ A medida sensorial obtida por essa escala se justifica quando há impossibilidade de medidas instrumentais ou quando o custo dessas medidas é proibitivo. As medidas a partir de escalas de Likert são também chamadas *psicométricas* e são indissociáveis de idiossincrasias do indivíduo respondente (tomado nesse caso como um instrumento de medida).

Escala Quantitativa: utiliza medidas discretas, intervalares ou contínuas coletadas por instrumentos, contadas ou calculadas a partir de uma unidade de medida numérica apropriada.

A escolha das escalas para medir as variáveis x e α definem dois modos principais para captar os impactos:

Modo Direto: Obter x e α a partir da avaliação sensorial sobre as mudanças ocorridas. Os extremos das escalas de Likert utilizada para medir as variáveis

⁸⁰ Escalas com medidas categóricas baseadas em juízos de intensidade ordenada e em oposição semântica em torno de um ponto médio.

qualitativamente podem diretamente definir os extremos de escala da medida de impacto do componente básico na estrutura de avaliação e seu ponto médio ser equivalente à situação de ausência de impacto. Isso traz a vantagem de referenciar a medida de impacto a juízos sobre a intensidade das mudanças dados pelos atores inseridos no contexto de aplicação da pesquisa/tecnologia.

Modo Indireto: Obter x a partir de medidas quantitativas diretas da variação dos indicadores. A conversão para os extremos de escala da medida de impacto associada e a medida de α é definida por especialistas. Nem sempre esses especialistas estão inseridos no contexto de aplicação da pesquisa/tecnologia.

A principal diferença entre essas duas formas está na medição conjunta ou disjunta dos efeitos e dos juízos sobre extensões limítrofes dos impactos. No primeiro modo, trabalha-se diretamente com o juízo (medida sensorial) dos atores impactados para medir a variação x e a participação α do programa/tecnologia na variação observada – sem passar necessariamente por alguma quantificação dos impactos. No segundo modo, a variável x é primeiro quantificada para, em seguida, ser cotejada frente a um referencial dado externamente – por especialistas – e só a partir daí adquirir seu significado e localização em relação aos extremos da medida de impacto.

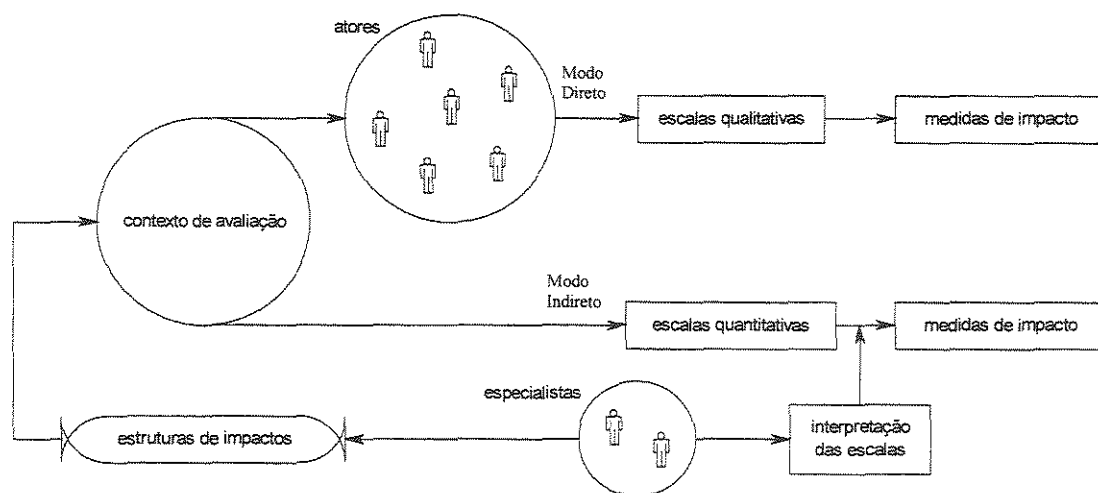


Figura 4.8: Modos Direto e Indireto para obtenção de medidas de impacto.

Os dois modos de medida não necessariamente indicarão o mesmo valor de impacto. Entretanto, será possível falar em uma medida correta, objetivamente definida para impacto? Um possível conflito entre as duas "racionalidades" subjacentes aos dois modos que levaria à indeterminação do impacto pode ser exemplificado na situação sugerida a seguir no Quadro 4.1. O exemplo considera um mesmo atributo *A* da realidade cuja variação *x* num determinado período de tempo tenha sido medida por diferentes escalas em duas avaliações distintas (p.e. numa avaliação comparativa entre duas tecnologias diferentes que impactaram atores em contextos para os quais a mesma estrutura de avaliação é aplicável):

<i>Casos</i>	<i>Variação de x qualitativa (juízo agregado dos atores)</i>	<i>Variação de x quantitativa</i>	<i>Variação de x relativa</i>
Programa/Tecnologia 1	Grande aumento	1.000	100%
Programa/Tecnologia 2	Aumento moderado	10.000	10,00%

Quadro 4.1: Exemplo de indeterminação aparente entre diferentes escalas de medida e impacto.

Se todos os indivíduos entre os entrevistados no universo de atores impactados e especialistas possuísem uma racionalidade perfeita seria possível – por definição – obter uma medida objetiva dos impactos, independentemente do modo de medida escolhido. Nessa situação hipotética, todo respondente no Modo Direto poderia ser visto como um instrumento preciso de medida capaz de prescindir de uma medida quantitativa em cada indicador para assim mesmo fornecer um juízo relativizado ao contexto de avaliação suficientemente representativo de sua situação particular em relação à geral. Por extrapolação, nesse caso seria possível até encontrar o valor quantificado a partir de um referencial externo. No caminho inverso, a inferência sobre como os impactados interpretam e sentem os impactos a partir da obtenção de dados quantitativos (examinados e valorados por especialistas) necessita da hipótese de racionalidade perfeita e da uniformidade dos atores.

Como definir então o impacto "correto" na escala da função $\phi: x \rightarrow [-1, 1]$? Cada tipo de escala impõe um tipo específico de distorção quando tomada isoladamente. Porém, não se justifica afirmar que há inconsistência nos valores mostrados no Quadro 4.1. A relação

entre juízos qualitativos, valores quantitativos e valores relativos depende do estado inicial do aspecto A (antes da existência dos programas/tecnologias 1 e 2) e também da dinâmica histórica desse aspecto no contexto ao qual os atores entrevistados fazem parte. Assim, para o caso 2, um aumento de 10%, significando 10.000 unidades de algo, é moderado, mas não é pequeno. Enquanto para o caso de introdução da tecnologia 1, 1.000 unidades equivalem a 100% e significam um grande aumento. A inconsistência aparente emerge ao se procurar correlação entre as comparações de 1 e 2 para as diferentes escalas. A comparação sempre precisará considerar as dinâmicas históricas e as diferentes características dos contextos aos quais se refere a tecnologia introduzida.

X. A ESCOLHA DOS VALORES REPRESENTATIVOS

Uma vez realizadas as medidas de campo para variação x e participação α em todos os componentes básicos da estrutura de avaliação e aplicadas as funções de transformação para as medidas de impacto ϕ e para a atribuição ρ diferentes padrões podem emergir dessas medidas. No caso de medidas qualitativas e quantitativas discretas ou intervalares, esses padrões podem ser caracterizados em uma *tabela de contingência*.⁸¹ Os diferentes tipos de distribuição de freqüências possíveis podem ser associados a diferentes *graus de coesão* das medidas.

Define-se a coesão (c) de uma distribuição de freqüências como "total" se todo valor para x e α para um componente básico $Di...i$ obtido numa amostra de a elementos forem iguais e constantes.

$$\begin{aligned}\forall (x_{tdi...i})_a = cte &\Leftrightarrow c[\phi_{tdi...i}(x_{tdi...i})] = total \\ \forall (\alpha_{tdi...i})_a = cte &\Leftrightarrow c[\rho_{tdi...i}(\alpha_{tdi...i})] = total\end{aligned}$$

Nessas condições os valores para $\phi(x)$ e $\rho(\alpha)$ são também constantes e correspondem diretamente aos *valores representativos* que devem ser inseridos na estrutura de avaliação. Entretanto, em situações da prática, essa condição quase sempre não se

⁸¹ Uma tabela de contingência organiza a freqüência medida para cada categoria disponível na escala (nominal ordinal, valores numéricos discretos ou valores numéricos intervalares) frente a todos os componentes básicos da estrutura de impactos.

verifica e é necessário escolher um valor representativo para a distribuição de valores obtida.

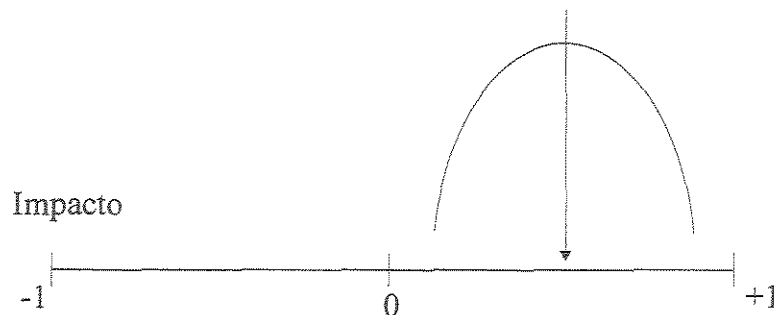
A dispersão das medidas pode ocorrer por erros de calibração dos instrumentos de medida, variações nos procedimentos de medidas e uso diferenciado das escalas qualitativas. Mas também pode ocorrer porque o aspecto medido realmente tem manifestações diferentes em situações diferentes que coexistem no mesmo contexto de avaliação, ou seja, porque o impacto se manifesta diferentemente para diferentes elementos da amostra avaliada. Em geral, a dispersão do primeiro tipo mantém as medidas de impacto $\phi(x)$ ou do lado positivo ou do lado negativo da escala $[-1,1]$ havendo apenas variação de intensidade. Se ocorrer inconsistência, isto é, o aparecimento simultâneo de valores positivos e negativos para um mesmo $\phi(x)$ isso pode indicar que o impacto tem manifestação oposta em condições específicas. Dispersões tendendo ao aleatório podem indicar má qualidade do instrumento de medida (p.e. da questão utilizada na entrevista) ou elementos amostrais de baixa qualidade (p.e. pessoas inabilitadas a responder as entrevistas).

Baseado no grau de coesão, foi adotado um dispositivo para a escolha dos valores representativos. Foram definidos diferentes graus de coesão a partir do formato da distribuição das frequências das medidas de impacto e atribuição. São levados em conta o fato de ocorrer ou não inconsistência (na distribuição de frequências de $\phi(x)$) e se a distribuição é ou não convexa (para $\phi(x)$ e $\rho(\alpha)$). Para cada tipo de coesão escolheu-se um procedimento de escolha tido como mais adequado, conforme justificado a seguir. Para fins do cálculo do grau de coesão, nos casos em que um componente básico não se aplicar – e a variação é substituída pela categoria N/A – é contabilizado o valor nulo para a medida de impacto.⁸²

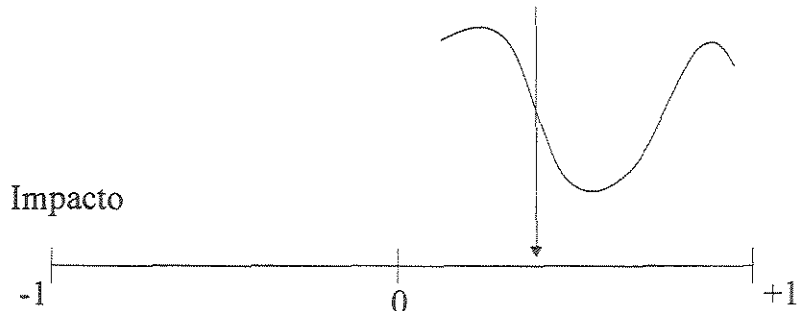
Caso 1: Coesão Forte ($c(x)=c(\alpha)=1$). Ocorre quando não há inconsistência e a distribuição das frequências é convexa. Trata-se do caso mais bem comportado, no qual a variação observada na distribuição provavelmente é tributável a variações aleatórias na obtenção dos dados. O valor representativo da amostra escolhido por esse critério para

⁸² O fato de um componente não ser aplicável revela imperfeição de aderência de uma estrutura de impactos a um contexto específico e também significa que não ocorre o impacto a que ele se refere. Do ponto de vista da medida de impacto, as duas situações são idênticas. Contabilizar ambas como $I = 0$ é necessário para manter a comparabilidade de duas avaliações que utilizem uma mesma estrutura de avaliação.

entrar no cálculo dos impactos é definido como a média de todas medidas obtidas para $\phi(x)$ ou $p(\alpha)$.

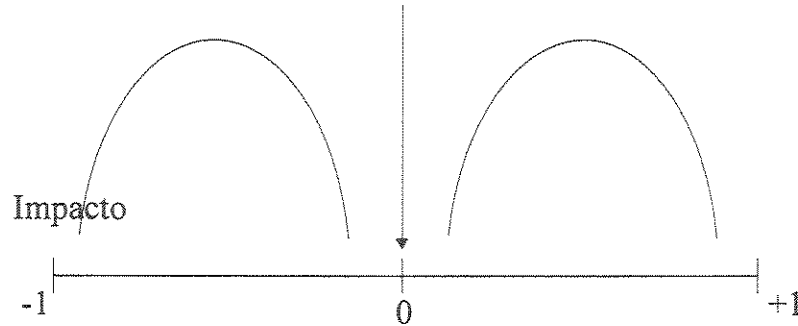


Caso 2: Coesão Fraca ($c(x)=0,75$; $c(\alpha)=0$). Ocorre quando a distribuição é consistente e não convexa. Neste caso, fica mais difícil saber se a distribuição está ou não em tendência para um valor central (já considerando a exclusão prévia de *outliers* e valores extremos) ou se haveria duas situações igualmente válidas tomando forma. Uma solução para definir um valor representativo é tomar partido da situação preponderante, adotando a mediana dos dados (assim o valor representativo estaria mais próximo da moda do que a média, no caso de distorção causada por algum subgrupo nas extremidades próximas a 0 ou a ± 1 , a mediana atenua sua importância).

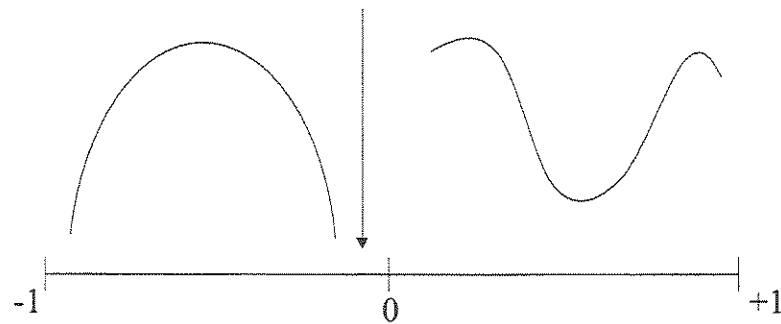


Caso 3: Coesão Dupla ($c(x)=0,5$) Ocorre quando a distribuição é inconsistente porém convexa nas duas extremidades – tomadas separadamente e incluindo os casos de $\phi(x)=0$ em ambas. É razoável supor que, descontados os *outliers*, ocorrem duas situações distintas e não apenas por conta de flutuações em torno da intensidade do aumento ou da diminuição, mas efetivamente controversa entre aumento ou diminuição. O valor representativo é dado pela média das médias de cada sessão convexa. A sobreposição dos valores nulos torna-se uma medida de cautela, pela incerteza em relação a baixa adequação do instrumento de medida utilizado. Se o tratamento das respostas isolar atores e categorias

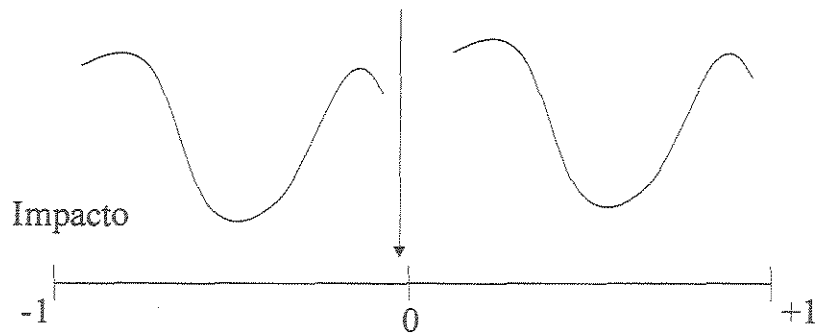
de atores ou retirar da amostra as respostas que causam a inconsistência, retorna-se para o primeiro caso e a distorção proposital em torno do zero desaparece. Se isso for possível, ganha força a hipótese de manifestações distintas do impacto para condições distintas.



Caso 4. Coesão Assimétrica ($c(x)=0,25$). Ocorre quando há inconsistência e apenas um lado do histograma de distribuições é convexo. Neste caso toma-se a média entre a média do lado convexo e a mediana do lado não convexo, buscando a combinação das propriedades justificadas nos dois casos anteriores.



Caso 5. Ausência de Coesão ($c(x)=0$). Ocorre quando há inconsistência e nenhum lado é convexo. Isso equivale a respostas quase aleatórias e a baixa adequação do instrumento de medida ou dos elementos da amostra utilizada. Neste caso, adota-se a média das medianas. Com uma distribuição de frequências tendendo a uniforme, esse critério faz o valor representativo tender a zero, anulando a medida do componente de impacto.



