KEMI
Report
Nr 8/07
A Report from the Swedish Chemicals Agency
www.kemi.se

The Substitution Principle

SWEDISH CHEMICALS AGENCY

The Substitution Principle

ISSN: 0284-1185
Order No. 360 878
Stockholm, November 2007
Publisher: Swedish Chemicals Agency©
Order address: Telefax +46 8 50 59 33 99, e-mail: kemi@cm.se

PREFACE

This report has been commissioned by the Swedish Chemicals Agency. Its purpose is to clarify what the substitution principle means and how it can reasonably be applied as part of policies based on the Swedish environmental objectives.

The project was performed as a literature study of the available international scientific literature in this area. This literature was identified through the Science Citation Index, Google-scholar, and Medline databases. Information from agency reports and various publications from other relevant actors, identified with the help from the Swedish Chemicals Agency and other contacts, has also been included.

The authors Sven Ove Hansson and Christina Rudén are both affiliated to the Royal Institute of Technology in Stockholm, Sweden. Hansson is professor in Philosophy. Hansson's scientific publications in the area of chemicals control are focused on regulatory issues including the interface between science and policy such as the burden of proof in toxicology and risk assessment, and decision-making under uncertainty. Rudén's research has focused on the chain of events leading from the generation of scientific data, via risk assessment, to risk management within the European chemicals legislation.

The report begins by defining the substitution principle (chapter 1). After that the authors "triangulate" it by showing how it relates to three other important principles in risk assessment and risk management, namely inherent safety (chapter 2), the precautionary principle (chapter 3) and risk analysis (chapter 4). This is followed by a chapter on comparative risk assessment, that is an essential component in any application of the substitution principle (chapter 5), and a chapter discussing practical measures that can be taken in order to increase the application of the substitution principle (chapter 6). Finally, the conclusions are summarized (chapter 7).

Swedish Chemicals Agency

CONTENTS

SUMMARY AND CONCLUSIONS

We propose to define the substitution principle as follows: *If risks to the environment and human health and safety can be reduced by replacing a chemical substance or product either by another substance or by some non-chemical technology, then this replacement should take place. All decisions on such substitutions should be based on the best available evidence. This evidence can be sufficient to warrant a substitution even if quantitative risk estimates cannot be made.* The substitution principle should not be understood as an absolute principle but as an efficient first-hand method to achieve the goals of chemicals risk management. The responsibility for achieving these goals, which is often best done by rigorously applying the substitution principle, rests primarily on the companies that produce and use chemical substances.

The substitution principle is an application of the principle of inherent safety, that has been developed in the chemical industry. Inherent safety, also called primary prevention, consists in the elimination of a hazard. It is contrasted with secondary prevention that consists in reducing the risk associated with a hazard. Proponents of inherent safety have shown that other things being equal, elimination of a hazard is a better option than to retain it and construct safeguards against it. The major reason for this is that as long as the hazard still exists, it can be realized by some unanticipated triggering event. Even with the best of control measures, some unforeseen chain of events can give rise to an unforeseen exposure. Even the best add-on safety technology can fail, or be destroyed in the course of an accident. This is of course an excellent argument for replacing a dangerous substance by a less dangerous one, even if it is believed that the hazard can be coped with.

Risk management decisions always have to be based on the available information. Waiting until we have all the information we want means to let potential damage occur in the meantime. The need to make decisions on incomplete information also includes decisions on substitution. Thus, if we have good reasons to believe that a substance is more dangerous than its alternatives, then we should avoid using that substance, even if the good reasons do not amount to full scientific proof.

To make a scientifically sound risk assessment, that enables a reasonably robust comparison of alternatives, an extensive amount of exposure and toxicity data are needed. For some groups of chemicals a set of such data are generated before the substance can be put on the market, e.g. for many new industrial chemicals, and for pesticides. However for the majority of general/industrial chemicals publicly available data are scarce. The most important improvement towards an efficient implementation of the substitution principle for general chemicals is thus to improve data availability. The new chemicals legislation, REACH, can be seen as a first, but insufficient, step towards this aim.

The uncertainties about the costs and benefits in the substitution decision must be considered case-by-case. In some cases the uncertainties may be so large that the potential benefits of a substitution are difficult to estimate. In other cases the benefits may be obvious. When uncertainties are large, substitution may not be the best option and alternative risk management strategies should be considered. The uncertain cases should however not be used as an argument against substitution decisions in the cases where the uncertainties are small.

In risk assessments of industrial chemicals, the prevailing praxis is still to treat a substance with unknown properties in the same way as a substance known to be harmless. This is irrational from the viewpoint of standard decision theory, and this practice should therefore be replaced by one that is not as extremely risk-taking. We propose that expected utility theory, that is generally recognized as the “risk-neutral” decision method, should be applied to this

problem. This means that a substance with unknown toxicity should be treated by risk managers in the same way as a substance with “average” toxicity. We can call this the method of *risk-neutral default*. It should be applied equally to substances in use and substitution candidates not yet in use.

Experience shows that the market will usually not automatically ensure that substitution takes place. For the substitution principle to be efficiently implemented, regulators and public authorities have to take the lead. However, it is impossible for them to decide on each particular substitution. Public decisions on specific substitutions have to be reserved for special cases. For the vast majority of cases, the role of regulators and authorities should instead be to create incentives for substitution. Many different method can be used for that purpose, including increased availability of toxicity data, increase availability of chemical composition data, increased the availability of information about technical functionality, technical help to enterprises, lists of unwanted substances, mandatory substitution plans and various economic incentives. It is important to conduct substitution work in such manner that it is accessible to outcome evaluation. By systematically evaluating the effects of various substitution-promoting measures undertaken by public agencies, we can learn which methods are the most effective ones.

In order to apply the substitution principle, knowledge and competences are needed that the chemical industry has to a much higher degree than most of their customers. The toxicological, chemical, and industrial considerations that are involved in successful applications of the principle can therefore to a large extent be provided as services by a competent supplier of chemicals to its customers. A chemical company can offer its customers less hazardous products for the chosen function or purpose, and it can in many cases offer to deliver a product in a less hazardous form. There does not seem to be any conflict between the substitution principle and the business interests of a chemical company that possesses the core competences of a supplier of chemical products.

SAMMANFATTNING

Vi föreslår följande definition av substitutionsprincipen: *Om miljö och hälsorisker kan minskas genom att en kemikalie eller en kemisk produkt byts ut mot en annan eller byts mot en teknologi som inte innefattar kemikalieanvändning, så skall sådan substitution genomföras. Alla substitutionsbeslut ska baseras på bästa tillgängliga kunskap. Beslut om substitution kan tas även om riskens storlek ej kan kvantifieras.* Substitutionsprincipen ska inte förstås som en absolut regel utan snarare som ett alternativ som det är effektivt att tillämpa i första hand i strävan mot målet om en säker kemikaliehantering. Ansvaret för att uppnå detta mål ligger i huvudsak hos de företag som producerar eller använder kemikalier.

Substitutionsprincipen kan ses som en tillämpning av principen om inneboende säkerhet (”inherent safety”), en princip som utvecklats inom den kemiska industrin. Inneboende säkerhet, eller primär prevention (”primary prevention”) som principen också kallas, fokuserar på att eliminera fara. Principen kan jämföras med sekundär prevention (”secondary prevention”) som istället går ut på att hantera de risker som en fara ger upphov till. Företrädare för principen om inneboende säkerhet har visat att, givet att alla andra faktorer är konstanta, är det mer effektivt att undvika en fara jämfört med att behålla faran och vidta säkerhetsåtgärder för att kontrollera de risker som den ger upphov till. Det huvudsakliga argumentet för att fokusera på primär prevention är att så länge som faran existerar så kan den realiseras genom oförutsedda händelser. Även de bästa säkerhetssystem kan fallera eller förstöras genom en olycka. Detta är alltså ett huvudargument för att byta en farlig kemikalie mot en mindre farlig, även om det *a priori* förefaller troligt att riskerna kan hanteras.

Beslut om riskminskning måste alltid baseras på den information som finns tillgänglig vid tiden för beslutet. Att skjuta upp ett beslut till dess att all relevant kunskap har genererats innebär att den potentiellt farliga verksamheten tillåts fortgå under tiden. Detta gäller även beslut om substitution. Om vi har goda skäl att tro att en kemikalie är farligare än dess alternativ, så bör användning av alternativet allvarligt övervägas även om dess fördelar för tillfället inte har bevisats vetenskapligt.

För att kunna göra en vetenskapligt försvarbar och robust jämförelse av riskerna med olika alternativa kemikalier krävs en stor mängd data om kemikaliernas egenskaper och användning inklusive vilken exponering användningen ger upphov till. För kemikalier som produceras för vissa användningsområden genereras sådan kunskap genom lagstadgade krav. Det gäller till exempel kemikalier som används som bekämpningsmedel och läkemedel. För många industrikemikalier är dock kunskaperna mycket begränsade. Den enskilt viktigaste åtgärden för att främja ett effektiv tillämpning av substitutionsprincipen för industrikemikalier är därför att förbättra kunskapsläget för dessa kemikalier. Den nya europeiska kemikalielagstiftningen, REACH kan ses som ett första steg i den riktningen.

Osäkerheter om kostnader och vinster med att fatta ett substitutionsbeslut måste hanteras från fall till fall. Ibland kommer dessa osäkerheter att vara betydande och leda till att de potentiella fördelarna med en substitution inte kan bedömas på ett meningsfullt sätt. I andra fall kommer det att vara uppenbart att fördelarna (eller nackdelarna) med en substitution överväger. När osäkerheterna är stora bör även andra metoder för riskhantering tas i beaktande. Det bör dock poängteras att osäkerheter i dataunderlaget i enskilda fall inte kan användas som ett generellt argument mot substitutionsprincipen som sådan.

Inom det regulatoriska ramverket för industrikemikalier är det fortfarande allmän praxis att behandla en kemikalie med okända egenskaper på samma sätt som man behandlar en kemikalie som har undersökts och befunnits vara relativt ofarlig. Denna praxis är extremt risktagande och därigenom irrationell ur en beslutsteoretisk synvinkel. Vi föreslår att vid

substitutionsbeslut bör denna praxis ersättas av den generella beslutsteoretiska princip som förordar maximerande av förväntad nytta, en beslutsprincip som allmänt anses vara riskneutral. En systematisk tillämpning av en sådan beslutsregel skulle innebära att kemikalier med okänd toxicitet skulle hanteras som om de var lika farliga som en kemikalie med en ”genomsnittlig” toxicitet. I avsaknad av detaljerad kunskap om en specifik egenskap antar man alltså att kemikalien har en egenskap som motsvaras av genomsnittskemikalien i en relevant grupp av ämnen. Denna princip ska tillämpas både för den substans som eventuellt ska bytas ut och för den kemikalie som övervägs som ersättning.

Erfarenheten visar att marknadsmekanismer är otillräckliga för att säkerställa en konsekvent tillämpning av substitutionsprincipen. För att åstadkommas detta krävs politiska initiativ och aktiva myndigheter. Att myndigheter fattar beslut om substitution i enskilda fall bör dock endast utgöra undantagsfall. Politiska aktörer och myndigheter bör istället verka på en övergripande nivå med att skapa ramverk och incitament som främjar substitutionsprincipens tillämpning. Det finns flera faktorer som kan bidra till att skapa sådana incitament. Exempel på sådana faktorer är: ökad tillgång på relevant information om kemikaliernas egenskaper och användning, ökad tillgång på information om innehållet i kemiska produkter, ökad tillgång på information om kemikaliers tekniska egenskaper, teknisk support till kemikalieanvändare, listor på kemikalier som har särskilt farliga egenskaper eller vars användning är förenad med betydande risker, krav på att företag ska upprätta substitutionsplaner, samt olika former av ekonomiska incitament. Åtgärder för att främja substitution bör alltid genomföras på ett sätt så att resultaten av åtgärderna kan utvärderas. Genom en systematisk utvärdering av sådana åtgärder kan kunskaper genereras om vilka metoder som är de mest effektiva.

För att kunna tillämpa substitutionsprincipen krävs kunskap och expertis som kemikalietillverkande industrier har tillgång till i betydligt högre utsträckning än de flesta av deras kunder. De toxikologiska, kemiska, och tekniska överväganden som behöver göras för att åstadkomma en framgångsrik substitution borde därför kunna erbjudas av den kemikalietillverkande industrin som en service till dess kunder. På så vis kan de kemikalieförsäljande företagen bidra till implementeringen av den säkraste lösningen för varje teknisk process. Det förefaller alltså inte finnas någon konflikt mellan en systematisk tillämpning av substitutionsprincipen och företagsekonomi för de kemikalieförsäljare som kan tillhandahålla relevant kunskap och expertis.

1 DEFINING THE SUBSTITUTION PRINCIPLE

To substitute (latin: substituere) means to replace, or to "put (one) in place of another" (OED). In chemicals regulation and risk management, by the *substitution principle* is meant a policy principle that requires the replacement of hazardous (or potentially hazardous) chemical substances by less hazardous alternatives. A less common synonym is the *product choice principle*. In what follows we will use the former terminology. We will use the term *original substance* for the chemical to be replaced *substitute* for that which replaces it.

The substitution principle has been discussed in different regulatory and risk management contexts. It has also been the subject of heated debates. Not surprisingly, it has been interpreted differently by various groups of appliers, proponents and opponents.[1] In order to make clear more precisely what we mean by the substitution principle, at least the following five aspects have to be specified.

- the purpose of substitution
- whether the substitution is chemical or functional
- the required functionality of the substitute
- whether substances or products are substituted
- whether substitution is based on hazard or risk

1.1 The purpose of substitution

There can be many motives for substituting one chemical by another. Chemicals are replaced for various economic and technical reasons. In most of the literature on the substitution principle, the motive of substitution is taken to be that the substance is "hazardous", without further specification. This is, in our view, one of the aspects of the principle that should be clarified.

Since the substitution principle is part of environmental policies, the substitutions that it refers to have motives related to the environment and to human health. We find in reasonable to include health-related motives that refer to other policy areas than environmental policies. Therefore, substitutions that aim at improving workplace health and safety should be included. This means that not only the toxic properties of substances should be included but also properties such as explosiveness and inflammability. This wide inclusion of motives seems to be well in accord with common practice, although this delimitation has not been much commented upon.

With such a wide definition, it is only to be expected that motives for substitution can sometimes come into conflict with each other. Hence, substance A may be preferable to substance B in terms of toxicity, whereas substance B is preferable to substance A in terms of inflammability. A substitution that takes only one of these factors into account can be a failure. Therefore it is important to emphasize that the substitution principle refers to all the

[1] As an example of a flagrant misunderstanding, we can refer to the Royal Society of Chemistry that said in 2001: "For example in some Nordic countries it [the substitution principle] is taken to mean that a chemical should not be used at all if a 'less hazardous' substance exists." (Royal Society of Chemistry, *Comments from the Royal Society of Chemistry on the scoping of the Chemicals Study*, 26 January 2001, www.rcep.org.uk/chemicals/RSC.htm) – Needless to say, no Nordic country has such a legislation.

dangers to human health and the environment that may be associated with the use of a chemical.

1.2 Chemical versus functional substitution

Some definitions of the substitution principle specify it as requiring the replacement of a (hazardous) chemical substance by another (less hazardous) chemical substance. Although Agenda 21 does not explicitly refer to the substitution principle, the following formulation can be read as a support of substitution of hazardous chemicals by less hazardous ones:

> "Adopt policies and regulatory and non-regulatory measures to identify, and minimize exposure to, toxic chemicals by replacing them with less toxic substitutes and ultimately phasing out the chemicals that pose unreasonable and otherwise unmanageable risk to human health and the environment and those that are toxic, persistent and bio-accumulative and whose use cannot be adequately controlled".[2]

Perhaps the clearest restriction of the substitution principle to replacement of chemicals by other chemicals can be found in a text issued by CEFIC.

> "Substitution is the replacement of one substance by another with the aim of achieving a lower level of risk."[3]

(Like many others who have written on the topic, CEFIC define "substitution" rather than "the substitution principle". This is unfortunate, since the term "substitution" should have a wider range of application. It should be possible, for instance, to talk about substitutions that do not comply with the substitution principle. In what follows we will assume that those who define "substitution" in the context of the substitution principle use this word as an abbreviation of "substitution that complies with the substitution principle".)

However, several authors have pointed out that the principle should not be restricted to replacement by another substance. From the viewpoint of environmental protection, it should be clear that "the hazardous substance does not necessarily have to be replaced by another substance. It can also be substituted by other means of fulfilling the function it had. Thus, a hazardous cleaning agent (e.g. a chlorinated solvent) can be replaced by a less harmful one, but [it] is also conceivable that the product or production process is redesigned in such a way that the cleaning step can be omitted."[4] This is a definitional issue with considerable practical importance. A definition that focuses on finding another chemical can limit the search process so that non-chemical options are not investigated. Hence, as pointed out by Jennifer Hall, it makes a difference if one says "We need a substance to replace BFR" or "We need an environmentally better way to protect a material from catching fire".[5] Most definitions of the substitution principle take this into account, and allow for substitution by non-chemical means:

> "Informed substitution is the considered transition from a chemical of particular concern to safer chemicals or non-chemical alternatives."[6]

---

[2] Agenda 21, §19.49, http://www.un.org/esa/sustdev/documents/agenda21/english/agenda21chapter19.htm.
[3] *CEFIC paper on substitution and authorisation under REACH*, 23 May 2005, p. 1.
[4] Frans Oosterhuis, *Substitution of hazardous substances. A case study in the framework of the project 'Assessing innovation dynamics induced by environmental policy'*. Institute for Environmental Studies, