An Analysis of a Sustainability Index

Stig Blomskog
Senior Lecturer in Economics
Södertörn University
e-mail: Stig.Blomskog@sh.se
Magnus Hjelmblom
Senior Lecturer in Computer Science
University of Gävle
e-mail: Magnus.Hjelmblom@hig.se

No abstract, key words.

* We wish to thank our discussants at the book workshop at Stockholm School of Economics (October, 2019) for their fruitful comments on our initial draft. We also wish to thank the editors of the book, Ranjula Bali Swain and Susanne Sweet, for the invitation, and Ulla Ahonen-Jonnarth and the editors for their insightful comments and valuable suggestions.

INTRODUCTION

Sustainable consumption is today an important policy goal, which is supposed to contribute to an overall sustainable development. It is therefore important that policy makers and other stakeholders have access to valid measurement or evaluation models regarding sustainable consumption. Today, there is also extensive research and development of measurement models regarding sustainable consumption. The main purpose of this chapter, which is primarily directed towards researchers and practitioners (e.g., policy makers) within the field of sustainability, is a critical analysis of the common usage of the concept sustainability and how it is measured. We focus on sustainability with respect to (household) consumption, and use a recently developed household sustainable consumption index for 28 EU countries as an example. The index determines a ranking of the countries regarding sustainable consumption, i.e., the index is a kind of ordinal measure regarding sustainable consumption of these countries. It should be noted that the construction of the index is supported by guidelines stated in the handbook about how to construct composite indicators issued by OECD (2008). The putative purpose of the handbook is to serve as a guideline for the construction of valid measures or indices regarding various aspects of sustainability.

We argue that the construction of the previously mentioned index and many others seems to be grounded on a misinterpretation of the meaning or functioning of the concept sustainability. The designers of the index seem to interpret sustainability as a descriptive concept, a mistake that we (following Odelstad 2017a: 31) refer to as the *descriptive* or *realistic mistake*. Instead, we argue that sustainability functions as an *intermediate concept*. Briefly, an intermediate concept functions as a link between different sorts of concepts, such as descriptive concepts and normative concepts. Intermediate concepts are common in legal, ethical and evaluative contexts. We are not aware of any previous studies where the concepts *sustainable consumption* and *sustainability* are explicitly treated as intermediate concepts. We therefore

stress that our analysis should be regarded as a preliminary treatment of a very complex conceptual problem. To make the analysis tractable, it is performed as far as possible by means of an informal non-algebraic language.

Finally, to avoid misinterpretation of our analysis, we stress that the purpose of our conceptual analysis is not to propose and/or elaborate any specific methods for the construction of a sustainable consumption index. As we argue in the chapter, the construction of a sustainable consumption index is a normative decision process. As such, it requires both (1) specific domain knowledge about the evaluation context and situation, and (2) adequate knowledge of the semantic function of the concept sustainability. Furthermore, since sustainability is a multidimensional concept, the process also benefits from (3) knowledge of multi-criteria decision analysis. As we point out in the chapter, the designer of a sustainability index should understand, for example, that the choice of a specific statistical method is a kind of normative decision, which should ultimately be grounded on normative reasoning. Since we do not possess the required domain knowledge, we make no attempt to prescribe a certain design of a sustainability index. Instead, our contribution focuses on (2) and (3).

The chapter is organised as follows: In the next section, we present evidence for our hypothesis about the descriptive mistake in the context of sustainable consumption. In the third section, we introduce the theory of intermediate concepts. In the fourth section, based on the theory of intermediate concepts we analyse the construction of a sustainable consumption index for 28 EU-countries. In the fifth section, we summarize the conclusions of the analysis.

COMMON USAGE OF THE CONCEPT SUSTAINABLE CONSUMPTION

Introduction

In this section, we examine how the concepts sustainable consumption and sustainability are commonly used. The examination functions as a background to the interpretation that we

suggest in the following sections. It is both practical and very common to talk about a living environment, or society, or some aspect of society such as household consumption, being sustainable. The notion of sustainable development could then be understood as a development that leads from a 'less sustainable' society to a 'more sustainable' society. Sustainable consumption is then used as a *comparative* concept, meaning that we can talk about different degrees of sustainability and compare different objects with respect to their sustainability. In such a context, the term 'sustainability with respect to consumption' is perhaps more natural than 'sustainable consumption'. Another common application of sustainable consumption is to label objects *sustainable* or *not sustainable* (i.e., *unsustainable*). Sustainable consumption is then treated as a *two-valued categorical concept*. We will in the following, despite a slight risk of confusion, refer to both applications of sustainability as *sustainable consumption*, and assume that the domain is a set of countries. When it is absolutely necessary to distinguish between the two applications, the term *sustainability with respect to aggregated household consumption* will be used for sustainable consumption as a comparative concept.

Construction of a Sustainable Household Consumption Index: A Case Study

Today there is vast literature on constructions of measurement and evaluation models regarding sustainable consumption. We take Bartolj, Murovec, Slabe-Erker (2018) as a case study and examine how the concept sustainable consumption is understood and used in this research field. The examination starts from our hypothesis that sustainable consumption is often misinterpreted as a descriptive concept.

Bartolj et al., (2018) develop an index named Household Sustainable Consumption Index (HSCI). The domain of the index is the aggregate household consumption in 28 EU countries in 2005, 2010 and 2015. The authors' starting point of the construction is as follows:

> "The *multidimensional phenomenon* of sustainable consumption, which is *measured* in our research, is understood in line with the definition given by the Oslo Roundtable (1994) and is divided in three major areas: (1) the use of goods and services that respond to basic needs and bring better quality of life, (2) minimizing the use of natural resources, toxic materials and emissions of waste and pollutants over the life cycle and (3) not jeopardizing the needs of future generations." (Bartolj et al., 2018: 36, italics ours).

Based on the three major areas (1–3), stipulated at the Oslo Roundtable Conference, and a method, named Social Progress Index methodology (OECD 2008), the authors design a formal decision process for selection of relevant variables for the construction of the HSC-index. They select 17 variables related to education, people's habits, material situation of households, health outcome, food and beverage consumption and the environment. These measures are inputs into a statistical algorithm named *factor analysis*, the outcome of which is that the variables are reduced into two dimensions. The index is then constructed by aggregating these two dimensions by means of a so-called geometric aggregation rule.