Pode-se verificar que cada tipo de coesão corresponde a uma faixa de variância dos dados medidos por variáveis categóricas, com sucessivas sobreposições. Portanto, o corte por coesão apesar de não levar em consideração diretamente uma medida de dispersão incorpora em alguma medida essa informação. Uma consequência é de que com a ampliação da amostra a coesão tende a crescer e depois se estabilizar. Pode-se derivar daí uma medida para a suficiência da amostra utilizada em uma avaliação.

XI. INDICADORES DE QUALIDADE DA ESTRUTURA DE AVALIAÇÃO

Os padrões característicos das medidas obtidas em campo permitem que se definam indicadores de qualidade para medidas de impacto calculadas.

Coesão das Respostas (Z)

$$Z = \frac{\sum_{i=1}^n c(x_i)}{n}$$

Equação 4.7

Na Equação 4.7, o parâmetro n é a quantidade de componentes básicos de uma estrutura de avaliação e $Z: \{0; 0,25; 0,5; 0,75; 1\} \rightarrow [0,1]$.

Quanto maior Z , mais coeso é o conjunto de medidas de variação x obtido. Pode-se verificar que $Z = 1$ sempre que esse conjunto contiver apenas um elemento. As medidas de impacto inseridas em uma estrutura de avaliação incorporam mais ou menos ambigüidade tanto menor ou maior for Z .

Pode-se encontrar as maiores coesões a partir do teste com diferentes combinações entre os elementos do conjunto de respondentes e assim encontrar subconjuntos latentes com elevado nível de consenso entre si.

Coesão da Atribuição (Z_α)

$$Z_\alpha = \frac{\sum_{i=1}^n c(\alpha_i)}{n} \quad \text{Equação 4.8}$$

Na Equação 4.8, o parâmetro n é a quantidade de componentes básicos de uma estrutura de avaliação e $Z_\alpha: \{0;1\} \rightarrow [0,1]$.

Quanto maior Z_α , mais coeso é o conjunto de medidas de participação α obtido. Pode-se verificar também que $Z_\alpha = 1$ sempre que esse conjunto contiver apenas um elemento. As medidas de impacto inseridas em uma estrutura de avaliação são mais ou menos controversas quanto à atribuição quanto menor ou maior for Z_α .

Aderência da Estrutura de Avaliação (A)

$$A = 1 - \frac{\sum(NA)}{\sum_{total}} \quad \text{Equação 4.9}$$

Na equação 4.9, $\sum(NA)$ é a contagem total da categoria "não se aplica" obtida da soma de todas as vezes que não foi possível obter medidas para os componentes da estrutura de avaliação numa dada avaliação e $\sum_{total}$ é o número total de respostas (x) para todas os componentes da estrutura.

A aderência da estrutura de avaliação mede o quanto seus componentes básicos foram adequados para captar impactos em um dado contexto de avaliação. É fácil verificar que A também está definido sobre $[0,1]$. Sua interpretação é: $A = 0$ para nenhuma aderência e $A = 1$ para aderência total, com os valores intermediários indicando situações proporcionais a esses dois extremos.

A medida da aderência é importante para verificar a adequação de uma determinada estrutura – definida para um certo contexto de avaliação – para avaliar um determinado objeto. Por consequência, comparar duas avaliações para as quais as aderências foram muito diferentes é arriscado.

Aderência da Atribuição (A_α)

$$A_\alpha = \frac{\sum_{\alpha(\alpha)=1, j=1}^{L, n} (\varphi(\alpha), \Sigma(\varphi(\alpha), j))}{\Sigma_{total}} \quad \text{Equação 4.10}$$

Na equação 4.10, A_α é calculado pela soma dos valores obtidos da multiplicação entre os valores assumidos pela função φ a partir das diferentes categorias de medida da variável α e a frequência Σ desses mesmos valores, sobre o número total de respostas para todas os componentes da estrutura de avaliação, Σ_{total} .

O indicador de aderência da atribuição dá a medida $[0,1]$ de quanto os componentes básicos da estrutura de avaliação foram capazes de distinguir no impacto geral medido o impacto do programa/tecnologia avaliado. É esperado que sempre $A_\alpha \leq A$. Essa medida se distingue da relação I_P / I_G porque é obtida do ponto de vista dos componentes básicos, ou seja, do ponto de vista do instrumento de medida, enquanto essa última relação fornece uma medida do ponto de vista da avaliação global de impacto.

XII. ASPECTOS PRÁTICOS DA AVALIAÇÃO DE IMPACTOS COM O MÉTODO MDM

Esta sessão organiza os passos necessários para a implementação de uma avaliação de impactos com o método MDM. A intenção é comentar de modo sequencial o processo de construção da avaliação para que fiquem mais evidentes os principais cuidados necessários e se discutam diversos aspectos práticos importantes. A inclusão de exemplos emprestados do teste prático de construção da metodologia de avaliação de impactos para as

tecnologias oriundas dos casos dos Programas Citros e Procana do IAC vai ilustrar seu uso.⁸³

Os passos gerais para a aplicação do método de avaliação são:

a. *delimitação do objeto de avaliação*

- descrever o contexto no qual se insere o objeto de avaliação;
- delimitar o objeto de avaliação no tempo e espaço;
- identificar especificidades setoriais, institucionais, sócio-culturais etc. do contexto de avaliação;

b. *verificar as dimensões e identificar atores relevantes*

- justificar a pertinência das diferentes dimensões para o objeto de avaliação segundo os objetivos motivadores da avaliação;
- identificar os atores relevantes por cadeias produtivas ou outros recortes;
- definir atores impactados por dimensão e categorizá-los;
- definir amostra intencional de representantes dos diferentes atores e categorias;

c. *construção das estruturas de avaliação*

- identificar especialistas nas dimensões escolhidas para o contexto da avaliação;
- construir estruturas de atributos para o contexto de avaliação;
- justificar conceitualmente a arquitetura das estruturas de avaliação;
- escolher as equações de agregação e definir seus parâmetros;

d. *definição e preparação dos instrumentos de medida de campo*

- escolher escalas de medida para as variáveis x e α ;
- definir as funções ϕ e ρ ;
- compor o questionário de campo;
- testar os instrumentos de medida;

e. *obtenção das medidas de campo*

- realizar entrevistas e medidas instrumentais no campo;

f. *avaliação dos dados*

- utilizar o algoritmo para o cálculo dos resultados;

⁸³ Todos os exemplos reproduzidos neste Capítulo foram obtidos de FURTADO *et al.* (2003).

- verificar a coesão dos resultados;
- realizar as partições necessárias no conjunto de respostas;
- avaliar os resultados dos cálculos conjuntamente com as demais informações disponíveis para o contexto de avaliação;
- fornecer retorno a todos os envolvidos no processo;
- retomar os passos anteriores de modo a aperfeiçoar o modelo.

O contexto de avaliação

Para a construção de uma avaliação com o método MDM é necessário fixar as referências que definem o contexto de avaliação e reunir informações gerais para a caracterização mais detalhada possível desse contexto:

i. *O objeto da avaliação.* O primeiro passo para a construção da metodologia é a definição do objeto motivador da avaliação. Em geral, uma nova tecnologia está associada a um programa de pesquisa e este, por sua vez, a uma ou mais organizações voltadas à pesquisa. Nem sempre é fácil dissociá-los e pode ser mais interessante do ponto de vista da inovação tomar o conjunto tecnologia, mais programa de pesquisa, mais organização de pesquisa como o objeto de avaliação.

ii. *Os lugares e os respectivos atores* relacionados com o objeto. Esse recorte pode ser feito a partir da análise da(s) cadeia(s) produtiva(s) e das redes de atores associadas ao uso da tecnologia em questão. No caso de impactos, é sabido que esses não necessariamente se manifestam homogeneamente nem com a mesma nitidez em diferentes pontos das cadeias produtivas e para diferentes atores. A definição desse universo é fundamental para o desenho amostral e para as subsequentes estratégias de obtenção de medidas de impacto. A *amostra intencional* escolhida deve representar o melhor possível a variação dos atores quanto a situações particulares que possam influenciar na intensidade em que se manifeste o impacto. Os atores podem ser sub agrupados ainda em diferentes categorias. Na amostragem, o número de elementos de cada categoria deve ser aproximadamente o mesmo.

iii. *A janela temporal.* Considerando a avaliação de impactos como uma atividade *ex post*, a escolha direta para o intervalo de tempo a ser analisado é daquele entre a

introdução da nova tecnologia e o momento em que ocorre a avaliação (presente). Avaliando outras opções, possíveis em tese, percebe-se que na prática a medição da variação dos indicadores, a menos que se estabeleçam procedimentos sistemáticos de monitoramento, é sempre facilitada se referenciada à situação anterior à introdução da nova tecnologia (conhecido na literatura como *base line*). Na realidade, a manifestação dos impactos não é instantânea, homogênea (entre indicadores) e nem constante (e muitas vezes dificilmente modelável no tempo) a partir da introdução da tecnologia. Assim, a medição em um tempo t_1 após a introdução em t_0 reflete o estado dos impactos naquele momento, mas nada indica sobre a variação da variação no tempo (derivada segunda no tempo). E caso se resolva, passado mais tempo, reavaliar os impactos em t_2 , do ponto de vista metodológico, será mais fácil fazer novamente em relação a t_0 , que é um momento mais bem definido que o contingente t_1 . Repetindo mais vezes a avaliação poderão surgir padrões temporais da evolução do impacto que permitam inclusive projeções do comportamento futuro dos impactos.

Essas definições do contexto de avaliação precisam ser anteriores ao desenho (ou à decisão de adoção) de uma estrutura de avaliação. Além disso, as informações do contexto da avaliação são importantes para dar ao avaliador e para os usuários da avaliação parâmetros para interpretar o significado dos impactos medidos.

As estruturas de avaliação

As estruturas de avaliação são o suporte matemático que possibilita a integração de todos os atributos avaliados em um marco integrado e logicamente articulado. Uma estrutura de avaliação é composta por diversos níveis hierarquizados de modo que a dimensão corresponda tanto a noção mais sintética como a medida mais agregada sobre o objeto. Quanto mais se desce na hierarquia de componentes, mais precisas se tornam as definições acerca dos efeitos que importam medir para a avaliação. Os últimos atributos desagregados na estrutura de avaliação são aqueles que efetivamente são medidos em campo.

Na prática, a construção é bastante flexível e pode ser realizada por meio de diferentes procedimentos. O primeiro passo para caracterizar uma estrutura de avaliação é

definir seu escopo. Ela pode estar referida a contextos mais ou menos amplos, ajustando-se ao contexto de avaliação previamente definido. A escolha desse "nível de especificidade" traz importantes consequências operacionais, uma vez que o detalhamento dos atributos a serem medidos está relacionado com a abrangência pretendida para a estrutura. Exemplificando: uma estrutura de impactos pode ser construída para um caso específico, desdobrando-a em impactos detalhados exatamente para o caso. Uma estrutura de impactos assim construída teria grau máximo de aderência ao universo empírico avaliado, porém seria extremamente limitada do ponto de vista comparativo e de difícil generalização. Seria análoga a um estudo de caso no qual as principais variáveis foram formalmente arranjadas em uma estrutura aditiva e hierarquizada. A opção de adotar um escopo mais amplo para a construção da estrutura de impactos, por exemplo, o contexto setorial, torna mais razoável estabelecer comparações embora haja perda de aderência a casos específicos.

Isto significa que quando se define uma estrutura de impactos para um contexto mais amplo, por exemplo, para a agricultura no Estado de São Paulo, essa estrutura não mais se refere a alguma tecnologia em particular, devendo ser aplicável a qualquer tecnologia no âmbito da agricultura. Desta forma, ela se descola do universo empírico de cada caso e passa a ser definida a partir de uma interpretação mais abstrata dos impactos, interpretação esta ancorada no contexto "agricultura no Estado de São Paulo".

A Figura 4.9 mostra, a título de ilustração, parte da estrutura de impactos econômicos construída para a avaliação dos programas de pesquisa do IAC, já mencionados. A tela reproduzida do *software* compilado para o método MDM mostra a decomposição da estrutura até seus componentes mais básicos (Q), aos quais são associadas questões. É possível ainda observar os pesos definidos para o nó que reúne todos os componentes do impacto em *custos de transação*.

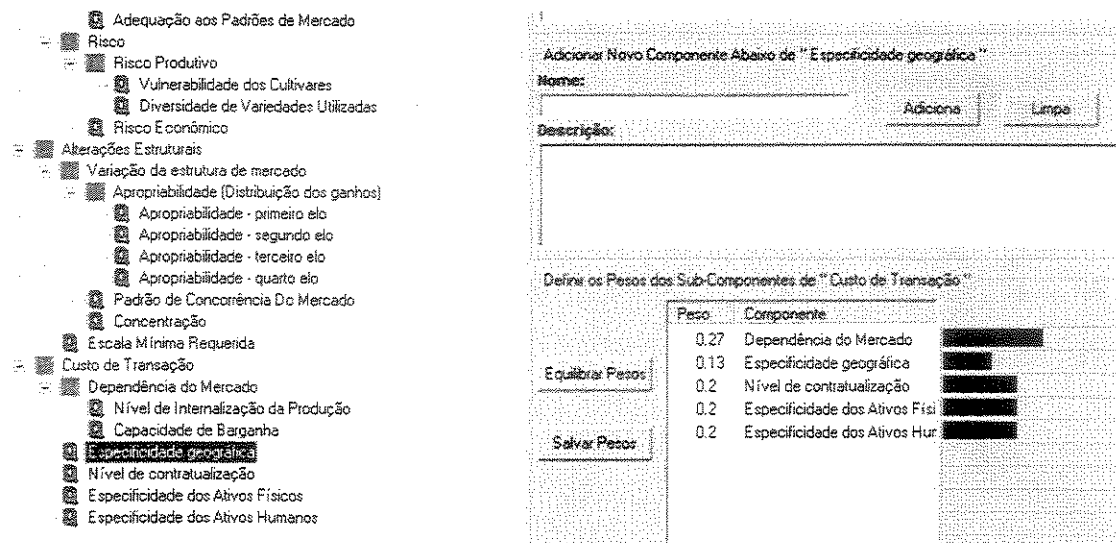


Figura 4.9: Parte da estrutura de avaliação para impactos econômicos de tecnologias na agricultura.

De todo modo, por conta das exigências impostas pelos axiomas de funcionalidade apresentados anteriormente e pela escala adotada para a medida de impacto, a definição prática dos componentes da estrutura de avaliação se torna uma tarefa não trivial. É preciso garantir:

- Adequação cognitiva que assegure que o conjunto de componentes proposto seja exato para descrever o "impacto" (Axiomas da Exaustividade e Não-Redundância);⁸⁴
- Nível de detalhamento compatível com o contexto de aplicação da estrutura de avaliação;⁸⁵
- Grau de objetividade suficiente na definição dos componentes básicos para que se obtenha medidas de variação (x) e participação (α) confiáveis;⁸⁶

O processo de construção é essencialmente interativo e deve contar com a participação de especialistas. A escolha e o encadeamento causal dos impactos para formar a hierarquia precisa ser tecnicamente justificada a partir de um entendimento comum que

⁸⁴ Evidentemente, obedecer esse requisito é inviável na prática, uma vez que sempre ocorre perda de significado ao se caminhar de noções genéricas para noções específicas. É preciso buscar um compromisso com os itens b e c seguintes.

⁸⁵ Observar que formalmente nada impede que exista uma estrutura de avaliação definida para um contexto mais geral que sofra detalhamento a partir daí para casos específicos.

⁸⁶ Isso nem sempre é viável dependendo da natureza do componente básico. Muitas vezes, um componente só pode recorrer à avaliação subjetiva e pode ser inescapável conviver com medidas ambíguas.

considere o estado atual dos conhecimentos sobre os efeitos cujas medidas possam caracterizar os impactos como positivos ou negativos, a pertinência desses efeitos para o universo avaliado⁸⁷ e outros aspectos operacionais para a modelagem, tais como as decisões necessárias para executar as medidas de campo.

Na construção da estrutura de avaliação os níveis intermediários fazem o papel de auxiliares conceituais que fornecem o encadeamento lógico para a agregação dos níveis inferiores em direção aos superiores, são os elementos conceituais que justificam a passagem do conjunto de atributos medidos em campo para medidas sintéticas nas dimensões escolhidas. É recomendável evitar o número excessivo de níveis e componentes na estrutura. É fácil verificar que a desagregação em 8 níveis com 8 desdobramentos por componente levaria a mais de 2 milhões de componentes por dimensão! Claramente isso seria de operacionalização impossível. Até 40 componentes é um bom limite prático.

Para cada ramo da estrutura (ou seja, para cada conjunto de elementos subordinado a apenas um impacto de nível superior na hierarquia) é definida uma fórmula de agregação. Em geral os impactos são aditivos, ou seja, suas medidas são somadas para compor a medida de impacto de nível superior a que se referem. Como nem sempre todos eles têm a mesma importância explicativa é necessário adotar ponderações específicas (ver exemplo da Figura 4.9). Assim, os pesos de cada impacto em seu ramo são ajustados de acordo com o entendimento conceitual de sua importância para a descrição do impacto diretamente acima. A arquitetura em camadas e aditiva da estrutura de avaliação permite que ela seja separada em módulos e que seu detalhamento, equações de agregação e respectivos parâmetros sejam definidos por grupos diferentes de especialistas.

Para atribuir ponderações a cada ramo da estrutura de avaliação que é agregada via equações aditivas, há pelo menos três opções:

1. *A ponderação por especialistas.* Um ou mais especialistas definem os pesos relativos a um determinado ramo. Cada ramo pode ser ponderado por indivíduos ou grupos independentes. Se a ponderação for feita por apenas um especialista, o procedimento é trivial. Se for por mais de um, pode-se usar procedimentos mais estruturados como o

⁸⁷ Pode ocorrer que certo efeito não seja tributável a uma determinada tecnologia mas monitorá-lo seja importante no contexto mais amplo no qual se insere a tecnologia (por exemplo, no contexto da agricultura, da indústria, da mineração etc.) para que possa haver comparação com outras tecnologias. Assim, a inclusão desse impacto na estrutura de impactos é justificável e necessária.

sugerido pelo método AHP⁸⁸ ou mesmo rodadas Delphi para se obter consenso. A contribuição de especialistas é mais indicada para os ramos inferiores da hierarquia, nos quais a especificidade é maior.

2. *A ponderação por atores.* Representantes de diferentes atores ponderam os ramos (é recomendado o uso de técnicas mediadoras como AHP ou Delphi). Pode-se obter opiniões agregadas, buscando valores de consenso, ou manter a diferenciação por grupos de atores, de modo a evidenciar eventuais conflitos na interpretação dos impactos. A ponderação por atores faz mais sentido nos níveis superiores da hierarquia, mais susceptíveis a diferentes posicionamentos políticos e ideológicos do ponto de vista da interpretação da importância conceitual dos diferentes componentes de impacto.

3. *Ponderação por intervalos de simulação.* Neste caso, definem-se combinações de pesos que polarizem interpretações estereotipadas e anota-se o comportamento dos resultados. Em seguida, força-se a variação dos pesos e estuda-se a mudança que isso acarreta nos impactos. Este procedimento é útil para analisar a sensibilidade do modelo e para aprofundar a análise dos impactos.

Uma consequência importante de se operar conscientemente com a possibilidade de convivência de diferentes racionalidades é que a interpretação e, portanto, a medida dos impactos pode ser considerada não única. Ela também seria dependente dos contextos tecnológicos, históricos e sociais. Se determinados fatos novos surgirem (inclusive por consequência da nova tecnologia avaliada), a estrutura de avaliação precisará ser atualizada e isso modificará o resultado por ela fornecido, inclusive para uma avaliação feita anteriormente (em outro contexto de interpretação). Assim, a medida de impactos não pode ser tomada como uma atividade puramente objetiva, trata-se de uma interpretação socialmente construída de um fenômeno também social. A partir desse ponto de vista evolutivo, a total consistência conceitual da estrutura não é fundamental. Pode-se dizer que uma estrutura de avaliação age como uma mediadora de compromissos, não necessariamente livre de contradições, porém suficiente para os propósitos de avaliação.

⁸⁸ Analytical Hierarchy Process. Cf. SAATY, 1997.

Medidas de Campo

Uma vez definidas as estruturas de impacto, seus componentes básicos correspondem aos aspectos que serão objeto de medidas de campo. Para cada componente básico é preciso definir uma variável x , sua escala de medida e a forma pela qual será medida. Nesse momento é preciso decidir sobre:

a) A definição da direção (positiva ou negativa) do impacto segundo as escalas adotadas para medir as variáveis de cada componente básico;⁸⁹

b) A definição da relação de correspondência entre cada variação de x e a medida para cada componente básico de impacto que deverá ser descrita em cada função ϕ ;⁹⁰

A definição desse último item é dependente da escolha entre os Modos Direto e Indireto de obtenção das medidas de impactos. Os dois Modos permitem a adoção de procedimentos diferentes para a definição de quais valores de variação corresponderiam aos máximos impactos positivos e negativos.⁹¹

O caminho do Modo Indireto assume um afastamento "técnico" da condição vivenciada por atores que inseridos no contexto de aplicação da nova tecnologia estão efetivamente convivendo com seus impactos. O Modo Indireto não prescinde de subjetividade, apenas a desloca para outro nível, identificado com os especialistas, muitas vezes externos ao contexto de avaliação. Por outro lado, é preciso controlar rigidamente a qualidade da amostragem para escapar de possíveis distorções que podem causar a "racionalidade limitada" de certos indivíduos no Modo Direto.

A saída pragmática de usar parte dos indicadores a partir de um Modo e parte de outro, conforme a facilidade de obtenção, pode também criar distorções pela dificuldade de calibrar homogeneamente a função $\phi: x \rightarrow [-1, 1]$ para as diferentes situações. Sendo os atores do Modo Direto e os especialistas do Modo Indireto integrantes do mesmo contexto de avaliação, essa dificuldade pode ser minimizada.

⁸⁹ Se existir controvérsia nesse ponto, provavelmente o indicador se relaciona a mais de um impacto e precisará ser considerado mais vezes na estrutura de impactos.

⁹⁰ Em geral, a partir dos valores extremos de variação de x faz-se I constante e igual a 1 ou -1 (a partir dali o impacto é sempre máximo).

⁹¹ É importante destacar que as definições dos extremos de impacto precisam guardar homogeneidade entre os indicadores para que a soma dos impactos faça sentido. A definição dos extremos de impacto pode ser, portanto, também um fator de ponderação (porém menos explícito que o peso k). A comparabilidade entre dois casos depende de se usar as mesmas definições de extremos de impacto, assim como é preciso usar o mesmo conjunto de pesos.

O uso de medidas de impacto oriundas de escalas do Modo Direto ou do Modo Indireto não exclui a coleta de sua complementar, ou seja, de medidas quantitativas para o Modo Direto e qualitativas para o Modo Indireto.

XIII. DIRETRIZES PARA A COLETA DE DADOS

Esse item apresenta sugestões de procedimentos gerais para a composição de um questionário para ser utilizado em trabalhos de coleta de dados no campo.⁹² Para cada dimensão, um questionário de avaliação de impactos pode ser composto de:

- a) Identificação da Avaliação (nome do projeto; tecnologia; programa; dimensão);
- b) Identificação do Respondente (nome; cargo; empresa; ator; categorias);
- c) Identificação da Entrevista (entrevistador; data; local; horário de início; horário de término);
- d) Campos de preenchimento obrigatórios e complementares para cada componente básico da estrutura de avaliação.

Nos questionários, a cada componente básico corresponde um conjunto de campos de preenchimento para contemplar:

- 1) Um juízo qualitativo (medida sensorial – Modo 1) sobre a variação observada no aspecto a que se refere no tempo transcorrido desde a adoção da nova tecnologia (x);

Para medir x , sugere-se recorrer a uma escala de Likert de sete pontos com duas opções de exceção, por exemplo:

- () grande aumento
- () moderado aumento
- () pequeno aumento
- () não houve alteração
- () pequena diminuição
- () moderada diminuição
- () grande diminuição
- () não se aplica
- () não sei

⁹² O *software* para o método MDM permita a aplicação via Internet de questionários, contemplando os itens apresentados nesta seção.

Se for o caso de se utilizar uma escala quantitativa (Modo Indireto), é mais prático usar uma escala intervalar que já explicita os juízos de especialistas – previamente estabelecidos – sobre a interpretação dos extremos da escala (Modo Indireto). Por exemplo uma escala do tipo: {10 ou mais, 5 a 10, até 5, 0, até -5, -5 a -10, -10 ou mais, não se aplica} etc.

2) Um juízo qualitativo (medida sensorial) sobre a participação (α) da nova tecnologia nas mudanças observadas no período;

Para determinar α , pode-se recorrer a uma escala de Likert do tipo:

“A participação da tecnologia na variação observada acima é:

- ☐ nenhuma
- ☐ quase nenhuma
- ☐ média
- ☐ quase total
- ☐ total
- ☐ não sei”

Além dessas duas questões com escalas qualitativas, o questionário deve conter de modo complementar:

- 3) Comentários e explicações sobre a resposta ao item 1.;
- 4) Comentários e explicações sobre a resposta ao item 2.;
- 5) Dados quantitativos que justifiquem 1. e/ou 2 (ou dados qualitativos sobre a importância da medida quantitativa se 1. for obtida pelo Modo Indireto);
- 6) outras observações.

A formulação das questões deve ser cercada de cuidados especiais no sentido de explicitar claramente ao entrevistado o aspecto que está sendo medido. É importante adotar uma linguagem familiar ao caso e ao ator investigado, porém sem distorcer o significado original do atributo. Deve-se usar terminologia e estruturas de frase buscando o máximo de neutralidade em relação a qualquer suposta "melhor resposta". As questões precisam ser formuladas para evitar ao máximo esse tipo de distorção e procurar amarrar a medida sensorial diretamente ao efeito analisado. O formato em duas questões qualitativas – sendo a primeira geral e a segunda restringindo a medida da primeira – seguidas por pedidos de explicação e dados quantitativos ajuda nesse sentido, porém a qualidade da construção do

período, a qualidade do entrevistador e a qualidade do respondente são fatores que também pesam na confiabilidade das medidas de campo.

A quantidade de entrevistas necessárias é função da qualidade desses fatores e também da quantidade de atores escolhidos. Se esses atores forem subdivididos segundo diferentes categorias, a amostra automaticamente cresce. O tipo de informação coletada para a avaliação de impactos não se traduz em eventos aleatórios (tampouco os erros induzidos por defeitos na confecção do questionário). A estratégia a ser adotada é sempre a de tomar uma amostra intencional que deve ser construída levando-se em conta fatores como a qualidade e a representatividade (política, econômica ou em termos de produção, área ocupada etc.) dos indivíduos escolhidos e os custos envolvidos. Como regra geral, é recomendável que a amostra contenha aproximadamente o mesmo número de representantes para cada tipo e categoria de ator.

As Figuras 4.10 e 4.11 abaixo mostram, respectivamente, as interfaces do *software* compilado para a composição de questionários e para a definição de atores participantes da avaliação.

Componente: Especificidade geográfica **Salvar Questão**

As variedades utilizadas alteraram a flexibilidade com respeito a mobilidade geográfica da produção?

Parâmetros do Instrumento de Medida

Grande aumento	-1
Médio aumento	
Pequeno aumento	
Sem alteração	0
Pequena diminuição	
Média diminuição	
Grande diminuição	1

Escala

☐ Simples

☒ Dupla

Pontos

☐ 2

☐ 3

☒ 4

☐ 5

☐ 6

Não se aplica 0

Prefiro não responder 0

Sentido

☐ Direto

☒ Inverso

Sensibilidade

☒ Normal

☐ Alta

☐ Baixa

☒ Pesquisar atribuição

Atribuição

Qual a participação da tecnologia gerada pelo PROCANA na alteração observada de flexibilidade em relação à mobilidade geográfica?

Opções de resposta

Nenhuma

Quase nenhuma

Média

Quase Total

Total

☒ Observações e comentários **Tipo**

☒ Texto

Figura 4.10: Interface para edição de questões para os componentes básicos da estrutura de avaliação do método MDM.

The interface is titled "Atores" and is divided into three main sections:

- Grupo de Atores:** A list box containing "Atores", "Fornecedor de Insu...", "Indústria", "Pesquisador", and "Agricultor". "Agricultor" is currently selected.
- Form Fields:**
 - Nome:** A text field containing "Agricultor".
 - Descrição:** A large text area for description.
 - Buttons:** "Adiciona", "Altera", and "Apaga" are located at the bottom of this section.
- Categorias de Agricultor:** A table with two columns: "Categorias" and "Nome".

Categorias	Nome
Pequeno	
Médio	Desc
Grande	

 A button labeled "Adi" is at the bottom right of this section.

Figura 4.11: Interface para definição dos atores participantes e suas categorias.

XIV. COMENTÁRIOS FINAIS SOBRE A AVALIAÇÃO EM MÚLTIPLAS DIMENSÕES

A avaliação em múltiplas dimensões modelada para a avaliação de impactos fornece três tipos de medidas de impactos:

- a) *impacto da tecnologia;*
- b) *impacto geral;*
- c) *impactos de outras causas* não decorrentes da tecnologia avaliada.

Essas medidas são disponibilizadas para todos os componentes da estrutura de impactos. Para o melhor entendimento dos impactos é adequado explorar os resultados seguindo a cadeia causal que define camada por camada da hierarquia de impactos e complementar a explicação de cada medida com informações colhidas de modo não estruturado durante a pesquisa de campo. Os diferentes níveis de agregação equivalem a diferentes níveis de detalhamento para a avaliação (ver Figura 4.12).

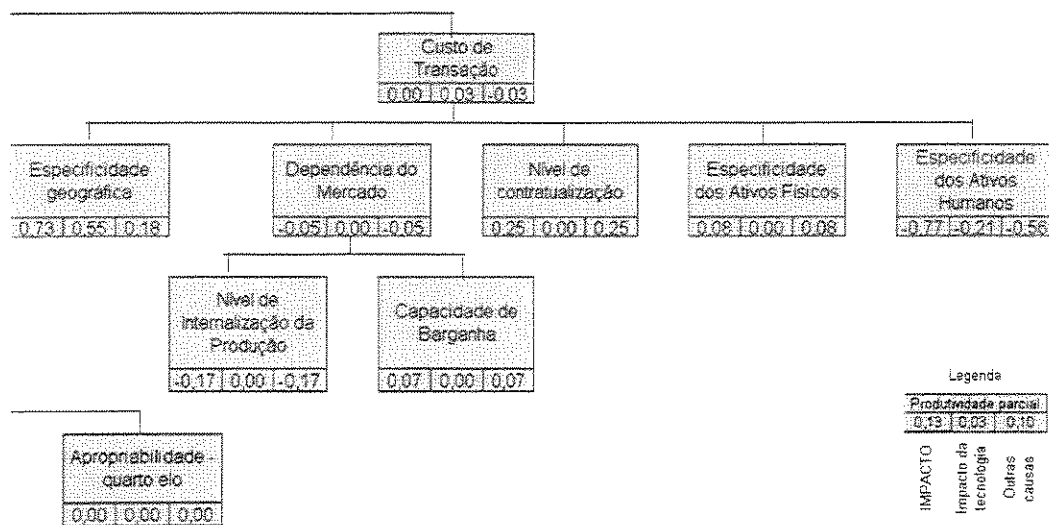


Figura 4.12: Resultados de campo para a avaliação dos impactos econômicos do PROCANA.

Num esquema de equações aditivas, a sensibilidade das medidas de impactos aos parâmetros de ponderação definidos (pesos) depende da existência ou não de impactos opostos em um mesmo ramo. Quando isso ocorre, a tendência é de compensação mútua. Por outro lado, quando os impactos em diferentes componentes são muito parecidos, a variação nos pesos perde importância na determinação do impacto subsequente na hierarquia.

A coexistência de dimensões é preservada ao não se fazer uma agregação final de seus impactos. Apresentar os valores para a medida de impacto das diferentes dimensões é normalmente o bastante para que quem observe os resultados aplique seu próprio esquema de ponderações para obter um juízo final sintético. Ademais, a importância dos resultados em cada dimensão está ligada diretamente à importância que cada dimensão teve na definição inicial de objetivos do programa tecnológico e às prioridades da agenda do avaliador.

A coexistência de ambigüidades na identificação de impactos é tolerada pelo método. Quando ocorre ambigüidade, os dispositivos de escolha do valor representativo o fazem de modo cauteloso e retiram importância do componente para a medida do impacto na dimensão. Na prática, sugere-se a adoção de $Z = 0,75$ ou $Z_{\alpha} = 0,5$ para o limite de

tolerância à ambigüidade. Para valores menores que isso, recomenda-se desdobrar os resultados, primeiro por atores e depois por categorias de atores, de modo a buscar por conjuntos de medidas que satisfaçam aquelas condições.

Isso significa que, para tais casos, a melhor medida para o impacto deve ser avaliada pelo recorte apropriado e não pelo contexto da avaliação original, que ao possuir $Z < 0,75$ obscurece muitas medidas ambíguas. Provavelmente, os impactos acontecem de forma diferenciada para diferentes categorias de atores sociais e só em partições mais coesas poderão ser adequadamente caracterizados.

Por exemplo, os resultados empíricos da aplicação do método MDM apresentados na Figura 4.12 mostraram coesão relativamente baixa, $Z = 0,56$. Por esse motivo, a análise dos resultados precisou ser feita a partir dos dados desagregados por diferentes categorias do ator Produtor. A Figura 4.13 mostra os resultados para os produtores da região de Ribeirão Preto, entre eles houve coesão maior que 0,75. O mesmo ocorreu entre os produtores das outras regiões consideradas, Piracicaba e Jaú. Cada um desses grupos apresentou alta coesão entre si, indicando que os impactos se manifestaram de modo diferente segunda cada região.

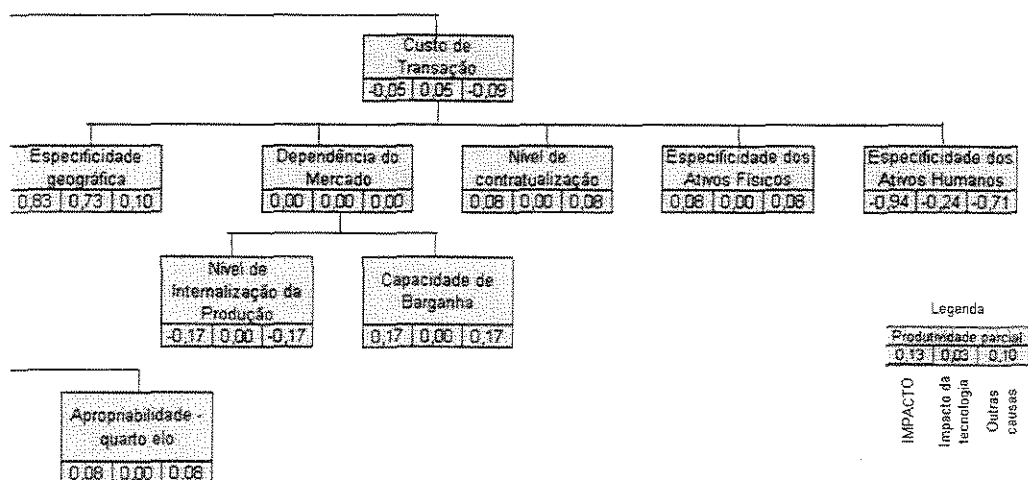


Figura 4.12: Resultados de campo para a avaliação dos impactos econômicos do PROCANA, desagregado para os produtores da região de Ribeirão Preto.

Finalmente, a medida de impacto é uma medida dinâmica. Ela não se dissocia do movimento geral das transformações sociais e das trajetórias tecnológicas em curso. O fato de se obter uma medida de impacto nula não significa que nada está mudando no contexto de aplicação de uma determinada tecnologia, mas sim que o impacto de um determinado programa/tecnologia nesse processo é nulo.

Além disso, uma medida obtida para uma avaliação entre t_0 e t_1 , com estruturas de impacto construídas para determinado contexto de avaliação, com a participação de um determinado grupo de especialistas, com a entrevista a uma determinada amostra de atores e com a interpretação de um determinado grupo de avaliadores é um retrato contingente a todos esses fatores e constituído do nível de conhecimento disponível em tal universo. Seguramente, a execução da avaliação de impactos desencadeia um processo de aprendizado que naturalmente leva à necessidade de se modificar os ingredientes inicialmente previstos, buscando o aperfeiçoamento do conjunto de premissas da avaliação. Esse processo é interminável e, em geral, o próprio objeto de avaliação também aprende e se modifica. Um certo grau de pragmatismo é sempre salutar no sentido de manter claro que o principal objetivo da avaliação é melhorar o entendimento dos impactos e prover meios para mitigar impactos negativos e para potencializar impactos positivos. Portanto, a utilização do método deve ser pautada em fornecer insumos de qualidade para a tomada de decisões e o aprendizado.

Conclusões

O método de Avaliação em Múltiplas Dimensões (MDM), apresentado no Capítulo 4, foi desenhado com o propósito de formalizar alguns procedimentos para avaliar a C&T em Rede. Essa forma institucional da ciência e tecnologia é caracterizada pela co-existência de múltiplos arranjos entre múltiplos atores da sociedade, orientados à produção de conhecimento e inovação. O Capítulo 3 estabeleceu as distinções entre a C&T em Rede e as formas que a antecederam, a Grande e a Pequena C&T, e discutiu as lógicas subjacentes aos métodos de avaliação empregados em cada uma delas.

Para C&T em Rede, premissas como a participação e o aprendizado associados aos processos de avaliação e a promoção de coordenação visando inovação são distintivas. Essas características precisam, de algum modo, ser incorporadas em qualquer proposta de método de avaliação que se pretenda adequada a essa nova forma institucional.

Entretanto, como mostrado nos Capítulos 1 e 2, participação e aprendizado são de difícil incorporação a métodos de avaliação. Para que isso seja possível é preciso entender quais são os limites da noção de método, os limites estruturais do esforço objetivador e da reprodutibilidade, e necessariamente flexibilizá-los. Se se exigir uma racionalidade objetivadora para um método, ele não poderá ser participativo, e o aprendizado será apenas efeito secundário, externo ao método empregado.

De partida, é oportuno resgatar a justificativa do presente esforço de sistematização teórica e metodológica, bem como do formato escolhido para apresentar seus elementos.

Minha experiência em diversos casos práticos de avaliação e, em particular, no desenvolvimento metodológico para avaliação de impactos da inovação, permitiu ao longo dessas experiências a constatação de dois fatos importantes. O primeiro deles é a relativa escassez de produção teórica sobre avaliação. Encontram-se inúmeros artigos com descrições de casos, relatórios técnicos e manuais, porém pouca reflexão teórica sobre avaliação. Essa, quando existe, surge no geral como crítica a métodos ou a aplicações particulares, sobre as quais é difícil generalizar. Um segundo ponto a ser destacado é a crescente insatisfação de avaliadores e avaliados com as ferramentas disponíveis para avaliar programas sociais ou outros objetos que envolvam conflitos de valor, aprendizado e

subjetividade nos julgamentos acerca dos fatos, como também é o caso dos impactos provocados pelas inovações.

É claro que existe correlação entre essas duas constatações. Quando as coisas parecem desajustadas no nível da aplicação, sente-se automaticamente a necessidade de investigar os conceitos, as premissas e os limites das teorias que sustentam as ferramentas usadas no dia a dia. Este trabalho surgiu exatamente da visão de que era urgente fazer essa investigação, para entender o porquê da sensação de desajuste, melhor delimitar a extensão desse desajuste e, eventualmente, propor outras abordagens para as práticas de avaliação.

Assim, a divisão em duas partes acomoda as principais vertentes da pesquisa. De um lado, os esforços de entendimento básico, abstrato, com intenção explícita de encontrar os elementos gerais para analisar as avaliações e para precisar os limites de seus métodos. Do outro, os esforços de localização, de contextualização dos métodos e das motivações da avaliação no universo da ciência e tecnologia. A descrição final de um método para avaliação dos impactos se insere na encruzilhada entre as duas vertentes; o método envolve a adoção de ferramentas que encontram fundamentação teórica nas discussões apresentadas na primeira parte e fundamentação de contexto baseada no que se desenvolve na segunda parte, incluindo a adequação ao contexto específico do para o qual o método foi concebido e ao panorama mais amplo do estágio atual do desenvolvimento organizacional e institucional da ciência e tecnologia. Esses aspectos de fundamentação teórica e de contexto são as bases para os principais tópicos que orientam as conclusões alcançadas por este trabalho. Para cada um desses tópicos, há desafios e trajetórias que se tornaram mais claros.

Estas notas finais retomam alguns elementos da argumentação apresentada no decorrer deste trabalho para esquematizar, de modo sucinto, o conjunto de desafios que hoje se colocam às atividades de avaliação em ciência, tecnologia e inovação. Em paralelo, os caminhos percorridos pelo método MDM para encontrar respostas a esses desafios são também discutidos.

As conclusões estão organizadas em seis itens. Primeiro são recuperadas algumas reflexões acerca da possibilidade de participação. Em seguida três pontos relativos à formalização do método de avaliação são revisitados, o uso de formas não exatas, sua estrutura em múltiplas camadas e a possibilidade de acomodar outros métodos em seus procedimentos. Um quinto item discute a promoção de coordenação e aprendizado e as

implicações para a Política Científica e Tecnológica. Finalmente, oferece-se um esboço de uma agenda de pesquisa destinada a dar continuidade aos esforços deste trabalho.

Sobre a participação

O problema da participação é talvez um dos mais difíceis e contraditórios para as avaliações. Afinal, entram em cena várias perspectivas sobre o que se entende por participação e sobre como incorporar de fato a participação do ponto de vista metodológico.

No plano geral da sociedade, participação significa tornar as avaliações mais democráticas, considerar a crítica e a oposição, inclusive das minorias, evitando assim decisões tecnocráticas baseadas apenas em critérios de otimização de alguns indicadores.

No plano conceitual, pode ser interpretada como a convivência de múltiplas perspectivas sobre um mesmo objeto. Essas perspectivas podem vir de diferentes atores ou de diferentes abordagens teóricas. Nesse sentido, um método capaz de lidar com múltiplas dimensões seria adequado tanto para congregar diferentes perspectivas teóricas quanto diferentes posicionamentos e preferências dos atores.

Entretanto, metodologicamente esse programa encontra limites importantes. Primeiro, como mostrou Arrow, nenhuma agregação de preferências é capaz de preservar a racionalidade objetiva. Segundo, se, além disso, as interpretações sobre os atributos forem divergentes, ou melhor, se os atributos a e b não forem exatamente os mesmos para as diferentes visões como acontece, por exemplo, no caso das interpretações de diferentes linhas teóricas, o problema da agregação fica ainda mais difícil.

O método MDM adota as seguintes atitudes em relação a esses problemas. Primeiro, em relação aos atores, o método entende que a participação somente é viável se a avaliação for entendida como coletiva, como método socialmente legítimo, mas que se limita a influenciar as decisões dos indivíduos. Assim, em diversos momentos a participação é necessária para a avaliação: na definição do contexto de avaliação, na definição dos objetivos e objeto(s), na construção das estruturas de avaliação, nas respostas coletadas e na interpretação dos resultados obtidos. O procedimento de agregação para a escolha de um valor representativo, por meio dos graus de coesão, ajuda a evidenciar as eventuais dissidências e se elas têm ou não ligação com determinadas características dos atores.

Porém é evidente que se perde objetividade nesse salto entre, por um lado, as racionalidades dos indivíduos e, por outro, a racionalidade coesa que fornece uma medida agregada para se calcular os critérios. Não poderia ser diferente.

O tratamento que o método MDM dá para as diferentes dimensões é mais simples: as dimensões são separadas e mantidas independentes. Este é um ponto para o qual caberia ainda muito mais pesquisa. Por exemplo, a) estudar aplicações de lógica *fuzzy* para estabelecer pontes entre as dimensões, talvez até mesmo entre seus atributos já decompostos; b) assumir que sobre os atributos de cada dimensão, outras dimensões também algo a dizer; c) tratar esses atributos como conceitos deliberadamente mais imprecisos, para acomodar as diferentes perspectivas e depois aplicar algum princípio similar aos graus de coesão para obter resultados delimitados; d) o procedimento de obtenção de um valor representativo das respostas dos atores a partir de seus graus de coesão é um tipo de "defuzificação", que também pode ser aperfeiçoado.

Por hora, o que se conseguiu com o método MDM, em suas aplicações práticas, mostrou que mesmo suas opções metodológicas, bastante simples, levam a resultados interessantes e, muitas vezes, surpreendentes.

O uso de métodos não exatos

Alguns dos recursos metodológicos introduzidos no MDM podem ser caracterizados como não exatos. Ser não exato significa que tais recursos não são fundamentados numa lógica estrita de objetivação. Isso é consequência inevitável das premissas de múltiplas dimensões e de participação.

Convenientemente, há formas matemáticas que ajudam nesse sentido. No método MDM, tanto a escolha dos graus de coesão, como a escolha do algoritmo para se obter os valores representativos são subjetivas. E também o são a medida dos atributos, a medida das atribuições, a construção das estruturas de avaliação, a definições dos pesos e as equações de agregação aditivas na estrutura de avaliação. Tudo isso é externo à matemática que opera as medidas através das estruturas de avaliação, pois faz parte de sua contextualização. Em resumo, são escolhas e parâmetros que precisam ser definidos por

alguém. Eles não são calculados a partir dos fatos, eles são construções sociais que precisam ser revistas caso a caso.

Desse modo, apenas os reticulados da estrutura de avaliação são matemática convencional. E precisam ser, porque de outro modo não haveria medidas. Mas, como tudo que neles entra, tanto em termos conceituais como numéricos, é fruto de processos de construção social, tudo que deles emerge também precisa ser lido com cautela.

Os procedimentos adotados possibilitam que se lide intersubjetivamente com a incerteza e a idiosincrasia dos diferentes aportes subjetivos que são necessários para a avaliação. Esse ponto é particularmente importante. O subjetivo é parte inescapável de qualquer avaliação. A fachada de objetividade que alguns métodos ostentam nunca exclui as escolhas anteriores e externas a eles. No método MDM, essas escolhas precisam ser explicitadas e discutidas antes que se tenha um modelo para realizar a avaliação.

Se não há como evitar o subjetivo, unamo-nos a ele, e aproveitemos para reproduzir o espírito da C&T em Rede, envolvendo especialistas com diferentes abordagens, gestores de inovação, técnicos, trabalhadores etc. A avaliação não se resume a coletar dados e analisar os resultados, ela se torna também um processo de aprendizagem que evolui para uma melhor apreensão de uma miríade de aspectos heterogêneos que não poderiam ser apreendidos de outro modo.

A arquitetura multicamadas

O método MDM é um procedimento que pressupõe a sedimentação de múltiplas camadas de entendimento sobre a realidade. Não só pelo fato de acomodar múltiplas dimensões do objeto avaliado, mas por se capaz de acomodar múltiplas avaliações (e os resultados intermediários de outros métodos de avaliação) e a participação de diferentes atores em diferentes momentos e com diferentes níveis de relevância.

Esse método pode ser interpretado como uma estrutura organizada para discussão em grupo, que pode levar a uma ou a várias conclusões. O indivíduo é sempre o decisor e, nesse sentido, o método não tem como garantir uma decisão.

A presença de múltiplas dimensões não implica que um decisor leve todas elas em consideração para sua decisão. Do mesmo modo, dentro de uma mesma dimensão,

determinados componentes (ou um determinado nível de desagregação) podem simplesmente ser irrelevantes para um decisor. O método fornece o quadro geral, e o resumo das avaliações em múltiplas instâncias, o que permite que se estude em detalhe esse quadro geral, inclusive em termos de suas controvérsias. Mas é óbvio que não pode substituir o embate político sobre as decisões de interesse da sociedade.

A inclusão de outros métodos

De um ponto de vista pragmático, mais especificamente focando nas organizações voltadas a ciência, tecnologia e inovação, a avaliação tem papel a cumprir em pelo menos três diferentes níveis, que revelam diferentes necessidades e características:

1. Quanto à adequação dos aspectos operacionais internos;
2. Quanto à consecução de objetivos e impactos dos resultados;
3. Quanto aos arranjos cooperativos e a capacidade de governança entre atores.

Para cada um desses níveis, mudam os atributos a serem pesquisados. Entretanto, o esquema geral do método MDM pode ser aplicado a todos eles. Assim, as prerrogativas de participação, de múltiplas dimensões de análise e de fomento da inovação continuam valendo para qualquer desses níveis.

Detalhar os atributos que devem ser empregados para aspectos organizacionais, adequação entre resultados e objetivos, impactos ou capacidade de governança é uma tarefa que precisa ser realizada em função do objeto avaliado. O Capítulo 4 enfatizou o caso da avaliação de impactos para tecnologias introduzidas na agricultura paulista.

Quando o detalhamento dos atributos levar a situações que podem ser tratadas por métodos tradicionais, o resultado obtido por meio deles pode ser perfeitamente inserido numa estrutura de avaliação, na qual será relacionado com outros resultados que não podem ser objetivamente obtidos.

Em geral, as avaliações dos aspectos operacionais e dos resultados podem ser mais facilmente objetivadas. Alguns tipos de impactos também podem ser aferidos por meio de métodos objetivados. Entretanto, como foi amplamente discutido nos Capítulos 3 e 4, muitos dos impactos e dos efeitos de coordenação e governança não o podem, necessitando de abordagens diferenciadas.

As técnicas existentes para *assessment* e *accountability* em geral falham quando se trata de levar em conta percepções ambíguas de diferentes atores impactados pelos resultados e outras incertezas inerentes dos processos de inovação. Nesses casos, a avaliação envolve necessariamente aspectos tácitos e subjetivos.

Isso tudo reforça a oportunidade para abordagens metodológicas capazes de transitar pelos diferentes níveis e integrar os procedimentos de avaliação típicos de cada um deles.

Aprendizado e Coordenação

As necessidades de cooperação e capacidade de governança entre atores heterogêneos que caracterizam a C&T em Rede implicam em novos atributos para as avaliações. Esses atributos são alvo de diversas metodologias de avaliação criadas para esse contexto institucional, conforme mostrou a seção V do Capítulo 3.

Porém, apenas medir esses atributos é insuficiente, uma vez que há rapidez nos processos de mudança e muita incerteza associada a essas mudanças. Mais do que diagnosticar as estruturas de cooperação e de governança é preciso criá-las e mantê-las azeitadas.

Isso vale para as organizações que participam da C&T em Rede e vale também para a Política Científica e Tecnológica que pretenda atuar sobre essa forma institucional.

É fato que estamos hoje numa transição (ou como diria Etzkowitz e Leydesdorff, numa *endless transition*). No nível mais agregado do planejamento não cabem mais os instrumentos tradicionais de planejamento. Para lidar com complexidade dos arranjos da C&T em Rede cabe, no máximo, esforços de *coordenação*.

E talvez a maior insuficiência – ou visto por outro lado, o maior desafio a ser superado – mesmo das abordagens que pretendem gerar o ambiente propício para essa coordenação (como é o caso da maior parte dos exercícios de *foresight*) é que elas ainda empregam para isso apenas procedimentos similares aos do planejamento convencional. Ou seja, apesar de reconhecerem a necessidade de envolvimento e de aprendizado dos atores, isso ainda se dá praticamente apenas no plano conceitual. Na prática, os métodos empregados ainda são os mesmos método objetivados que trazem consigo (explícita ou tacitamente) a velha concepção. Entremeados a esses métodos, procedimentos que buscam

convergência e coordenação são adotados de modo muito pouco sistemático e têm dificuldade de alcançar na prática os objetivos pretendidos em teoria.⁹³

Novamente, a tentativa de formalização do método MDM incorpora explicitamente essa preocupação. Um método de avaliação no contexto de C&T em Rede deve servir, prioritariamente, para promover aprendizado e gerar auto organização. Sendo assim, o processo de avaliação é também uma oportunidade de gerar coordenação, é uma intervenção política. Não mais no sentido normativo, mas no sentido operacional. A existência da avaliação é uma oportunidade para canalizar as demandas e as percepções de distintos atores, o que pode implicar na criação de agendas oficiais e extras oficiais tanto de ação e como de reação. E é essa precisamente a dinâmica que precisa ser fomentada para incentivar a inovação. As inovações nunca são minuciosamente planejadas e não podem ser, elas nascem antes de um contexto favorável e da fertilização mútua de competências.

Especulações teóricas e agenda de pesquisa

A investigação que gerou este trabalho se caracterizou por investidas em diversas disciplinas e correntes teóricas que de alguma forma se relacionam com avaliação e suas conexões com ciência, tecnologia e sociedade.

Como não poderia deixar de ser, as possibilidades não foram esgotadas e diversas frentes de pesquisa continuam abertas. Muitas dificuldades ainda subsistem e, com certeza, novas surgirão a partir do uso continuado do método de avaliação construído, especialmente após a difusão do recém concluído *software* para o método.

A possibilidade de construir *softwares* para operacionalizar os procedimentos metodológicos traz algumas vantagens. É possível usar algoritmos relativamente complicados e que envolvam grande capacidade de processamento (como, por exemplo, ferramentas de mineração de dados e de textos que podem ser empregadas para fornecer

⁹³ Autores como RAPPERT (1999) e DAVID (2002) apresentam críticas interessantes aos exercícios de *foresight*. O primeiro argumenta que as abordagens de *foresight* representariam, na verdade, um esforço de racionalização do futuro da evolução do conhecimento e que isso ao invés de promover a inovação a estancaria em médio prazo. Há nesse, argumento, um componente de *ethos* de Pequena C&T reclamando sua liberdade de pesquisa e criação, mas há também a percepção de que os resultados práticos do *foresight* não corresponderiam ao seu discurso conceitual. Paul David, por sua, vez enxerga nos exercícios de *foresight* apenas um "modelo linear invertido", no qual a Política Científica estaria subordinada às demandas do mercado. Ambos discursos são fundamentados nas fraquezas metodológicas dos exercícios realizados na prática.

medidas e indicadores a partir de respostas discursivas ou procurar por padrões intrincados no conjunto das respostas de todos os atributos avaliados). É possível também realizar diversos estudos por simulação de parâmetros como peso dos atributos, interdependência, busca por grupos coesos, entre outros. Essas linhas de desenvolvimento certamente colaborariam com o aperfeiçoamento do método.

Do ponto de vista mais teórico, chama a atenção o mar de possibilidades que as aplicações de modelos baseados em lógica *fuzzy* têm a explorar nas ciências sociais. Por exemplo, a partir da modelagem das preferências como estruturas de comparabilidade parciais (Capítulo 2), por que não repensar as curvas de indiferença dos modelos de comportamento microeconômicos da economia neoclássica? Certamente, algumas dificuldades dessa teoria seriam evitadas.

Mais especificamente no que se refere à avaliação, a aplicação de estruturas *fuzzy* ao método pode avançar muito mais. A determinação dos pesos poderia ser estudada a partir dos conceitos dos conjuntos aproximativos (*rough sets*) e os estudos das distribuições das respostas e da coesão a partir dos conceitos de integral *fuzzy* ou de medidas de possibilidade. Cada uma dessas frentes significaria importantes avanços em relação aos procedimentos um tanto quanto simplificados até aqui adotados.

A propósito da condução dos processos de avaliação, é preciso avançar nas técnicas de construção de questões e questionários, procurando adequar esses instrumentos às peculiaridades psicológicas e sociais dos indivíduos inseridos no ambiente institucional da C&T em Rede. Porém, especialmente, é preciso adequar esses instrumentos também aos indivíduos ainda não plenamente inseridos nesse ambiente. Uma das dificuldades importantes é superar a barreira e a desconfiança com avaliações, herdada de experiências nem sempre agradáveis com avaliações sob outros contextos. É preciso estudar as causas da recorrentemente baixa motivação em participar e criar estratégias para reverter esse quadro e promover a inserção das pessoas à nova lógica de avaliação que se propõe. Um caminho que pode ser interessante nessa direção seria estudar a aplicação do modelo do jogo agentes-principal para arranjos típicos da C&T em Rede e para as situações de transição da Pequena ou Grande C&T para a C&T em Rede. Claramente, teríamos uma situação de múltiplos jogadores com agendas contraditórias e altamente mutáveis. A situação de instabilidade e incerteza precisaria ser mais bem modelada (novamente com o uso de

modelos *fuzzy* ou de aprendizado, como redes bayesianas ou mesmo redes neurais) para tentar a formulação de procedimentos robustos de participação, capazes de favorecer os processos coletivos de avaliação e de promoção de efeitos de coordenação, inclusive como forma de reduzir a incerteza e a instabilidade a níveis toleráveis.

Há outras conexões, portanto, que emergem dessa agenda. As questões de contratualização entre atores e de apropriação de resultados. Os gargalos de financiamento e das estratégias setoriais. Enfim, diversos temas focais da Política Científica e Tecnológica estão intimamente relacionados com os processos de avaliação num contexto de C&T em Rede.

É minha opinião que cada um desses temas não deva ser estudado isoladamente. Não podemos nos dar o luxo de perder o essencial, a oportunidade de produzir estudos e intervenções em múltiplas dimensões, de modo a mobilizar pessoas e mobilizar organizações para estratégias mais efetivas e socialmente interessantes de inovação.

Referências Bibliográficas

- ABRÃO, B. S. *História da Filosofia*. Editora Nova Cultural Ltda., São Paulo, 1999.
- ALSTON, J. M., NORTON, G. W., PARDEY, P. G. *Science Under Scarcity: Principles and Practice for Agricultural Research Evaluation and Priority Setting*. Cornell University Press. Ithaca & London, 1995.
- ARMSTRONG, J. S. (ed.). *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Norwell, MA: Kluwer Academic Publishers, 2001.
- ARROW, K. *Social Choice and Individual Values*. 2nd Ed. John Wiley & Sons, 1963.
- AVILA, A. F. D., EVENSON, R. Total Factor Productivity Growth in Brazilian Agriculture and the Role of Agricultural Research. *Anais do XXXIII Congresso Brasileiro de Economia e Sociologia Rural*. Curitiba, 1995.
- BACH, L. Emerging common knowledge about the evaluation of public R&D programmes in teh context of European Union programmes. *Texto para Discussão*. DPCT, Unicamp. Campinas, 2001.
- BACH, L.; FURTADO, A. & LAMBERT, G. Variété des programmes de R&D, variété des méthodes d'évaluation, variété des effets économiques – quelques enseignements tirés de l'application de la méthode du BETA à différents programmes de R&D. *Avaliação de Programas Tecnológicos e Instituições de P&D*, Workshop Convênio CAPES–COFECUB. Campinas: UNICAMP/mimeo, 9 de novembro de 1998.
- BACH, L.; LAMBERT, G. Evaluation of the economic effects of large R&D programmes: the case of the European space programme. *Research Evaluation*, v. 2, n.1, pp. 17-26, 1992.
- BACH, L.; LAMBERT, G. Une Etude Transversale des Facteurs d'Innovation. *Avaliação de Programas Tecnológicos e Instituições de P&D*, Workshop Convênio CAPES–COFECUB. Campinas: UNICAMP/mimeo, 9 de novembro de 1998.
- BARBA-ROMERO, S. & POMEROL, J. C. *Decisiones Multicriterio: fundamentos teóricos y utilización práctica*. Colección de Economía. Universidad de Alcalá. Espanha, 1997.
- BAYES, T. An Essay Toward Solving a Problem in the Doctrine of Chances. *Philosophical Transactions of the Royal Society of London*. v. 53, p. 370-418, 1764.

- BELL, D. E., KEENEY, R. L., RAIFFA, H. (eds.). *Conflicting Objectives in Decisions*. Grã-Bretanha: John Willey & Sons, 1977.
- BENNETT, C. A., LUMSDAINE, A. A. Social Program Evaluation: Definitions and Issues. BENNETT, C. A., LUMSDAINE, A. A. (eds.) *Evaluation and Experiments – Some Critical Issues in Assessing Social Programs*. Academic Press, Inc. New York, 1975.
- BERLOZNIK, R., VAN LANGENHOVE, L. Integration of Technology Assessment in R&D Management Practices. *Technological Forecasting and Social Change*, v. 58, pp. 23-33, 1998.
- BERTALANFFY, L. VON. *Teoria Geral dos Sistemas*. Editora Vozes, Petrópolis, 1973.
- BIRKHOFF, *Lattice Theory*. 3rd Ed. Colloquium Publications, vol. 25, American Mathematical Society, 1967.
- BOLTANSKI, L., CHIAPELLO, E. *Le Nouvel Esprit du Capitalisme*. NRF Essais, Éditions Gallimard, France, 1999.
- BORUCH, R. F., RINDSKOPF, D. On Randomized Experiments, Approximation to Experiments, and Data Analysis. RUTMAN, L. (ed.) *Evaluation Research Methods: A Basic Guide*. Sage Publications, Beverly Hills, EUA, 1977.
- BOZEMAN, B., DIETZ, J. S., GAUGHAN, M. Scientific and Technical Human Capital: An Alternative Model for Research Evaluation. Paper prepared for presentation at the *American Political Science Association*, Atlanta, Georgia, September, 5, 1999.
- BRAVEMAN, H. *Trabalho e Capital Monopolista – A Degradação do Trabalho no Século XX*. 3a Edição. LTC – Livros Técnicos e Científicos Editora S. A. Rio de Janeiro, 1987.
- BRESCIANI FILHO, E. & D'OTTAVIANO, I. M. L. Conceitos Básicos de Sistêmica. In: D'OTTAVIANO, I. M. L. & GONZALES, M. E. Q. (org.) *Auto-Organização*. Coleção CLE n. 30, p. 283-306. Campinas, 2000.
- CALLON, M. Society in the Making: The Study of Technology as a Tool for Sociological Analysis. BIJKER, W. E. *et al.* (eds.) *The Social Construction of Technological Systems*. The MIT Press. Cambridge, 1990.
- CALLON, M. The dynamics of techno-economic networks. In: COOMBS, R.; SAVIOTTI, P.; WALSH, V. (Eds.) *Technological change and company strategies*. London, Academic Press. pp. 72-102, 1992.

- CALLON, M. Actor-network theory – the market test. LAW, J., HASSARD, J. *Actor Network Theory and After*. Blackwell Publishers. Oxford e Malden, 1999.
- CALLON, M.; LAREDO, P. MUSTAR, P. *La gestion stratégique de la recherche et de la technologie. L'évaluation des programmes*. Paris: Economica, 1995.
- CAMPBELL, D. T. "Degrees of Freedom" and the Case Study. COOK, T. D., REICHARDT, C. S. (eds.) *Qualitative and Quantitative Methods in Evaluation Research*. Sage Publications, Beverly Hills, EUA, 1979.
- CASTELLS, M. *The rise of the Network Society*. USA: Blackwell Publications, 1996.
- CASTELLS, M. Hacia el Estado Red? Globalization economica e instituciones politicas en la era de la information. *Seminário Internacional Sociedade e a Reforma do Estado*. São Paulo, 26 a 29 de março, 1998.
- CHUBIN, D. E., HACKETT, E. J. *Peerless Science: Peer Review and US Science Policy*. State University of New York Press, Alabny, NY, 1990.
- CIGNOLI, R. L. O. D'OTTAVIANO, I. M. L. MUNDICI, D. *Álgebras das Lógicas de Lukasiewicz*. Coleção CLE, v. 12. Centro de Lógica, Epistemologia e História da Ciência – Unicamp, Campinas, 1995.
- CLEMEM, R., REILLY, T. *Making Hard Decisions with Decision Tools*. Duxbury Thomson Learning. USA, 2001.
- COHON, J. L. *Multiobjective Programming and Planning*. Academic Press, New York, 1978.
- COLE, S., RUBIN, L., COLE, J. Peer Review and the Support of Science. *Scientific American*, v. 237, n. 4, pp. 34-41, 1977.
- CONANT, J. B. *Science and Common Sense*. Yale University Press, New Haven, 1963.
- COOK, K. S., LEVI, M. (Ed.) *The Limits of Rationality*. The University of Chicago Press, Chicago, 1990.
- COOK, T. D., COOK, F. L., MARK, M. M. Randomized and Quasi-Experimental Designs in Evaluation Research: an Introduction. RUTMAN, L. (ed.) *Evaluation Research Methods: A Basic Guide*. Sage Publications, Beverly Hills, EUA, 1977.
- COZZENS, S. E. *Literature-Based Data in Research Evaluation: A Manager's Guide to Bibliometrics*. Department of Science and Technology Studies, Rensselaer Polytechnic Institute. Troy, New York, 1989.

- COZZENS, S. E. Assessing federally-supported academic research in the United States. *Research Evaluation*, v. 8, n. 1, pp. 5-10, 2000.
- CROZIER, M. *A Sociedade Bloqueada*. Coleção Sociedade Moderna, v. 8. Editora Universidade de Brasília, 1983.
- CRUZ, E. R., PALMA, V., AVILA, A. F. D. *Taxas de Retorno dos Investimentos da EMBRAPA: investimentos totais e capital físico*. Departamento de Diretrizes e Métodos de Planejamento – EMBRAPA. Brasília, 1982.
- DA COSTA, N. C. A. O Ambiente Matemático no Século XIX e a Lógica do Século XX. ÉVORA, F. R. R. (ed.). *Século XIX: O Nascimento da Ciência Contemporânea*. Coleção CLE, v. 11. Centro de Lógica, Epistemologia e História da Ciência – Unicamp, Campinas, 1992.
- DA COSTA, N. C. A. *O Conhecimento Científico*. 2a. Ed. Discurso Editorial, São Paulo, 1999.
- DA COSTA, N. C. A., BÉZIAU, J. Y., BUENO, O. *Elementos de Teoria Paraconsistente de Conjuntos*. Coleção CLE, v. 23. Centro de Lógica, Epistemologia e História da Ciência – Unicamp, Campinas, 1998.
- DANTAS, C. A. B. *Probabilidade: Um Curso Introdutório*. 2ª edição. Editora da Universidade de São Paulo. São Paulo, 2000.
- DAVID, P. The Political Economy of Public Science: a Contribution to the Regulation of Science and Technology. SMITH, H. L. (ed.). *The Regulation of Science and Technology*. Palgrave Publishers Ltd, 2002.
- DAVID, P. A. & HALL, B. H. *Heart of Darkness: Public-Private Interactions Inside the R&D Black Box*. New York, 1999.
- DAVID, P. A.; HALL, B. H.; TOOLE, A. A. *Is Public R&D a Complement or Substitute for Private R&D? A Review of the Econometric Evidence*. Draft Version, 1999.
- DICKEN, P. *Global Shift: transforming the world economy*, 3rd ed. The Guilford Press. New York, 1998.
- DILLON, W. R., GOLDSTEIN, M. *Multivariate Analysis: Methods and applications*. New York: John Wiley e Sons, 1984.
- DRUKER, P. F. *Knowledge Work and Knowledge Society – The Social Transformations of this Century*. The 1994 Edwin L. Godkin Lecture. Harvard University's John F. Kennedy School of Government, 1994.

- EDWARDS, W., GUTTENTAG, M. Experiments and Evaluation: A Reexamination.
- BENNETT, C. A., LUMSDAINE, A. A. (eds.) *Evaluation and Experiments – Some Critical Issues in Assessing Social Programs*. Academic Press, Inc. New York, 1975.
- ETZKOWITZ, H. & LEYDESDORFF, L. The dynamics of innovation: from National Systems and “Mode 2” to a Triple Helix of university – industry – government relations. *Research Policy*. n. 29, p. 109-123, 2000.
- EVENSON, R. E. Productivity decomposition methods for evaluation of agricultural research systems impacts. In: EVENSON, R. E.; CRUZ, E. R. da; ÁVILA, A. F. D.; PALMA, V. *Economic evaluation of agricultural research: methodologies and Brazilian applications*. New Haven: Embrapa e Yale University, 1987.
- FERREIRA, C. *Tendências de Reorganização da Pesquisa: um estudo a partir de experiências internacionais*. Dissertação de Mestrado no Departamento de Política Científica e Tecnológica – DPCT/UNICAMP, 2001.
- FREEMAN, C. *Measurement of Output of Research and Experimental Development: a review paper*. Unesco, 1969.
- FURTADO, A. Apresentação. *Avaliação de Programas Tecnológicos e Instituições de P&D*, Workshop Convênio CAPES–COFECUB. Campinas: UNICAMP/mimeo, 9 de novembro de 1998.
- FURTADO, A. T. (coord.) *Políticas Públicas para a Inovação Tecnológica na Agricultura do Estado de São Paulo: Métodos para Avaliação de Impactos de Pesquisa – Relatório Final*. DPCT, Unicamp. Campinas, 2003.
- FURTADO, A. T., SOUZA, J. H. Levantamento das Principais Metodologias de Avaliação de Projetos e Programa de P&D no Eixo Rio-São Paulo. *VIII Seminario Latino-Iberoamericano de Gestión Tecnológica*, 1999.
- FURTADO, A.T., SILVEIRA, J. M. Avaliação de Impactos Econômicos. SALLES FILHO, S. (coord.). Revisão da Literatura: Impactos Econômicos, Sociais, Ambientais e de Criação de Competências. Documento 1. *Projeto Políticas Públicas para a Inovação Tecnológica na Agricultura do Estado de São Paulo: Métodos para Avaliação de Impactos de Pesquisa*. DPCT, Unicamp. Campinas, 2000.
- GAFFARD, J. L. Évaluation de la Recherche et Création de Technologie. In: BANDT, J. & FORAY, D. *L'Évaluation Économique de la Recherche et du Changement Technique*. Paris: Editions du CNRS, 1991.

- GARRET-JONES, S. International trends in evaluating university research outcomes: what lessons for Australia? *Research Evaluation*, v. 8, n. 2, pp. 115-124, 2000.
- GEORGUIOU, L. Research evaluation in European national science and technology systems. *Research Evaluation*. v. 5, n. 1, p. 3-10, 1995.
- GEORGUIOU, L., ROESSNER, D. Evaluating technology programs: tools and methods. *Research Policy*, v. 29, p. 657-678, 2000.
- GIBBONS, M., LIMOGES, C., NOWOTNY, H., SCHWARTZMAN, S., SCOTT, P., TROW, M. *The New Production of Knowledge: the Dynamics of Science and Research in Contemporary Societies*. Sage. London, 1994.
- GOLDBARG, M. C. & LUNA, H. P. *Otimização Combinatória e Programação Linear: modelos e algoritmos*. Rio de Janeiro: Editora Campus, 2000.
- GOMES, L. F. A. M. & DUARTE, V. C. A. Indicadores Tecnológicos em um Contexto Multicritério: fundamentação teórica e aplicação prática. *Anais do 5º Seminário Modernização Tecnológica Periférica*, pp. 168-185. Recife, 1997.
- GRÄTZER, G. *Lattice Theory – First Concepts and Distributive Lattices*. W. H. Freeman and Company. San Francisco, 1971.
- GRECO, S., MATARAZZO, B., SLOWINSKI, R. Rough sets theory for multicriteria decision analysis. *European Journal of Operational Research*, n. 129, p. 1-47, 2001.
- GRECO, S., MATARAZZO, B., SLOWINSKI, R. Rough sets methodology for sorting problems in presence of multiple attributes and criteria. *European Journal of Operational Research*, n. 138, p. 247-259, 2002.
- GUEDES, T. Redes de Inovação Tecnológica e Política de C&T. *Ciclos de Seminários de Pesquisa*. Núcleo de Política e Gestão Tecnológica da Universidade de São Paulo. São Paulo, 1999.
- HAIR, J. F., ANDERSON, R. E., TATHAM, R. L., BLACK, W. C. *Multivariate Data Analysis*. 5 ed. Upper Saddle River: Prentice Hall, 1998.
- HAM, C., HILL, M. *The Policy Process in the Modern Capitalist State*. St. Martin's Press, New York, 1984.
- HARNAD, S. (ed.) *Peer commentary on peer review: A case study in scientific quality control*. Cambridge University Press, NY, 1982.
- HARNAD, S. The Invisible Hand of Peer Review. *Exploit Interactive*, issue 5, April 2000.

- HERTZFELD, H. Space as an investment in economic growth. LOGSDOWN, J. M. *Exploring the Unknown, Selected Documents in the History of the US Civilian Space Program*, v. III. Using Space, The NASA History Series. NASA, Washington DC, 1998.
- HOGWOOD, B. W. & GUNN, L. A. *Policy Analysis for the Real World*. Oxford University Press, 1984.
- INGLEHART, R. *Culture Shift in Advanced Industrial Society*. Princeton University Press. Princeton, New Jersey, 1990.
- INGLEHART, R. *Modernization and Postmodernization – Cultural, Economic, and Political Change in 43 Societies*. Princeton University Press. New Jersey, 1997.
- JANTSCH, E. *Technological Forecasting in Perspective: A Framework for Technological Forecasting, its Techniques and Organisation*. OCDE. Paris, 1967. 401p.
- KEENEY, R. L. & RAIFFA, H. *Decisions with Multiple Objectives: preferences and value tradeoffs*. John Wiley & Sons. New York, 1976.
- KLEIN, H. A. *The Science of Measurement: a historical survey*. Dover Publications, Inc., New York, 1974.
- KNAPP, M. S. Ethnographic Contributions to Evaluation Research: The Experimental Schools Program Evaluation and Some Alternatives. COOK, T. D., REICHARDT, C. S. (eds.) *Qualitative and Quantitative Methods in Evaluation Research*. Sage Publications, Beverly Hills, EUA, 1979.
- KOSTOFF, R., AVERCH, H., CHUBIN, D. Research impact assessment: introduction and overview. *Evaluation Review*, v. 18, n. 1, pp. 3-10, 1994.
- KUHN, T. S. *A Estrutura das Revoluções Científicas*. 5ª Edição. Editora Perspectiva, São Paulo, 2000.
- LAAT, B. & LARÉDO, P. Foresight for research and technology policies: from innovation studies to scenario confrontation. In: EDUARD ELGAR (Ed.). *Technological Change and Organization*. p. 150-179, 1998.
- LATOUR, B. From the world of science to the world of research? *Science*, n. 280, p. 208-209, 1998.
- LATOUR, B. On recalling ANT. LAW, J., HASSARD, J. *Actor Network Theory and After*. Blackwell Publishers. Oxford e Malden, 1999.

- LINDBLOM, C. E. *O Processo de Decisão Política*. Tradução de Sérgio Bach. Editora Universidade de Brasília, Brasília, 1980.
- LINSTONE, H. A. *Complexity Science: Implications for Forecasting*. Technological Forecasting and Social Change, n. 62, p. 79-90, 1999.
- LOVERIDGE, D. Experts and Foresight: Review and experience. *Discussion Paper Series*. PREST, June 2002.
- LUMSDAINE, A. A., BENNETT, C. A. Assessing Alternative Conceptions of Evaluation. BENNETT, C. A., LUMSDAINE, A. A. (eds.) *Evaluation and Experiments – Some Critical Issues in Assessing Social Programs*. Academic Press, Inc. New York, 1975.
- LUNDVALL, B. A. (ed.) *National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning*. London: Pinter, 1992.
- MAKRIDAKIS, S.; WHEELWRIGHT, S. C.; MCGEE, V. E. *Forecasting: Methods and Applications*, 3. ed.. New York, John Wiley & Sons, 1998.
- MANSFIELD, E. *Industrial research and technological innovation : an econometric analysis*. 235p. W. W. Norton. New York , 1968.
- MARLOW, W. h. *Mathematics for Operational Research*. John Wiley & Sons, 1978.
- MARTIN, M. E. C., RÓDENAS, M. C. E., BARBERIS, G. F. Evolución Histórica de la Decisión Multicriterio. *Archives Internationales d'Histoire des Sciences*, v. 51, n. 147, 2001.
- MARTÍNEZ, E., ALBORNOZ, M. Indicadores de ciencia y tecnología: balance y perspectivas. MARTÍNEZ, E., ALBORNOZ, M. (eds.). *Indicadores de ciencia y tecnología: estado del arte y perspectivas*. Editorial Nueva Sociedad, Caracas, Venezuela, 1998.
- MARTÍNEZ, E., ESCUDEY, M. (org.) *Evaluación y Decisión Multicriterio: Reflexiones y Experiencias*. Editorial Universidad de Santiago/UNESCO. Santiago, 1998.
- MAYR, O. The science-technology relationship. BARNES, B., EDGE, D. (eds.) *Science in Context*. The Open University Press. Milton Keynes, 1982.
- MCCUBBINS, M. D., THIES, M. F. Rationality and the Foundations of Positive Political Theory (or, Rational Choice Without Apology). *Rebaisan*, n. 19, p. 7-32, 1996.
- McKENZIE, R. N., NULTY, G. F., TAYLOR, W. F. *Algebras, Lattices, Varieties Vol.1*. Wadsworth & Brooks/Cole Advanced Books & Software, Monterey, California, 1987.

- MENDELSON, E. *Introduction to Mathematical Logic*. The University Series in Undergraduate Mathematics. Van Nostrand, 1987.
- MERTON, R. K. El efecto Mateo en la Ciencia. MERTON, R. K. *La Sociologia de la Ciencia*. Alianza Editorial. Madrid, 1977.
- MERTON, R. K. Os Imperativos Institucionais da Ciência. DEUS, J. D. (org.). *A Crítica da Ciência*. Zahar Editores. Rio de Janeiro, 1979.
- MILES, I., KEENAN, M., KAIVO-OJA, J. *Handbook of knowledge society foresight*. PREST, Manchester, 2002.
- MOIGNE, J. L. *La théorie du système général: théorie et modélisation*. 4e ed. Paris: Presses Universitaires de France, 1994.
- MOULIN, H. *Axioms of cooperative decision making*. Cambridge University Press, Cambridge, 1988.
- MULLER, D. C. Public Choice in Perspective. In: MULLER, D. C. *Perspectives on public choice: a handbook*. Cambridge University Press, Cambridge 1997.
- MUNDLAK, Y. *Agriculture and Economic Growth: Theory and Measurement*. Harvard University Press. Cambridge, Massachusetts, 2000. cap 5, 11, 12, 13.
- NELSON, R. (ed.) *National Innovation Systems – a comparative analysis*, New York: Oxford University Press, 1993.
- NELSON, R., WINTER, S. *A evolutionary theory of economic change*. Harvard University Press. Cambridge, Massachusetts, 1982.
- NGUYEN, H. T., WALKER, E. A. *A First Course of Fuzzy Logic*. 2nd edition. Chapman & Hall/CRC, Boca Raton, 2000.
- NIETZSCHE. *Além do Bem e do Mal: prelúdio a uma filosofia do futuro*. Companhia das letras, São Paulo, 1999.
- NOWOTNY, H., SCOTT, P., GIBBONS, M. *Re-Thinking Science: Knowledge and the Public in an Age of Uncertainty*. Polity Press. London, 2001.
- NYE, M. J. *Before Big Science – The Pursuit of Modern Chemistry and Physics (1800 – 1940)*. Harvard University Press. Cambridge, Massachusetts, 1996.
- OCDE. Science, Technology and Society: Public Opinion and Assessment Mechanisms. *Science and Technology Policy*. Paris, 1992.
- OCDE. *Improving Evaluation Practices: Best Practice Guidelines for Evaluation and Background Paper*. Paris, 1999.

- OFFE, C. The present historical transition and some basic design options for societal institutions. *Seminário Internacional Sociedade e a Reforma do Estado*. São Paulo, 26 a 29 de março, 1998.
- PARETO, V. *Manual de Economia Política*. v. 1. Trad. João G. Vargas Netto. Coleção Os Economistas. São Paulo, Abril, 1984.
- PATTANAIK, P. K. Some paradoxes of preference aggregation. In: MULLER, D. C. *Perspectives on public choice: a handbook*. Cambridge University Press, Cambridge 1997.
- PEREIRA, J. C. R. *Análise de dados qualitativos: Estratégias Metodológicas para as Ciências da Saúde, Humanas e Sociais*. São Paulo, Editora da Universidade de São Paulo, 1999.
- POPPER, K. *A Lógica da Pesquisa Científica*. Tradução de Leonidas Hegenberg e Octanny Silveira da Mota. Editora Cultrix e Editora da Universidade de São Paulo, São Paulo, 1975.
- PORTER, M. *The Competitive Advantage of Nations*. New York: The Free Press, 1990.
- POSSAS, M. L. As principais abordagens contemporâneas à economia da mudança tecnológica: um comentário. *Projeto Mudança Técnica e Natureza do Trabalho – uma avaliação crítica dos efeitos da informatização*. NPCT/Unicamp, 1987.
- PRAHALAD, C. K., HAMEL, G. A competência essencial da corporação. In: MONTGOMERY, C., PORTER, M. (Eds.). *Estratégia – A busca da vantagem competitiva*. Ed. Campus, 1998.
- PRICE, D., J., S. *O Desenvolvimento da Ciência – Análise Histórica, Filosófica, Sociológica e Econômica*. Trad. de Simão Mathias. Livros Técnicos e Científicos Editora S. A. Rio de Janeiro, 1976.
- QUINE, W. O. *O Sentido da Nova Lógica*, 2a edição. Editora UFPR, Curitiba, 1996.
- RAPPERT, B. Rationalising the future? Foresight in science and technology policy coordination. *Futures*, v. 31, pp. 527-545, 1999.
- REICHARDT, C. S., COOK, T. D. Beyond Qualitative Versus Quantitative Methods. COOK, T. D., REICHARDT, C. S. (eds.) *Qualitative and Quantitative Methods in Evaluation Research*. Sage Publications, Beverly Hills, EUA, 1979.
- ROCHE, C. A. *Avaliação de Impacto dos Trabalhos de ONGs: aprendendo a valorizar as mudanças*. Cortez Editora, Oxfam/ABONG. São Paulo, 2000.

- ROSENBERG, N. *Inside the black box: technology and economics*. Cambridge University Press, 1982.
- ROWLAND, F. *The Peer Review Process: A Report to the JISC Scholarly Communications Group*. Loughborough University, 2003.
- ROY, B. Partial Preference Analysis and Decision-Aid: the fuzzy outranking relation concept. In: BELL, D. E., KEENEY, R. L., RAIFFA, H. (eds.). p. 40 – 75. *Conflicting Objectives in Decisions*. Grã-Bretanha: John Willey & Sons, 1977.
- ROY, B., BOUYSSOU, D. *Aide Multicritère à la Décision: Méthodes et Cas*. Economica, Paris, 1993.
- ROY, R. Funding science: the real defects of peer review and an alternative to it. *Science, Technology & Human Values*, v. 10, n. 3, pp. 77-81, 1985.
- RUELLE, D. *Acaso e Caos*. Editora da Unesp. São Paulo, 1993.
- RUSSEL, B. *História da Filosofia Ocidental*. Vol. 1. Companhia Editora Nacional, Rio de Janeiro, 1960.
- RUSSEL, B. *História do Pensamento Ocidental: a aventura das idéias dos pré-socráticos a Wittgenstein*. Tradução de Laura Alves e Aurélio Rebello. Ediouro, Rio de Janeiro, 2002.
- RUTMAN, L. Planning an Evaluation Study. RUTMAN, L. (ed.) *Evaluation Research Methods: A Basic Guide*. Sage Publications, Beverly Hills, EUA, 1977.
- SAATY, T. L. *Mathematical Methods of Operational Research*. MacGraw-Hill Book Company, 1959.
- SAATY, T. L. *Toma de decisiones para líderes*. Pittsburgh, RSW Publishing, 1997.
- SALLES-FILHO, S. (coord.) et al. *Ciência, Tecnologia e Inovação – A reorganização da pesquisa pública no Brasil*. Editora Komedi. Campinas, 2000.
- SALLES FILHO, S. (coord.). Revisão da Literatura: Impactos Econômicos, Sociais, Ambientais e de Criação de Competências. Documento 1. *Projeto Políticas Públicas para a Inovação Tecnológica na Agricultura do Estado de São Paulo: Métodos para Avaliação de Impactos de Pesquisa*. DPCT, Unicamp. Campinas, 2000a.
- SALOMON, J. J. Maîtrise sociale de la technologie : l'enjeu démocratique. *Futuribles*. n. 157, pp 3-18, sept. 1991.
- SALOMON, J. J. *Le Destin Technologique*. Collection Folio/Actuel, Gallimard, 1992.

- SALTER, A. J., MARTIN, B. R. The economic benefits of public funded basic research: a critical review. *Research Policy*, v. 30, p. 509-532, 2001.
- SCHUMPETER, J. A. *Capitalism, Socialism and Democracy*. G. Allen & Unwin. Londres, 1976.
- SCHUMPETER, J. A. *Teoria do Desenvolvimento Econômico – Uma Investigação Sobre Lucros, Capital, Crédito, Juro e o Ciclo Econômico*. Coleção Os Economistas. Tradução de Maria Sílvia Possas. Abril Cultural. São Paulo, 1982.
- SCOTT, W. R. *Institutions and Organizations*. 2nd Edition. Sage Publications, Thousand Oaks, 2001.
- SEN, A. *Choice, Welfare and Measurement*. Harvard University Press, Cambridge, 1997.
- SILVA, R. T. *Eficiência e Eficácia da Ação Governamental: uma análise comparativa de sistemas de avaliação*. Relatório Técnico. Fortalecimento da Função Avaliação nos Países da América do Sul. Projeto de Cooperação Técnica BID – Ipea. Brasília, 2002.
- SIMON, H. A. *El Comportamiento Administrativo: Estudio de los Procesos Decisorios en la Organización Administrativa*. 2ª Ed. Aguilar, Madrid, 1962.
- SO, C. Y. K. Citation ranking versus expert judgment in evaluating communication scholars: effects of research specialty size and individual prominence. *Scientometrics*, v. 41, n. 3, pp. 325-333, 1998.
- SOUZA, J. A. Fomento da Tecnologia Industrial. *Avaliação e Perspectivas em Ciência e Tecnologia*. Relatório Técnico n. 12. Diretoria de Planejamento e Gestão, CNPq, 1990.
- SCHMÉDER, G. Prévision technologique, rétrospective critique. *Futuribles*. n. 124, 1988.
- STERN, E. Ongoing and participative evaluation: purpose, design and role in the evaluation of a large-scale R&D programme. *Research Evaluation*. v. 3, n. 2, p. 75-82, 1993.
- SUDMAN, S., BRADBURN, N. M., SCHWARZ, N. *Thinking About Answers – The Application of Cognitive Processes to Survey Methodology*. Jossey-Bass Publishers. San Francisco, 1996.
- TARSKI, A. *Introduction to Logic and to the Methodology of Deductive Sciences*. Translated by Olaf Helmer. Dover Publications, Inc., New York, 1961.
- TEECE, D. Profiting from technological innovation: implications for integration, collaboration, licensing and public policy. *Research Policy*, v. 15, n. 6, p. 285-305, 1986.

- TOMIZAWA, H. NIWA F. Evaluating overall national science and technology activity: General Indicator of Science and Technology (GIST) and its implications for S&T policy. *Research Evaluation*, v. 6, n. 2, pp. 83-92, 1996.
- VALLÉ, D., ZIELNIEWICZ, P. *Electre III-IV: Aspects Méthodologiques (tome 1 et 2)*. Document n. 85. Laboratoire d'Analyse et Modélisation de Systèmes pour l'Aide à la Décision, Université Paris-Dauphine, 1994.
- VAN DEN ENDE, J., MULDER, K., KNOT, M., MOORS, E. VERGRAGT, P. Traditional and Modern Technology Assessment: Toward a Toolkit. *Technological Forecasting and Social Change*, v. 58, pp. 5-21, 1998.
- VAN DER MEULEN, B. Science policies as principal – agent games: institutionalization and path dependency in the relation between government and science. *Research Policy*, n. 27, p. 397-414, 1998.
- VAN RAAN, A. F. J. Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises. *Scientometrics*, v. 36, n. 3, pp. 397-420, 1996.
- VANBERG, V. J. *Rules and choice in economics*. Routledge, London, 1994.
- VEBLER, T. *A Teoria da Classe Ociosa*. Coleção Os Economistas. Abril Cultural, São Paulo, 1983.
- VEBLER, T. The Place of Science in Modern Civilization. MITCHELL, W. C. (ed.). *What Veblen Taught: Selected Writings of Thorstein Veblen*. Reprint of Economic Classics. Augustus M. Kelley, New York, 1964.
- VELHO, L. Avaliação acadêmica. A hora e a vez do "baixo clero". *Ciência e Cultura*. Revista da Sociedade Brasileira para o Progresso da Ciência, v. 41, n. 10, p. 957-968, 1989.
- VELHO, L. Estratégias para um sistema de indicadores de C&T no Brasil. *Parcerias Estratégicas*, n. 13, pp. 109- 121. Brasília, dezembro de 2001.
- VELHO, L. Indicadores científicos: aspectos teóricos y metodológicos e impactos en la política científica. MARTÍNEZ, E., ALBORNOZ, M. (eds.). *Indicadores de ciencia y tecnología: estado del arte y perspectivas*. Editorial Nueva Sociedad, Caracas, Venezuela, 1998.
- VELHO, L. Indicadores de C&T e seu uso em Política Científica. *Sociedade e Estado*, v. 7, n. 1-2, pp. 63-77, 1992.

- VON NEUMANN, J. & MORGENTHAU, O. *Theory of Games and Economic Behavior*. 3rd ed. Princeton University Press, Princeton, 1953.
- WAGNER, H. M. *Pesquisa Operacional*. 2o ed. Rio de Janeiro: Prentice/Hall do Brasil, 1986.
- WEBER, M. *Sobre a Teoria da Ciências Sociais*. Tradução de Rubens Eduardo Frias. Editora Moraes, São Paulo, 1991.
- WEBER, M. *Economía y Sociedad – Esbozo de sociología comprensiva*. Segunda edición en español de la cuarta en alemán, 15ª reimpresión. Fondo de Cultura Económica. México D. F., 2004.
- WHISTON, T. (ed.). *The Uses and Abuses of Forecasting*. Sussex: SPRU, 1979.
- WHOLEY, J. S. Evaluability Assessment. RUTMAN, L. (ed.) *Evaluation Research Methods: A Basic Guide*. Sage Publications, Beverly Hills, EUA, 1977.
- WILLIAMS, D., RANK, A. D. Measuring the economic benefits of research and development: the current state of art. *Annals of the APEC Symposium of S&T Programmes*. Wellington-New Zealand, 2-4 December, 1998.
- WORTHEN, B. R., SANDERS, J. R., FITZPATRICK, J. L. *Avaliação de Programas: concepções e práticas*. Tradução de Dinah de Abreu Azevedo. Editora Gente e Edusp. São Paulo, 2004.
- WYNNE, B. The rhetoric of consensus politics: a critical review of technology assessment. *Research Policy*, n. 4, p. 108-158, 1975.
- YOUNG, H. P. Group choice and individual judgments. In: MULLER, D. C. *Perspectives on public choice: a handbook*. Cambridge University Press, Cambridge 1997.
- ZACKIEWICZ, M. *A Definição de Prioridades de Pesquisa a Partir da Abordagem de Technological Foresight*. Dissertação de Mestrado. DPCT/Unicamp, 2000.
- ZACKIEWICZ, M. *Modelos para Avaliação de Impactos da Pesquisa: introdução a uma discussão axiomática*. Seminário de Tese. DPCT/Unicamp, 12 de dezembro de 2001.
- ZACKIEWICZ, M. Coordenação e organização da inovação: perspectivas do estudo do futuro e da avaliação. *Parcerias Estratégicas*, n. 17, p. 193-214, 2003.
- ZACKIEWICZ, M., ALBUQUERQUE, R. H. P. L., SALLES-FILHO, S. L. M. Multicriteria Analysis for the selection of priorities in the Brazilian Program of Technological Prospection – Prospectar. *Innovation: Management, Policy & Practice*. n.7, v.1, 2005.

- ZADEH, L. A., FU, K. S., TANAKA, K., SHIMURA, M. (eds.) *Fuzzy Sets and Their Applications to Cognitive and Decision Process*. New York: Academic Press, 1975.
- ZUFFO, A. C. *Seleção e aplicação de métodos multicriteriais ao planejamento ambiental de recursos hídricos*. Tese de Doutorado. São Carlos: EESC/USP, 1998.

Anexo 1 – Detalhamento do Método Multicritério Electre III

O método ELECTRE III – ELimination Et Choix Traduisant la REalité – foi criado por Bernard Roy em 1978. Pertence a uma família mais ampla de métodos multicritério e está inserido na chamada tradição francesa de métodos multicritério de apoio à decisão. Essa escola se destaca pela sofisticação dos algoritmos e pela premissa de ajusta-los às sutilezas dos processos de decisão em situações reais. Comparando com os métodos de otimização linear (e seus derivados) mostram-se mais flexíveis e mais abertos à interferência do usuário, com o inconveniente de não serem facilmente comunicáveis devido à grande quantidade de parâmetros e passos em seus algoritmos.

Este texto tem o objetivo de detalhar e esclarecer o algoritmo utilizado pelo método ELECTRE III. Para isso, inicialmente todos seus parâmetros serão listados e brevemente comentados. Em seguida, o algoritmo será explicado seguindo os passos de aplicação do método. Os conceitos necessários para o entendimento dos procedimentos de uso serão apresentados e comentados no decorrer da exposição.

AS DEFINIÇÕES NECESSÁRIAS

O método ELECTRE III é particularmente indicado para problemas de hierarquização: definições de prioridades e ordenação das alternativas de ação segundo seu desempenho relativo umas às outras. Para gerar uma hierarquia de alternativas o método precisa de:

1. *As Alternativas*. Codificadas como a, b, ..., n, as alternativas são elementos discretos e independentes entre si, de preferência situadas no "mesmo nível", isto é, sem grande variação em termos de particularidade ou generalidade entre si, candidatas à oferecer resposta ao problema de decisão posto. Em problemas mais triviais as alternativas são objetos, pessoas ou projetos que precisam ser hierarquizados segundo sua capacidade de responder aos critérios de decisão colocados. Entretanto, muitas vezes, dado um problema complexo é difícil de identificar *a priori* soluções alternativas a ele, nesses casos, a

hierarquização é parte da solução de um problema, indicando quais elementos são mais importantes para essa solução. Nessas situações, a definição de alternativas se torna mais abstrata, é preciso criar uma partição no conjunto de elementos do problema identificados como invariantes (ou como variável independente) e tomar os demais elementos como variáveis de decisão, ou seja, como critérios.

2. *Os Critérios.* Codificados como $g_1, g_2, \dots, g_j$ os critérios são as variáveis pelas quais se mede cada alternativa segundo os aspectos relevantes ao problema de decisão. Assim para cada alternativa a , um vetor com as valorações $g_1(a), g_2(a), \dots, g_j(a)$ precisa ser obtido. O conjunto F de todos os critérios é chamado de família de critérios e um conjunto F coerente é aquele que obedece a três axiomas fundamentais: exaustividade, não-redundância e coesão. Obedecer o axioma de exaustividade significa que a família de critérios é suficiente para gerar a decisão, a não-redundância determina que nenhum critério é ocioso nem tampouco se sobrepõe aos demais, por fim a coesão pressupõe que o instrumento de medida utilizado para cada critério é estável e independente das alternativas que mede.

Obs 1. No método ELECTRE III cada critério g pode ter sua própria escala para medir as diferentes alternativas. Conforme ficará claro adiante, a comparação entre alternativas prescinde da "unidade" em que foi definida medida do critério e ocorrendo apenas internamente a cada critério não há necessidade de normalizar as escalas (como ocorre nos métodos de otimização linear).

3. *Os Pesos.* A cada critério g se associa um peso k , que denota a importância relativa que este critério tem no conjunto F . O modelo de decisão precisa ter um ou mais vetores de pesos $[k_1, k_2, \dots, k_j]$ determinados. Os pesos são atribuídos pelo decisor. Desta forma, ele é obrigado a pensar sobre a importância dos critérios e não diretamente na importância das alternativas. Construir mais de um vetor de pesos é equivalente a estabelecer mais de uma "visão" sobre o problema, que depois podem ser comparadas em busca de alternativas robustas.

Obs 2. Os pesos não precisam somar 1 como nos métodos de otimização linear.

Obs 3. Se uma alternativa a possuir um vetor de valoração $[g_1(a), g_2(a), \dots, g_j(a)]$ tal que $\forall g_i(a) \geq g_i(b)$ então a sempre irá superar b independentemente do vetor peso associado ao problema (a será sempre uma alternativa *dominante*).

4. *Os Limites de Indiferença (q)*. Parâmetro que modifica a relação de indiferença entre as alternativas a e b, denotada aIb . O limite de indiferença é modelado a partir da escala de critério g_i e do valor que este assume em relação a a e b. Na prática, escreve-se $q_i[g_i(a)] = \alpha_i g_i(a) + \beta_i$. Desse modo, para cada critério i tem-se um parâmetro α_i dependente do valor assumido pela alternativa a e um parâmetro β_i que é independente. Se $q_i[g_i(a)] = q_i[g_i(b)] = 0$ tem-se o caso trivial de indiferença, isto é, $aIb \Leftrightarrow g_i(a) - g_i(b) = 0 \Leftrightarrow bIa$. Por outro lado, quando o limite de indiferença é ativado, aIb se $q_i[g_i(a)] \geq g_i(a) - g_i(b) \geq -q_i[g_i(a)]$ e bIa se $q_i[g_i(b)] \geq g_i(b) - g_i(a) \geq -q_i[g_i(b)]$. É fácil notar que nem sempre aIb e bIa valem simultaneamente.

5. *Os Limites de Preferência (p)*. Parâmetro que modifica a relação de preferência aPb entre as alternativas a e b. O limite de preferência é modelado como $p_i[g_i(a)] = \delta_i g_i(a) + \phi_i$. Desse modo, para cada critério i tem-se um parâmetro δ_i dependente do valor assumido pela alternativa a e um parâmetro ϕ_i independente. Se $-p_i[g_i(a)] \geq g_i(a) - g_i(b) \geq p_i[g_i(a)]$ diz-se que aPb e se $-p_i[g_i(b)] \geq g_i(b) - g_i(a) \geq p_i[g_i(b)]$ diz-se que bPa . Nem sempre aPb e bPa valem simultaneamente.

Obs 4. Por definição, $p_i[g_i(a)] \geq q_i[g_i(a)]$.

Obs 5. Os parâmetros p e q quando aplicados a um critério g tornam este um *pseudocritério*. Isto porque, pela definição convencional, um critério é sempre capaz de discriminar entre os elementos de um conjunto. Tecnicamente, um critério g induz sobre o conjunto das alternativas duas relações (I, P), sendo a primeira reflexiva, simétrica e transitiva e a segunda irreflexiva, anti-simétrica e transitiva. Essas relações são necessárias e suficientes para a definição de critérios capazes de particionar o conjunto das alternativas. A introdução dos parâmetros p e q altera essas relações. As relação I torna-se apenas reflexiva e a P irreflexiva, isto é, aIa e $\neg(aPa)$ são as únicas certezas remanescentes. Na prática, é como se o critério tenha se esmaecido, incluindo novas nuances de comparação antes não presentes; o designativo de pseudo critério faz referência à incapacidade da medida escolhida em distinguir perfeitamente entre as alternativas. O recurso ganha relevância prática porque pode evitar que uma alternativa avaliada em 1,2 pontos seja considerada

melhor que outra com 1,1 pontos considerando que a terceira colocada tenha obtido 0,7, por exemplo.

Obs 6. Além das relações I e P aparece uma nova relação Q, chamada *preferência fraca*, para modelar a região da escala quando a diferença $g_i(a) - g_i(b)$ cai no intervalo entre $p_i[g_i(a)]$ e $q_i[g_i(a)]$, ou seja, não se trata de indiferença, mas também ainda não é estritamente preferência.

6. *Os Limites de Veto (v)*. O parâmetro de veto também é definido sobre a escala de cada critério. Tem-se para a alternativa a $v_i[g_i(a)] = \gamma_i \cdot g_i(a) + \eta_i$ de modo similar aos casos anteriores. O valor mínimo para o parâmetro veto é $v_i[g_i(a)] \geq p_i[g_i(a)]$, que consiste na forma mais intensa de veto. A forma mais branda, equivalente a não considerar o efeito de veto, ocorre quando $v_i[g_i(a)] = \max g_i(x)$, isto é, é igual ao maior valor obtido para a escala do critério (seu fundo de escala).

Obs 7. O efeito de veto é uma forma complementar ao peso para indicar a importância de um critério. No algoritmo do método, se a diferença $g_i(a) - g_i(b)$ for maior que o valor $v_i[g_i(a)]$ no critério i, mesmo que para todos os outros $j - i$ critérios bPa , b não poderá superar a, porque o veto no critério i impedirá.

7. *O Limite de Discriminação do Índice de Credibilidade ($s(\lambda)$)*. Trata-se de um parâmetro técnico global do algoritmo que tem a função de regular o passo das destilações realizadas para chegar à hierarquia final. O método, como será visto em detalhe adiante, produz duas hierarquias intermediárias: a destilação descendente, que ordena as alternativas a partir das melhores para as piores, e a destilação ascendente, que ordena das piores para as melhores alternativas. Estas duas pré ordens completas são cruzadas e a intersecção dá origem a uma pré-ordem parcial que consiste na hierarquia final das alternativas. O parâmetro é constituído de outros dois parâmetros, de modo que $s(\lambda) = \mu \cdot \lambda + \rho$, com $\mu + \rho > 0$ e $0,5 \geq \rho$. O método utiliza como padrão $s(\lambda) = -0,15 \cdot \lambda + 0,3$ e os autores não recomendam que seja mudado. A variável λ está definida no intervalo de $1 \geq \lambda \geq 0$ e é identificada com o índice de credibilidade da relação de superação aSb , denotado $\sigma_s(a,b)$, utilizado pelo algoritmo. A função $s(\lambda)$ regula o tamanho do decréscimo no valor de $\sigma_s(a,b)$ considerado como referência para determinar a relação aSb durante o processo de destilação.

Obs 8. Para valores mínimos de μ e ρ , por exemplo $\mu = 0$ e $\rho = 0,01$, as pré-ordens completas obtidas tenderão a ter tantos níveis quanto forem as alternativas. Para valores máximo de μ e ρ , por exemplo $\mu = 0$ e $\rho = 0,5$, haverá poucos passos de destilação e o número de alternativas por nível tenderá a crescer excessivamente enquanto avança a destilação, tornando-a pouco discriminante. Valores negativos de μ indicam que no processo de destilação cada passo subsequente para a extração do (s) representante(s) do nível se torna maior.

ALGORITMO DA RELAÇÃO DE SUPERAÇÃO (S)

O método ELECTRE III é baseado na relação de superação. Esta relação é inspirada numa afirmação de Condorcet de 1785: “Quando uma alternativa A é ao menos tão boa quanto outra B em uma maioria de critérios, e não há nenhum critério em que A seja notoriamente inferior a B, podemos afirmar sem risco que A supera B”.

Fica claro que a relação de superação aS_b pode tolerar que em alguns critérios b tenha melhor desempenho que a, desde que esse desempenho não seja "notoriamente" superior. Esse tipo de sutileza, comum em casos reais de decisão, pode ser adequadamente modelado pelo algoritmo do método.

Uma vez definidos todas alternativas, os critérios e seus parâmetros associados e obtidas as valorações que cruzam alternativas e critérios constrói-se uma matriz de decisão completa do tipo:

	critérios					
	1	2			j	
	k_1	k_2	.	.	k_i	
	q_1	q_2	.	.	q_i	
	p_1	p_2	.	.	p_i	
	v_1	v_2	.	.	v_i	
alternativas	a	$g_1(a)$	$g_2(a)$	.	.	$g_i(a)$
	b	$g_1(b)$	$g_2(b)$	.	.	$g_i(b)$
	.	.	.	.	.	.
	.	.	.	.	.	.
	n	$g_1(n)$	$g_2(n)$	.	.	$g_i(n)$

A matriz de decisão é o ponto de partida para o algoritmo do método ELECTRE III. O algoritmo consiste em duas etapas principais: 1. obtenção da matriz dos índices de credibilidade e 2. obtenção da pré-ordem parcial que representa a hierarquia final.

Etapa 1 – Obtenção da matriz dos índices de credibilidade

Para cada par de alternativas a e b, uma medida $\sigma_s(a,b)$ definida entre 0 e 1 fornece o grau de credibilidade da afirmação "a supera b". Calculado σ_s para todos os pares de alternativas, obtém-se a matriz dos índices de credibilidade:

	a	b	.	.	n
a	1	$\sigma_s(a,b)$	.	.	$\sigma_s(a,n)$
b	$\sigma_s(b,a)$	1	.	.	.
.	.	.	1	.	.
.	.	.	.	1	.
n	$\sigma_s(n,a)$	.	.	.	1

Nessa matriz, cada índice do tipo $\sigma_s(a,b)$ mede a credibilidade da relação aSb e possui uma contrapartida $\sigma_s(b,a)$ que mede a relação bSa. Por definição $\sigma_s(a,a) = 1$ uma vez que relação S é reflexiva. O valor assinalado para o índice de credibilidade vai depender dos diversos parâmetros contidos na matriz de decisão.

Desse modo, o método ELECTRE III calcula $\sigma_s(a,b)$ como sendo:

$$\sigma_s(a,b) = c(a,b) \cdot \prod_{j \in D_c(a,b)} \frac{1 - d_j(a,b)}{1 - c(a,b)} \quad \text{Equação A.1}$$

onde $c(a,b)$ é o grau de concordância global da relação aSb e $d_j(a,b)$ é o grau de discordância da relação aSb para o (pseudo)critério j.

Na Equação A.1, o termo mais à direita (produtório) só é considerado para os critérios da família F cujos graus de discordância de que aSb sejam maiores que o grau de concordância sobre a mesma afirmação. Esses critérios compõem o conjunto $D_c(a,b)$ definido como:

$$D_c(a, b) = \{j \in F : d_j(a, b) > c(a, b)\}$$

É fácil notar da Equação A.1 que, se:

$$\exists d_j(a, b) = 1 \Rightarrow \sigma_s(a, b) = 0, \text{ e se}$$

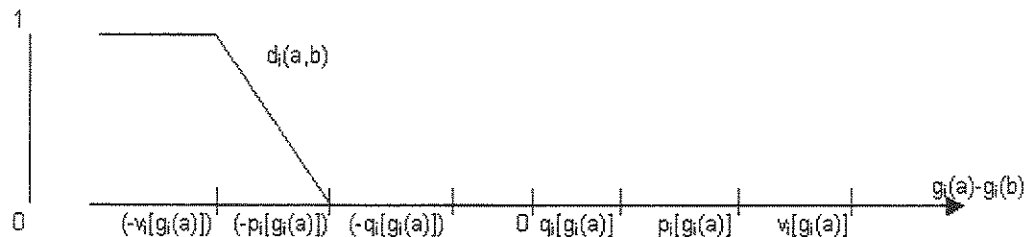
$$\forall d_j(a, b) = 0 \Rightarrow \sigma_s(a, b) = c(a, b)$$

Esses são os dois casos extremos da Equação 1. Se houver pelo menos um $d_j(a, b) = 1$ a expressão se anula e se não houver nenhum $d_j(a, b) > 0$ o termo contendo o produtório se iguala a 1. Nas situações intermediárias, a Equação 1 desconta progressivamente o valor obtido para o grau de concordância, tanto mais critérios apresentarem graus de discordância elevados.

O cálculo do grau de discordância depende essencialmente dos valores dos limites de preferência e de veto.

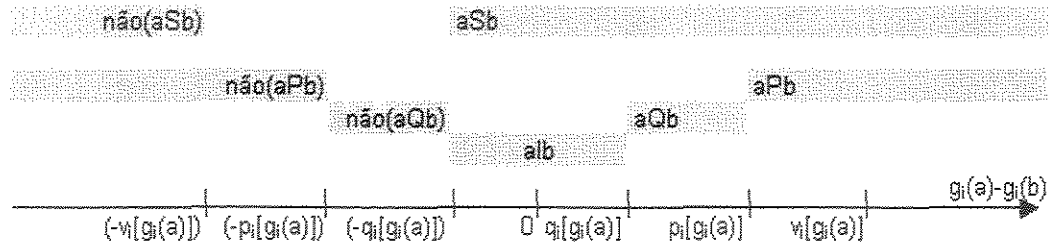
$$d_j(a, b) = \frac{g_j(b) - g_j(a) - p_j[g_j(a)]}{v_j[g_j(a)] - p_j[g_j(a)]} \quad \text{Equação A.2}$$

A interpretação da Equação A.2 pode ser feita graficamente. A medida do grau de discordância varia necessariamente entre 0 e 1. Suponha a reta dos valores $g_i(a) - g_i(b)$:



O grau de discordância $d_i(a,b)$ passa a ser diferente de 0 quando a relação $\text{não}(aPb)$ entra em vigor e se iguala a 1 quando a diferença $g_i(a) - g_i(b)$ suplantar o limiar imposto pelo limite de veto.

Resumindo graficamente, temos as seguintes estruturas de preferências definidas para a comparação entre duas alternativas a e b.



Note que para a comparação entre b e a, os limites q, p e v podem ser definidos diferentemente. Portanto não é um fato necessário que $\text{não}(aPb)$ implique bPa , por exemplo.

O grau de concordância global (c) entre duas alternativas a e b é composto pelos pesos atribuídos aos critérios nos quais aSb acrescidos dos pesos onde $\text{não}(aQb)$, estes últimos descontados progressivamente a medida que o valor da diferença $g_i(a) - g_i(b)$ se aproxime do limite de preferência a favor de $\text{não}(aPb)$.

$$c(a, b) = \frac{K[C(aSb), C(\text{não}(aQb))]}{K[F]} = \frac{\sum_{j \in C(aSb)} k_j - \sum_{j \in C(\text{não}(aQb))} \phi_j k_j}{\sum_{j \in F} k_j}$$

Equação A.3

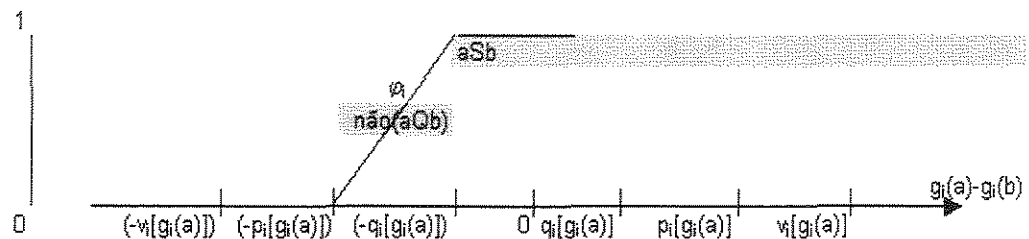
O operador K soma seletivamente os pesos para diferentes subconjuntos da família de critérios F. O resultado é dividido pela soma de todos os pesos da família F. Os critérios pertencentes ao subconjunto $C(aSb)$ são somados diretamente e os pertencentes ao subconjunto $C(\text{não}(aQb))$ são descontados do fator ϕ . Esses conjuntos e o fator ϕ são definidos como segue:

$$C(aSb) = \{j \in F : g_j(a) + q_j[g_j(a)] \geq g_j(b)\}$$

$$C(\text{não}(aQb)) = \{j \in F : g_j(a) + q_j[g_j(a)] < g_j(b) \leq g_j(a) + p_j[g_j(a)]\}$$

$$\varphi_j = \frac{g_j(a) + p_j[g_j(a)] - g_j(b)}{p_j[g_j(a)] - q_j[g_j(a)]}$$

Os elementos da Equação A.3 podem ser visualizados graficamente. Tudo se passa como se o fator φ se aplicasse a todos os critérios de F . Nos casos que nem aSb nem $\text{não}(aQb)$ valem $\varphi = 0$ e os pesos desses critérios não são computados. Nos demais seguem a Equação A.3 já discutida.



Etapa 2 – Obtenção da hierarquia final

A hierarquia final das alternativas é uma pré-ordem parcial Z formada pela interseção de duas pré-ordens completas Z_d e Z_a , chamadas respectivamente de *destilação descendente* e *destilação ascendente*. A hierarquia final é uma pré-ordem parcial porque nela nem todos os elementos estabelecem relações de comparação claras entre si. Na hierarquia final, tem-se que uma alternativa a ou é preferível a b (aPb), ou é preterida por b ($aP'b$), ou é indiferente a b (aIb) ou não é comparável a b (aRb).

O algoritmo para o cálculo das destilações descendente e ascendente parte da matriz dos índices de concordância construída na Etapa 1.

	a	b	.	.	n
a	1	$\sigma_s(a,b)$	.	.	$\sigma_s(a,n)$
b	$\sigma_s(b,a)$	1	.	.	.
.	.	.	1	.	.
.	.	.	.	1	.
n	$\sigma_s(n,a)$	.	.	.	1

As destilações são processos de escolha que em sucessivas interações retiram do conjunto A que contém todas as alternativas a hierarquizar as melhores e as piores alternativas. O algoritmo define classes dC e aC que as contém as retira de A, reduzindo o conjunto A para a próxima interação, até construir duas pré-ordens completas com tais classes e esgotar o conjunto A. O Limite de Discriminação do Índice de Credibilidade ($s(\lambda)$) regula o processo de destilação. O processo completo consiste em sete operações com dois laços e é bastante similar para as duas pré-ordens, por isso ambas as destilações serão descritas simultaneamente, apesar de serem realizadas de modo independente (em particular, o número k de passos para chegar a uma classe e o número m de classes obtidas ao final são diferentes para cada destilação).

Condições de partida:

- $m = 0$ (parâmetro de controle do laço principal, presente em A_m e C_m)
- ${}^dA_0 = A$ e ${}^aA_0 = A$. Ambas as destilações partem do mesmo conjunto A que contém todas as alternativas do problema de decisão.

i) Faz-se:

- $\lambda_0 = \max_{a, b \in A_m, a \neq b} \sigma_s(a, b)$. O parâmetro λ_0 é o máximo índice de credibilidade encontrado na matriz de credibilidade, excetuando os valores da diagonal principal. Trata-se do patamar inicial de partição para o conjunto A_m .
- $k = 0$ (parâmetro de controle do laço interno, presente em λ_k e D_k)
- $D_0 = {}^dA_m$ ou $D_0 = {}^aA_m$.

ii) $\lambda_{k+1} = \max_{\substack{a, b \in D_k \\ a \neq b}} \sigma_s(a, b)$. O cálculo de λ_{k+1} define o próximo patamar de partição para o conjunto D_k segundo o valor calculado para $s(\lambda_k)$. O tamanho do salto dado aqui depende

dos parâmetros já discutidos μ e ρ que compõem a função $s(\lambda)$. Importante: se para todo a e b de D_k não houver $\sigma_s(a,b) > \lambda_k - s(\lambda_k)$ então, por definição $\lambda_{k+1} = 0$.

iii) Matriz de Superação. Para todo par de alternativas a e b de D_k define-se:

$$aS_{D_k}^{\lambda_k} b \Leftrightarrow [\sigma_s(a,b) > \lambda_k] \wedge [\sigma_s(a,b) > \sigma_s(b,a) - s[\sigma_s(a,b)]]$$

Com isso é possível calcular a λ_{k+1} -qualificação de cada alternativa a . Por definição:

- λ_{k+1} -qualificação: $q_{D_k}^{\lambda_{k+1}}(a) = p_{D_k}^{\lambda_{k+1}}(a) - f_{D_k}^{\lambda_{k+1}}(a)$, onde
- λ_{k+1} -potência: $p_{D_k}^{\lambda_{k+1}}(a) = |\{b \in D_k : aS_{D_k}^{\lambda_{k+1}} b\}|$, isto é, o número de vezes que a λ_{k+1} -supera b (a cardinalidade do conjunto que contém todos os b que são λ_{k+1} -superados por a).
- λ_{k+1} -fraqueza: $f_{D_k}^{\lambda_{k+1}}(a) = |\{b \in D_k : bS_{D_k}^{\lambda_{k+1}} a\}|$, isto é, o número de vezes que b λ_{k+1} -supera a .

iv) Uma vez realizado o cálculo das λ_{k+1} -qualificações para todas alternativas de D_k seleciona-se as com λ_{k+1} -qualificação máxima ou as com λ_{k+1} -qualificação mínima dentre todas as n alternativas. Obtém-se o conjunto:

$$\bullet \quad {}^d D_{k+1} = \left\{ a \in D_k : q_{D_k}^{\lambda_{k+1}}(a) = \max_{x \in D_k} q_{D_k}^{\lambda_{k+1}}(x) \right\}$$

ou

$$\bullet \quad {}^a D_{k+1} = \left\{ a \in D_k : q_{D_k}^{\lambda_{k+1}}(a) = \min_{x \in D_k} q_{D_k}^{\lambda_{k+1}}(x) \right\}$$

v) Se $|{}^d D_{k+1}| = 1$ ou $|{}^a D_{k+1}| = 1$ ou $\lambda_{k+1} = 0$ passa-se à operação vi. Senão, faz-se $k = k + 1$ e retorna-se à operação ii.

vi) Faz-se:

${}^d C_{m+1} = {}^d D_{k+1}$, que representa a classe de alternativas retida pela $(m+1)$ -ésima destilação descendente.

ou

${}^a C_{m+1} = {}^a D_{k+1}$, que representa a classe de alternativas retida pela $(m+1)$ -ésima destilação ascendente.

vii) Faz-se:

${}^d i_{m+1} = {}^d i_m \cup {}^d C_{m+1}$ (v.l.) ${}^a i_{m+1} = {}^a i_m \cup {}^a C_{m+1}$, retirando-se as classes obtidas do conjunto das alternativas ${}^d A_m$ ou ${}^a A_m$.

Se: ${}^d A_{m+1} \neq \varnothing$ ou ${}^a A_{m+1} \neq \varnothing$ faz-se $m = m + 1$ e volta-se a i. Senão, termina a destilação.

A destilação descendente produz a pré-ordem completa

$$Z_d = \left\{ {}^d C_1 > {}^d C_2 > \dots > {}^d C_n : \bigcup_1^n {}^d C_i = J \right\}$$

A destilação ascendente produz a pré-ordem completa

$$Z_a = \left\{ {}^a C_n > {}^a C_{n-1} > \dots > {}^a C_1 : \bigcup_1^n {}^a C_i = J \right\}$$

A hierarquia final procurada é dada por $Z = Z_d \cap Z_a$. Assim, na hierarquia final:

- uma alternativa a é melhor classificada que uma alternativa b (aPb) se a é melhor classificada em uma das pré-ordens Z_d e Z_a e pelo menos tão bem classificada quanto b em outra;
- uma alternativa a é indiferente a uma alternativa b (aIb) se elas são indiferentes nas duas pré-ordens;
- uma alternativa a é incomparável a uma alternativa b (aRb) se a é melhor classificada que b em uma pré-ordem e b é melhor classificada que a em outra.

A partir dessas três regras é possível construir uma matriz que assinala as relações P, I ou R a todos os pares de alternativas existentes. Dessa matriz de classificação final pode-se obter ainda três formas diferentes de se visualizar os resultados:

a) *Grafo*. Trata-se de uma representação gráfica das classificações obtidas. A alternativa dominante, que é preferível ou indiferente a todas as outras, encabeça o grafo. Uma flecha indica a próxima dominante, isto é, aquela que domina as demais mas é dominada pela primeira, e assim por diante. As alternativas indiferentes entre si aparecem juntas (sem flechas entre elas) e as alternativas incomparáveis são representadas num mesmo nível, são dominadas pelo nível acima e dominam o nível abaixo, mas não estabelecem nenhuma ligação entre si;

b) *Níveis da Pré-Ordem Parcial*. Trata-se de uma lista com o número de classes obtido na interseção das duas pré-ordens e a indicação de quais alternativas compõem cada nível. Tem o inconveniente de não distinguir alternativas indiferentes e incomparáveis, listadas conjuntamente como elementos de um mesmo nível. Das alternativas listadas conjuntamente apenas se sabe que são dominadas pelas que estão acima e que dominam as que estão abaixo;

c) *Pré-Ordem Mediana*. Trata-se de uma pré-ordem completa obtida a partir da pré-ordem parcial Z com o uso do seguinte recurso: as ações incomparáveis em cada nível são pontuadas pelo valor da diferença de colocação nas pré-ordens descendente e ascendente. As que tiverem menores diferenças são melhor classificadas. A comparação entre a pré-ordem parcial Z e a pré-ordem mediana fornece em decorrência um indicador da estabilidade relativa entre as pré-ordens descendente e ascendente.