A central term in the quotation above is "multidimensional", which means "having or relating to multiple dimensions or aspects" (Merriam-Webster, n.d.). In the following, the terms *dimension*, *variable*, and *aspect* will be used more or less interchangeably, usually preferring the latter. Examples of aspects include area, temperature, age, loudness, archaeological value (cf. Odelstad, 2019:105), and carbon dioxide emissions. Note the important distinction between aspects that are *descriptive* (empirical or factual; e.g., temperature or CO_2 levels in the atmosphere) and *non-descriptive*, i.e., normative or evaluative concepts or *intermediaries* that function as links descriptive and normative conceptual systems. Odelstad (2019:106) gives a number of examples of so-called "policy relevant concepts", i.e. concepts of importance for planning as well as decision and policy making, that are multidimensional intermediate

concepts: measurements of national income, inequality, poverty, gross domestic product, inflation, unemployment, gender equality, and "other 'indicators' defined with normative motivation incorporating interpersonal weighting in some easily tractable way" (Sen, 1977: 53). Understanding the nature of an aspect is crucial for understanding what it means to measure it and how to interpret the result of the measurement.

An Interpretation of the Constructed HSC-index

The HSC-index determines a ranking over the 28 EU-countries regarding sustainable household consumption. The index implies, for example, that regarding sustainable household consumption, the Netherlands is in 2015 ranked above Latvia, i.e., that household consumption in the Netherlands is *more sustainable than* household consumption in Latvia (Bartolj et al., 2018, Figure 3: 43).

A crucial question is: How do Bartolj et al. interpret such a statement implied by the index? Is it interpreted as a *descriptive* or as a *kind of normative statement*? *On the one hand*, (i) it seems that the authors interpret the concept 'sustainable household consumption' as a kind of descriptive concept. It is unclear how the authors interpret the term "phenomenon", which occurs in the quotation above, but a common view in the scientific context is that this term has a descriptive connotation, i.e., it refers to some empirical magnitude (Merriam-Webster, n.d.). This suggests that the authors assume that the sentence, "The household consumption in the Netherlands is much more sustainable than the household consumption in Latvia," corresponds to a factual statement regarding some kind of empirical relationship.

On the other hand, (ii) Bartolj et al., (2018) seem to assume that the index implies guidelines and action-guiding norms for policy makers and other stakeholders:

> "The results should be at the disposal of policymakers as well as the general public in order to give everybody a rough idea about whether a country is on *the right, sustainable path* to well-being." (Bartolj et al., 2018: 45, italics ours).

According to the index, the household sustainable consumption in the Netherlands is on the right path, or at least closer to it compared to Latvia. But the claim that the household consumption in Latvia is "not on the right path", and should be adjusted such that the consumption patterns is approaching the right path, is obviously a *normative statement* (in the form of an *action-guiding norm*). A reasonable interpretation is that the index implies that the household consumption in Latvia should be adjusted such that the consumption pattern becomes more consistent with a sustainable consumption pattern.

The problem that arises for Bartolj et al. is how to justify these action-guiding norms if the authors believe that sustainable consumption is a descriptive concept. According to the view (often referred to as Hume's law[i]) that no ought-judgment may be correctly inferred from a set of premises expressed only in terms of 'is', either their reasoning relies on a set of hidden normative premises, or sustainability is not a descriptive concept.

Measuring Descriptive and Non-descriptive Concepts

Another central concept in the quotation by Bartolj et al. (2018: 36) is "measure". What does it mean to measure sustainable consumption? As Odelstad (2017a: 31) remarks, the notion of measurement is not unproblematic. The common, classical view is that to measure how much some object x has of some aspect α, involves applying a method for determining if x has more or less of α than some other object. That is, measurement in this narrow sense presupposes the existence of an 'objective' *decision method* that is tied to the meaning of the concept, and that produces the same result if performed by different competent users. The term measurement, however, may also be used in the broader sense of representing an aspect (either descriptive,

i.e., some empirical phenomenon, or normative, e.g., a valuation) with a numerical structure. Inspired by Odelstad (2017b: 26), we will refer to the former (narrower) kind of measurement, which is relevant only for descriptive aspects, as *measurement*-M (M for decision Method). The measure that is the result of measurement-*M* will be called an M-*measure*. Examples include measures of length, area, temperature, age, velocity and the level of CO_2 in the atmosphere. Measuring these concepts means applying a well-specified measurement method that is part of the meaning of the concept, like placing unit-length sticks on a straight line, edge-to-edge, alongside the object whose length is to be measured, and counting the sticks.

Furthermore, we will use the term *measurement*-R (R for numerical Representation) for measurement in the broader sense of representing an aspect (either descriptive or normative) with a numerical structure. A measure constructed by means of measurement-*R* will be called an R-*measure*. Measures of the previously mentioned multidimensional concepts—national income, inequality, poverty, gross domestic product, inflation, unemployment and gender equality—are examples of *R*-measures, i.e., measures that represent valuations and not descriptive phenomena. We will later return to the issue of what it means to measure multidimensional concepts, and argue that such concepts can be measured in the sense of measurement-*R* but not in the sense of measurement-*M*. Note that, since measurement-*M* is a subset of measurement-*R*, it is a fallacy to conclude that an aspect is measurable in the sense measurement-*M*, and hence is descriptive, from the fact that an aspect is measurable in the sense measurement-*R* (Odelstad, 2017b: 27). However, measurement-*R* of multidimensional concepts usually presupposes measurement-*M* of various underlying descriptive dimensions.

Returning to the notion of sustainability with respect to household consumption and the HSC-index, it is not unreasonable to believe that at least some of the 17 variables that constitute the basis for the HSCI are descriptive phenomena, and thus can be measured in the sense of applying a decision method, i.e., in the sense of measurement-*M*. But even if all variables

represent descriptive phenomena (and thus are measured by *M*-measures), it does not mean that the subsequent application of factor analysis and geometric aggregation to the measures of these variables, constitutes a decision method that is tied to the meaning of some phenomenon called 'sustainability with respect to household consumption'. While we certainly agree with Bartolj et al., (2018) that this concept depends on many dimensions or aspects, we believe that by treating this phenomenon as an empirical phenomenon that can be measured in the sense of measurement-*M*, one commits what we previously referred to as the descriptive mistake. One may of course, like Bartolj et al. (2018), propose a particular well-specified procedure for measuring sustainability, but since the procedure itself is the result of normative decisions regarding how to valuate and aggregate several dimensions, it represents a valuation and is not tied to the meaning of the concept. Thus, it is an *R*-measure and not an *M*-measure, and sustainability is not a descriptive concept. Furthermore, the decision to adopt this measurement procedure means taking a normative stance, and accepting its normative consequences (e.g., that a lower ranked country is to a lesser degree "on the right path to sustainable well-being").

Conclusions of the Analysis of the Index Construction

We conclude that the quotation of Bartolj et al. (2018) above supports the hypothesis that the concept 'sustainable consumption' is often misinterpreted, in that it is assumed to refer to some kind of a multidimensional yet descriptive phenomenon. This mistake gives rise to the conceptual problem of how this descriptive concept can imply action-guiding norms. So, if sustainable consumption is neither a purely descriptive concept, nor a purely normative concept, what kind of concept is it and how should a measure of it be understood? In the following sections we will discuss these questions. The conceptual framework that is required may be unfamiliar to some readers, but we will use it informally and as far as possible avoid the technical and mathematical notions necessary for a more detailed formal analysis.

A BRIEF INTRODUCTION TO THE THEORY OF INTERMEDIATE CONCEPTS

In this section, we explain in an informal way the functioning of intermediate concepts and argue that 'sustainable consumption' should be understood as an intermediate concept, more precisely an *intermediate aspect* (Odelstad, 2019: 105).

We will start by a reinterpretation of a fictitious example by Odelstad (2019: 107f). Let us suppose that some traffic policy stipulates that the air pollution caused by traffic should decrease by 5 per cent each year. To evaluate the effects of this policy, it is necessary to examine whether this objective (which is expressed in quantitative terms) is met. This means that 'air pollution caused by traffic' must be measured. It is, of course, possible to separately measure (in the sense measurement-*M*) traffic-related emission of different kinds of gases (e.g. nitrous gases) and harmful particles of different sizes. But it is less obvious how to trade different kinds of emissions off against each other. One can argue that the stipulated objective is met if the emissions within each individual category decreases by 5 per cent, but what if, for example the emission of nitrous gases is reduced by 20 per cent while the emission of a particular kind of harmful particles is not reduced at all? Furthermore, if we also include fossil fuel CO_2 emissions as part of air pollution, how do we trade off such emissions against emission of nitrous gases and harmful particles, respectively? Apparently, *air pollution* is a multidimensional concept, and there is no straightforward way to measure it.

In the original example, it is some bus traffic policy that stipulates that the punctuality of buses should increase. As Odelstad (2019: 108) remarks, one could, for example, select an easily measurable aspect of punctuality and regard it as an operationalization of punctuality, or one could select some important aspects of punctuality and aggregate them, or one could aggregate as many aspects of punctuality as possible. Whichever may be the case, the choice involves taking a normative stance. Thus, bus punctuality is not a purely descriptive concept, and therefore not measurable in the sense of measurement-M. On the other hand, it is not a purely normative concept either. Instead, it is an *intermediate concept* with descriptive grounds and

normative consequences, and to measure it means to evaluate its grounds with regard to the normative consequences. Due to the structural similarity with bus punctuality, the same argument can be applied to the concept of air pollution. Both concepts depend on measurable facts, but *how* they depend on those facts is a normative issue, i.e., involves valuations.

The notion of intermediate concepts is also known under names such as *intermediaries*, *ground-consequence-terms*, *middle terms* or *coupling terms*, and is related (cf. Odelstad, 2019: 105) to the notion of *thick concepts* discussed in as diverse areas as ethics (e.g. concepts like *cruelty*, *courageousness*, *kindness*; see for example Kirchin, 2013; Väyrynen, 2016) and risk analysis (e.g., *risk* and *safety*; see for example Möller, 2009). A thorough overview of the discourse on intermediate concepts is given in, for example, Lindahl & Odelstad (2013, section 1.7).

A classic example of an intermediate concept is ownership, whose meaning is tied to its function as a syntactic tool for formulating legal rules and a 'vehicle of inference' for legal reasoning. Its role is to link factual grounds for ownership with legal consequences of ownership. In this view, ownership is attached to certain facts, and different legal or normative positions are attached to ownership. The term *being the owner of* functions as a bridge between different conceptual systems, one containing facts (e.g., events, actions, or circumstances) and one containing normative positions like obligations, claims, legal powers, etc. In other words, the term ownership connects legal information of two different sorts, factual (descriptive) and normative, and ownership is in itself neither a purely descriptive nor a purely normative concept. A schematic view of ownership as an intermediate concept is given in Figure 2 in Hjelmblom et al. (2019), referring to Lindahl & Odelstad (2013: 553). Other examples of intermediaries in a legal context are *citizenship*, *relationship similar to be married* (see Lindahl & Odelstad, 2013: 557), *work of equal value* (see Odelstad, 2017a), and *enduringly suited to its purpose* (see Hjelmblom et al., 2019).

Deciding on how to measure a multidimensional aspect such as air pollution or bus punctuality means taking a normative stand, and is part of clarifying the meaning of the concept (Odelstad, 2019: 105–107). Figure 1 shows a simple *aggregation tree* that visualizes the structure of the aggregation problem in Bartolj et al. (2018).

[Figure 1 here]

The bottom factors b_1 - b_{17} could represent a number of basic aspects, the aggregate aspect S_0 at the top could represent 'sustainability with respect to household consumption', while S_1 and S_2 in the middle stratum are intermediate concepts. Note that, in some sense, higher up in the tree means 'more normative', while lower down means 'more descriptive'.

As will be further discussed in later, intermediate concepts often form networks (like chains or trees; see for example Lindahl & Odelstad, 2013; Odelstad, 2019), so that what constitutes a consequence of a certain concept in turn constitutes a ground for another concept. In Figure 1, S_1 and S_2 are intermediate concepts (more precisely, intermediate aspects) whose grounds are the basic aspects b_1, b_2, …, b_{10} and b_{11}, b_{12}, …, b_{17}. These intermediate concepts constitute in turn the grounds for S_0, sustainability with respect to household consumption. This concept may in turn be one of the grounds for a broader notion of sustainability, taking also other factors (like sustainability with respect to production) into account.

AN ANALYSIS OF THE CONSTRUCTION OF THE HSC-INDEX

Introduction

In this section we perform a critical analysis of the construction of the HSC-index based on our claim that sustainable consumption has a similar function as the intermediate concepts, 'bus punctuality' and 'air pollution', discussed in the previous section, i.e., that it functions neither as a purely descriptive nor a purely normative concept. Thus, we believe that under the design of the HSC-index lies an insufficient understanding of the nature of the concept 'sustainable

consumption', that negatively affects the validity of the index. To measure this concept, i.e., to construct an index such as the HSC-index, means to evaluate the grounds for sustainability with respect to household consumption with regards to its normative consequences. This evaluation is a paradigmatic example of a normative multi-attribute decision process. Such a process cannot reasonably be founded on objective decision rules. This means that the HSC-index cannot be interpreted as an *M*-measure, but must be interpreted as an *R*-measure that is a numerical representation of the outcome of a multidimensional normative decision process.

We start the reconstruction of the HSC-index by introducing a simple formal conceptual framework. The domain of sustainable consumption consists of 28 EU countries c_i, and is denoted as:

$C = \{c_1, c_2, \ldots\ldots, c_{28}\}$.

The HSC-index is apparently intended only to be used for ranking the objects in *C*, which means that "distances between the countries measured on the horizontal axis do not represent the differences in the absolute level of sustainability household consumption" (see Bartolj et al., 2018: 42). We thus treat the HSC-index as an ordinal measure, denoted HSC, that determines a ranking of the countries regarding *sustainable consumption.* That is,

c_i is more sustainable (with respect to household consumption) than c_j if and only if

$\mathrm{HSC}(c_i) > \mathrm{HSC}(c_j)$,

and

c_i and c_j are equally sustainable (w.r.t household consumption) if and only if

$\mathrm{HSC}(c_i) = \mathrm{HSC}(c_j)$.

Our claim that 'sustainable consumption' functions as an intermediate concept means that statements such as $\mathrm{HSC}(c_i) > \mathrm{HSC}(c_j)$ are neither purely descriptive (empirical) statements

nor purely normative statements. Thus, such statements do not refer to a kind of multidimensional empirical phenomenon. Briefly stated, we claim that an adequate interpretation is that a statement such as $\mathrm{HSC}(c_i) > \mathrm{HSC}(c_j)$ functions as a *link* or *bridge* between grounds in terms of facts about the two countries and possible normative consequences in terms of various action guiding norms. A conceptual and logical analysis of such links requires the introduction of a formal framework, but here we simply perform an informal analysis focusing on two questions:

What are the possible normative consequences of a statement such as $\mathrm{HSC}(c_i) > \mathrm{HSC}(c_j)$?

What are the grounds for a statement such as $\mathrm{HSC}(c_i) > \mathrm{HSC}(c_j)$?

These questions will be discussed in the following two subsections.

For example, Figure 3 (in Bartolj et al., 2018: 43) holds that $\mathrm{HSC}(F) > \mathrm{HSC}(P)$, where *F* and *P* denote France and Poland, respectively. What are the possible normative consequences of this statement for (e.g.) Poland's government? What are the grounds for this statement? A reasonable claim is that a policy maker should be able to understand the kinds of index design decisions that have been taken that lead to this particular ranking, otherwise the policy maker might take irrational policy decisions. That is, it should be possible for Poland's government to figure out why France is ranked above Poland, in order to take action. Furthermore, suppose that Poland's government is to decide between two suggested policy changes, one which leads to an improvement in aspect b_1 (i.e., the leftmost arrow in figure 1) but a deterioration in aspect b_3, and one which leads to an improvement in aspect b_2 but a deterioration in aspect b_4, and status quo with respect to the other aspects. It should then be possible to reason about which of these two actions leads to the highest degree of sustainability with respect to household consumption. Thus, a reasonable requirement is that the decision process, including the trade-offs made, should be transparent for the user of the index.

The Consequence-side of Household Sustainable Consumption

Although the intention of the HSC-index designers is that the index should be used as an ordinal measure of sustainability as a comparative concept, it could in principle also serve as the basis of a corresponding two-valued categorical concept. One way is the following: Imagine a (real or fictitious) reference object in domain *C* that is considered to be 'just barely' sustainable. For example, let us assume that this applies to France. Then for each *c* in *C* it holds that,

c is sustainable if and only if $\mathrm{HSC}(c) \geq \mathrm{HSC}(F)$,

otherwise it is not sustainable. Note that 'at least as sustainable as' is a binary relation (a *predicate*) on *C*, while 'is sustainable' is a unary relation (a *property*). On the consequence-side of sustainable consumption we might find various action-guiding norms that prescribe or put restrictions on certain actions. The basic idea behind the comparison of two countries (e.g. the Netherlands, *NL*, and Latvia, *L*) with respect to sustainability seems to be the following: If $\mathrm{HSC}(NL) > \mathrm{HSC}(L)$ on one measurement occasion, but $\mathrm{HSC}(NL) < \mathrm{HSC}(L)$ on a later occasion, then this can be seen as a signal to the policy-makers of the Netherlands to increase their efforts towards higher sustainability. Or if $\mathrm{HSC}(NL') < \mathrm{HSC}(NL)$, where *NL'* represents the Netherlands at some later occasion, then this suggests that the Netherlands is "not on the right sustainable path to well-being" and should take action.

Sustainable consumption as a two-valued categorical aspect may imply, in its turn, various action-guiding norms. For example, if some country *c* is found to be 'not sustainable' with respect to household consumption, then a possible normative consequence is that *c*'s government should (for example by legislation, by creation of economic incentives or by other means) try to affect the household consumption patterns of *c* towards higher sustainability with respect to household consumption:

If *c* is *not sustainable*, then the government of *c* should try to change the consumption patterns of *c*.

Furthermore, Bartolj et al. argue that "the sustainable path to well-being should be built upon sustainable consumption" (2018: 34). A plausible interpretation in terms of the theory of intermediate concepts is that sustainable consumption is one of the grounds for a broader notion of sustainability that also takes other aspects (e.g. sustainability with respect to production) into account. This broader notion of sustainability (e.g. 'overall sustainability' or 'sustainability all things considered') is then an aggregate of several intermediate concepts that constitute its grounds, and it may in turn imply various action-guiding norms that regulate decision making on various levels. That is, the concept 'overall sustainability' is found on the consequence-side of sustainable consumption. In the aggregation tree in figure 1, this aggregate aspect would represent a fourth stratum at the top of the tree, above S_0. Naturally, the aggregation of these different sustainability dimensions into a higher-level notion of sustainability requires more weighing decisions, where different aspects of sustainability must be traded off against each other.

A theoretical (algebraic) tool for the rational reconstruction of complicated conceptual systems containing intermediaries is the so-called Theory of Joining Systems, developed by Lindahl and Odelstad (see for example Lindahl & Odelstad, 2013; Odelstad, 2019). However, an account of the formal treatment of complex networks or strata of intermediate concepts of various kinds, and the logical form of action-guiding norms, is beyond the scope of this investigation.

The Ground-side of Household Sustainable Consumption

In this subsection we examine what kind of design decisions that give rise to statements such as $\mathrm{HSC}(F) > \mathrm{HSC}(P)$. Since we treat sustainable consumption as a multidimensional intermediary, we understand the construction of the index as a normative multi-criteria decision

process consisting of a number of stages. A good starting point, to get an overview of the decision process, is an aggregation tree (see figure 1) that visualises the structure of the aggregate aspects involved. The tree displays three 'decision points', one regarding the aggregation of basic aspects $b_1, b_2, \ldots, b_{10}$ into S_1, one regarding the aggregation of $b_{11}, b_{12}, \ldots, b_{17}$. into S_2, and one regarding the aggregation of S_1 and S_2 into S_0 (i.e. sustainable consumption).

The design decisions express valuations (i.e. normative standpoints) that should be linked to the intended normative consequences implied by the index. A problem here is the 'open' nature of the HSC-index, that is, that its consequences are not entirely specified (regarding so-called *open intermediaries*, see for example Lindahl & Odelstad, 2013). Again, we stress that we do not possess the kind of extensive domain knowledge of empirical facts, cultural values, formal legal rules as well as informal normative rules etc. that is required to make substantial comments on the normative process to construct a sustainability index. Thus, our analysis is intended as a formal analysis of the normative decision process, not as a material analysis of the 'best' or 'correct' way to make trade-offs between different dimensions or what kind of evaluation model that would be appropriate.

As previously described, the construction of the index takes place mainly in two stages: 1) Selection of basic aspects, and 2) Aggregation of basic aspects.

Selection of basic aspects. In the first stage, the designers select seventeen variables, which we name basic aspects and denote as:

$$b_1, b_2, \ldots, b_{17}.$$

These basic aspects describe various facts about the countries, which are or should be relevant for the construction of the HSC-index. The starting point for the selection of these basic aspects is the so-called working definitions stipulated at the Oslo Roundtable conference (1994). The

working definitions are stated in three areas, which function as broad guidelines for the selection of the basic aspects. But it would be misleading to regard the working definitions as a set of objective rules for the selection of relevant basic aspects. The selection of the basic aspect must reasonably be context sensitive. This is also noted by the designers claiming that:

> "It is therefore possible that the developed index is not suitable for other countries due to its bias towards the European lifestyle." (See Bartolj et al., 2018: 45).

The notion "European lifestyle" might be interpreted as referring to a broad normative background knowledge, which is relevant for constructing the HSC-index in a European context. In another context a different set of basic aspects might be relevant.

We denote the measures of the basic aspects as:

$$m_1, m_2, \dots, m_{17}.$$

It is not unreasonable to believe that some of these measures may be of the kind *M*-measures, that is, measures of descriptive phenomena, but it is likely that others are representations of some valuations and normative decision processes, that is, *R*-measures but not *M*-measures. In the next stage we comment on the aggregation process.

Aggregation of the basic aspects. The construction of the HSC-index is based on an aggregation of the measures regarding the basic aspects. Formally stated, the measure HSC is a function of the measures $m_1, m_2, \dots \dots, m_{17}$:

$$\text{HSC} = f(m_1, m_2, \dots, m_{17}).$$

The construction of f corresponds to an aggregation process, which is obviously an intricate and complicated normative decision process. The function f can be interpreted as an aggregation rule that determines relationships between the aggregated aspect 'sustainable

consumption', measured by HSC, and the basic aspects measured by $m_1, m_2, \dots, m_{17}$. The relationship between the aggregated concept *sustainable consumption* and its basic aspects is in the general case very complicated. We summarize some of these relationships in the Appendix, but a thorough specification of these relationships requires a more elaborated conceptual framework (see for example Odelstad, 2002; Bouyssou, Marchant, Pirlot, Tsoukias, & Vincke 2006) than can be introduced in this chapter.

As illustrated in Figure 1, the aggregation of the basic aspects takes place in two stages. In the first stage, two measures (representing "dimensions" or "factors" in Bartolj et al., here denoted HSC_1 and HSC_2) are constructed. HSC_1 is a measure of an intermediate aspect S_1 which is an aggregation of ten of the seventeen selected basic aspects. Formally, HSC_1 is a function of the measures $m_1, m_2, \dots, m_{10}$, i.e.,

$$\mathrm{HSC}_1 = f_1(m_1, m_2, \dots, m_{10}).$$

HSC_2 is a measure of an intermediate aspect S_2 which is an aggregation of the remaining seven of the seventeen selected basic aspects. Formally, HSC_2 is a function of the measures $m_{11}, m_{12}, \dots, m_{17}$, i.e.

$$\mathrm{HSC}_2 = f_2(m_{11}, m_{12}, \dots, m_{17}).$$

How should the measures be interpreted? It should be noted that the two measures represent *partial evaluations* of the countries regarding sustainable consumption. This means that the measures represent at least two partial rankings of the 28 EU countries, as follows:

c_i is (as regards $b_1, b_2, \dots, b_{10}$) at least as sustainable with respect to consumption as c_j if and only if

$$\mathrm{HSC}_1(c_i) \geq \mathrm{HSC}_1(c_j)$$

and

c_i is (as regards b_{11}, b_{12}, … , b_{17}) at least as sustainable with respect to consumption as c_j

if and only if

$\mathrm{HSC}_2(c_i) \geq \mathrm{HSC}_2(c_j)$.

In Figure 1, the two partial evaluations represented by HSC_1 and HSC_2 correspond to the arrows denoted S_1 and S_2, respectively (i.e., the two "dimensions" in Bartolj et al.). S_0 represents the aggregated evaluation, that is, the aspect 'sustainability with respect to household consumption'.

We have previously pointed out that the measures are determined by applying a statistical algorithm named factor analysis. It is essential to understand in this context that the application of a certain statistical method has normative consequences since sustainable consumption is an intermediate concept. The choice of using factor analysis should not only be grounded on various kinds of statistical principles and criteria, but also on normative reasoning regarding (for example) how an improvement in one basic aspect relates to a deterioration in another basic aspect, in terms of overall sustainability. We argue that rational normative reasoning presupposes both a correct interpretation of the function of 'sustainable consumption' as well as adequate domain knowledge. It is clear that the designers of the HSC-index have extensive domain knowledge, but we believe that they make a conceptual mistake that erodes the validity of the index.

In the next stage, these two measures are aggregated by means of a geometric aggregation rule defined as:

$$\mathrm{HSC}(x) = \sqrt{\mathrm{HSC}_1(x) \cdot \mathrm{HSC}_2(x)}$$

The overall rank-order of 28 EU countries is finally determined as follows:

c_i is at least as sustainable with respect to consumption as c_j

if and only if

$\mathrm{HSC}(c_i) \geq \mathrm{HSC}(c_j)$.

In Figure 1, the rank-order is denoted as the arrow S_0. It should be noted that there are many different aggregation rules that could be applied in this context. The question is—at least from a normative point of view—why the designers prefer this aggregation rule. For example, another popular aggregation rule is the additive aggregation rule, which in this case would be defined as:

$$\mathrm{HSC}(x) = w_1 \cdot \mathrm{HSC}_1(x) + w_2 \cdot \mathrm{HSC}_2(x),$$

where w_1 and w_2 are scaling constants coordinating the two aggregated measures. The ranking determined by this additive aggregation rule might of course not be consistent with the ranking determined by the geometric aggregation rule. The choice of the aggregation rule is obviously an intricate normative decision problem. But it should be noted that while additive aggregation requires that the measures to be aggregated are interval scales, geometric aggregation requires ratio scales, which is a higher demand.

For example, the choice between the two aggregation rules might imply inconsistent solutions of so-called value conflicts, illustrated in Figure 2. Let us assume that the construction of the two measures HSC_1 and HSC_2 give rise to the following outcome (where, for example, *F* and *P* denote France and Poland, respectively):

$\mathrm{HSC}_1(F) > \mathrm{HSC}_1(P)$, i.e. $\mathrm{HSC}_1(F) - \mathrm{HSC}_1(P) > 0$,

and

$\mathrm{HSC}_2(P) > \mathrm{HSC}_2(F)$, i.e. $\mathrm{HSC}_2(P) - \mathrm{HSC}_2(F) > 0$.

This gives rise to a value conflict, illustrated in Figure 2, since in this example France is more sustainable than Poland with respect to the basic aspects b_1, b_2, ... , b_{10}, while Poland is more sustainable than France with respect to b_{11}, b_{12}, ... , b_{17}.

[Figure 2 here]

Obviously, there are three possible outcomes of this value conflict:

1. $\mathrm{HSC}(F) > \mathrm{HSC}(P)$,
2. $\mathrm{HSC}(F) = \mathrm{HSC}(P)$,
3. $\mathrm{HSC}(P) > \mathrm{HSC}(F)$.

Now, let us assume that the additive aggregation implies that $\mathrm{HSC}(F) > \mathrm{HSC}(P)$, which is the outcome illustrated in Figure 2, whereas a geometric aggregation implies that $\mathrm{HSC}(P) > \mathrm{HSC}(F)$. The inconsistent solutions of the value conflict might in turn give rise to different policy recommendation regarding measures about the consumption patterns in the two countries. Policy makers that intend to use the index should be able to understand such crucial weighing or trade-off decisions. As we claimed earlier, a policy maker that takes decisions based on statements such as $\mathrm{HSC}(c_i) > \mathrm{HSC}(c_j)$, but does not understand what decisions (e.g., which tradeoffs) the statement is grounded on, might take irrational policy measures not consistent with the policy maker's own values.

To sum up: By means of the reconstruction of the HSC-index, we have identified and commented on three kinds of normative decisions taken by the designers.

1) Selection of seventeen basic aspects and assignment of measures to the basic aspects.
2) Construction of two partial measures grounded on the basic aspects by means of factor analysis.
3) Aggregation of the two partial measures by means of a geometric aggregation rule.

The construction of the index is based on an extensive number of decisions taken by the designers. A problem with the decision process is the lack of transparency. This means that an external judger as a policy maker (that intends to use the index as guidelines for policy measures) is not able to understand and assess the decisions that have been taken by the designers. The designers might have taken decisions that would not be consistent with the policy maker's values. For example, various trade-offs or weighing decisions taken more or less implicitly in the construction process might not be consistent with the policy maker's views of reasonable trade-offs. If that would be the case, the index has a low external validity, that is, a low validity from the policy maker's point of view. However, as the index is constructed it seems difficult or maybe impossible to assess the external validity of the HSC-index. It should be noted that the index might have a low external validity even if the index has a high degree of internal validity, that is, the decisions taken in the construction process are from the designers' point of view well-argued.

We end the chapter by emphasizing that to determine a ranking of objects as the 28 EU countries regarding sustainable consumption is a normative multi-attribute decision problem and not a statistical and empirical problem. We do not, however, argue that the use of statistical methods and geometric aggregation is necessarily a bad design choice, but the choice needs to be grounded on normative reasoning. Such reasoning could include the kind of trade-offs illustrated in Figure 2, for example supported by decision support tools based on a conceptual framework elaborated within multi-attribute decision theory. For an extensive overview and treatment, see for example, Belton & Stewart, 2002, the Handbook by Figueira, Greco & Ehrgott, 2005, and classical works like Keeney & Raiffa, 1976 and Keeney, 1993.

CONCLUSION

This chapter is directed towards researchers and practitioners within the field of sustainability, particularly those interested in the design of measures or indices of sustainability and those interested in the application and/or interpretation of such measures. Its relevance to, for example, policy makers lies in that it contributes to a deeper understanding of what kind of concept sustainability is and what it means to measure such concepts. Without a proper understanding of the nature of this concept, there is an obvious risk of low validity of its measurement. In the chapter, we perform a case study where we point out potential misinterpretations of the notion of sustainability in general and sustainable consumption in particular. The case study consists of a critical analysis of the construction of the HSC-index, a household sustainable consumption index for 28 EU countries.

A starting point for the analysis is our claim that it is a conceptual mistake (which we refer to as the descriptive or realistic mistake) to treat sustainable consumption as a descriptive concept. A consequence of this mistake is that the construction of the index is treated as a kind of statistical and empirical problem only. Instead, we argue that sustainable consumption is a specific kind of value concept named *intermediate concept*, whose function is to link descriptive grounds with normative consequences, and thus that the construction of the index is a normative problem. That is, the construction of a sustainable consumption index should be regarded as a normative multi-attribute decision process. We discuss what it means to measure multidimensional intermediate concepts, in light of the distinction between measurement-*R* (the wider notion of measurement as numerical representation) and measurement-*M* (the narrower notion of measurement as application of a decision method), and argue that measurement of sustainability is measurement in the former sense but not in the latter. That is, a measure of sustainable consumption is a numerical representation of normative decisions taken in the index design process.

Obviously, the conceptual mistake made by the designers hollows the validity of the constructed sustainable consumption index. Naturally, this also applies to similar sustainability measures or indices, and to some extent even to the guidelines in the OECD (2008) handbook "Constructing Composite Indicators" that the HSC-index designers refer to. The other main contribution of the chapter is the presentation of an alternative approach that aims to avoid this conceptual mistake, employing key concepts from multi-attribute decision theory (e.g. aggregation, component relations, utility difference comparisons) and the theory of intermediate concepts. We suggest a deeper and more formal conceptual analysis of the concept sustainable consumption, together with more general analyses of sustainability indices constructed by means of the guidelines in the OECD handbook. Further, we suggest that decision support tools should be developed and tested in the context of sustainable consumption as well as in other kinds of sustainability contexts. Such decision support tools can be constructed by means of the conceptual framework elaborated within multi-attribute decision theory. Using decision support tools, it is possible to explicitly treat the construction of sustainability indices as normative multi-attribute decision process.

[i] The Oxford Dictionary of Philosophy: Hume's law. A name for the contested view that it is impossible to derive an 'ought' from an 'is'. In other words: There is no logical bridge over the gap between fact and value. (Blackburn, 1994: 180.)

REFERENCES

Bartolj, T., Morovec, N., & Slabe-Erker, R. 2018. Development of a Household Sustainability Consumption Index and Its Application to EU-28. *Sustainable Development*, 26: 34–50.

Belton, V., & Stewart, Th. 2002. *Multiple Criteria Decision Analysis*: *An Integrated Approach.* Dordrecht: Kluwer Academic Publishers.

Blackburn, S. 1994. *The Oxford Dictionary of Philosophy*. Oxford: Oxford University Press.

Bouyssou, D., Marchant, Th., Pirlot, M., Tsoukias, A., & Vincke, Ph. 2006. *Evaluation and Decision Models with Multiple Criteria*. Dordrecht: Kluwer Academic Publishers.

Figueira, J., Greco, S., & Ehrgott, M. 2005. *Multiple Criteria Decision Analysis: State of the Art Surveys.* New York: Springer.

Hjelmblom, M., Paasch, J. M., Paulsson, J., Edlund, M., & Bökman, F. 2019. Towards Automation of the Swedish Property Formation Process: A Structural and Logical Analysis of Property Subdivision. *Nordic Journal of Surveying and Real Estate Research*, 14 (1): 29–63. doi:10.30672/njsr.78170

Keeney, R. L. 1993. *Value Focused Thinking: A Path to Creative Decision Making*. Cambridge, Massachusetts: Harvard University Press.

Keeney, R. L. & Raiffa, H. 1976. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. New York: Wiley.

Kirchin, S. 2013. *Thick Concepts*. Oxford: Oxford University Press.

Lindahl, L., & Odelstad, J. 2013. The Theory of Joining-Systems. In D. Gabbay, J. Horthy, X. Parent, R. van der Meyden, & L. van der Torre (Eds), *Handbook of Deontic Logic and Normative Systems*, vol. 1: 545–634. London: College Publications.

Merriam-Webster. (n.d.). Multidimensional. In *Merriam-Webster.com dictionary*. https://www.merriam-webster.com/dictionary/multidimensional. Accessed March 3, 2020.

Merriam-Webster. (n.d.). Phenomenon. In *Merriam-Webster.com dictionary*. https://www.merriam-webster.com/dictionary/phenomenon. Accessed March 3, 2020.

Möller, N. 2009. *Thick Concepts in Practice: Normative Aspects of Risk and Safety*. Unpublished PhD Thesis, Royal Institute of Technology, Stockholm.

OECD. 2008. *Handbook on Constructing Composite Indicators*: *Methodology and User Guide*. Paris: OECD Publications.

Odelstad, J. 2002. *Intresseavvägning*: *En beslutsfilosofisk studie med tillämpning på planering. (Weighing of Interests*: *A Study in the Philosophy of Decision Making with Applications to Planning.*) Stockholm: Thales.

Odelstad, J. 2017a. Likvärdigt arbete och teorin om mellanbegrepp. (Jobs of Equal Value and the Theory of Intermediate Concepts. Research report No. 46.). Sweden: Gävle University Press.

Odelstad, J. 2017b. Om beslutsteoretiska verktyg vid tillståndsprövning av vindkraft. (Decision Support Tools and Permission Processes for Wind Power. Research Report No. 47.). Sweden: Gävle University Press.

Odelstad, J. 2019. Joining Conceptual Systems: Three Remarks on TJS. *Filosofiska Notiser* (Philosophical Notes), 1: 77–131.

Sen, A. K. 1970. *Collective Choice and Social Welfare.* San Francisco: Holden-Day.

Sen, A. 1977. Social Choice Theory: A Re-Examination. *Econometrica,* 45(1): 53–88.

Väyrynen, P. 2013. *Thick Ethical Concepts.* Stanford: Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/entries/. Accessed March 2, 2020.

APPENDIX: DEFINITIONS OF SELECTED COMPONENT RELATIONS

Odelstad (2017a, referring to Sen, 1970) discusses a number of conditions that may apply to the relationship between an aggregated aspect (like sustainable consumption) and its basic aspects, and that can be of relevance for normative reasoning and argumentation. These 'component relations' include *equality preservation*[ii], *positive response*[iii], *non-negative response*, *global non-negative response*, and *in accordance with*, *ceteris paribus* (see for example, Odelstad, 2002, section 9; Odelstad, 2017a: 26f) and similar principles. Below we describe some of the component relations by means of a simple two-dimensional example, where m_1 and m_2 are measures of the underlying aspects b_1 and b_2 and m_0 is a measure of an aggregated aspect Σ_0.

Stated in terms ofm_1, m_2 and m_0 : Σ_0 is *equality-preserving* in C in relation to b_1 and b_2 if, for all c_i and c_j in C:

If $m_n(c_i) = m_n(c_j)$, for all n, $1 \leq n \leq 2$, then $m_0(c_i) = m_0(c_j)$.

Σ_0 exhibits *positive response (type 1)* in relation to the aspects b_1 and b_2 if, for all c_i and c_j in C:

If $m_n(c_i) > m_n(c_j)$ and $m_p(c_i) = m_p(c_j)$ for all p, $1 \leq p \leq 2$ and $p \neq n$, then $m_0(c_i) > m_0(c_j)$.

Σ_0 exhibits *global non-negative response* in relation to the aspects b_1 and b_2 if, for all c_i and c_j in C:

If $m_n(c_i) \geq m_n(c_j)$ for all n, $1 \leq n \leq 2$, then $m_0(c_i) \geq m_0(c_j)$.

Strong-positive response can be stated as:

$$\text{if } m_k(c_i) > m_k(c_j), \text{ then } m_0(c_i) > m_0(c_j).$$

To generalize these component relations to the case of an arbitrary number of basic aspects is in some sense straightforward but requires nevertheless a stronger algebraic language, which cannot be introduced in this chapter.

[ii] "Pareto-wise indifference" in the terminology of Sen (1970). (See Odelstad, 2017a: 26).

[iii] "Pareto-wise better" (Odelstad, 2017a: 26).

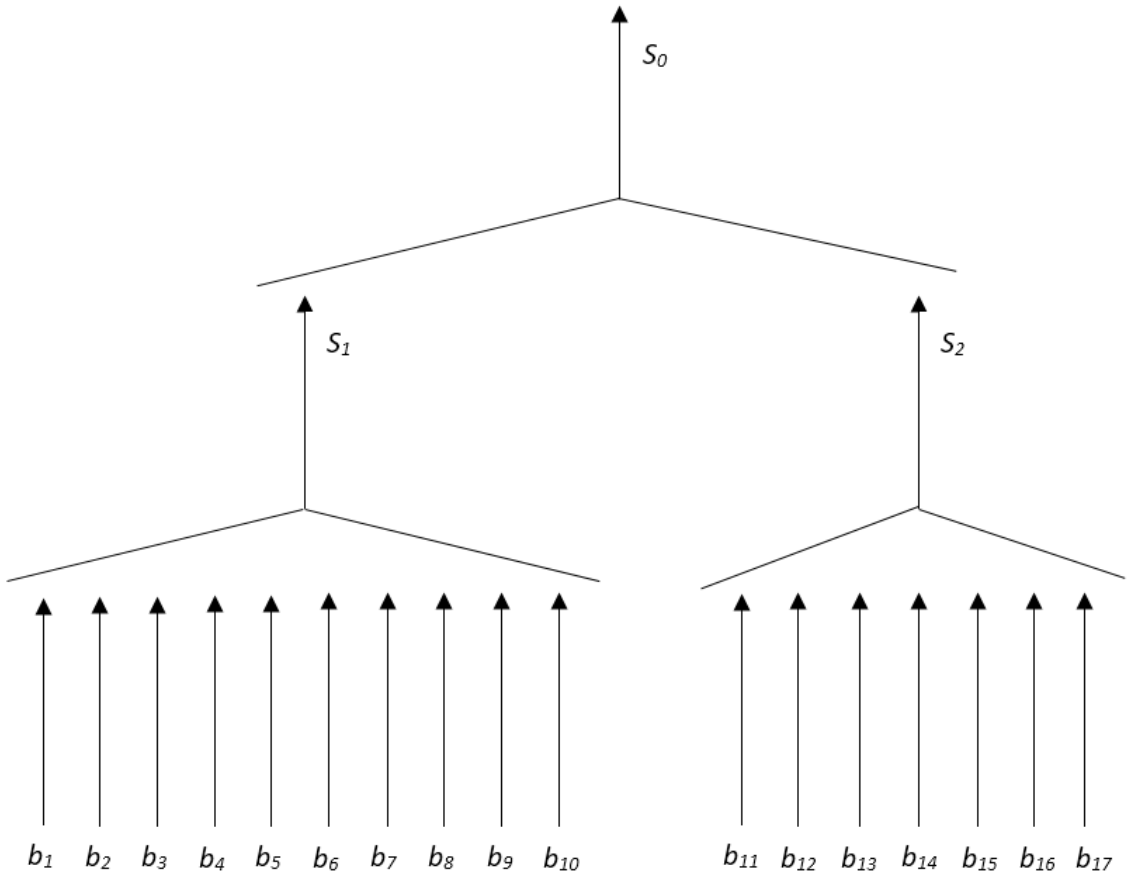

Figure 1. An example of an aggregation tree with three strata.

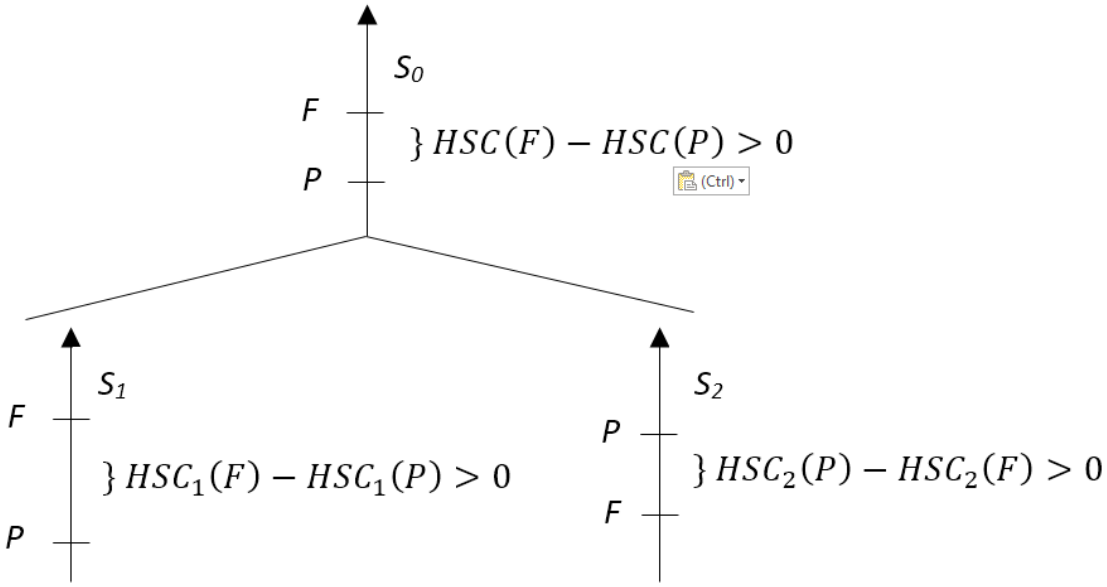

Figure 2. A value conflict. The difference between *F* and *P* in S_1 outweighs the difference between *P* and *F* in S_2.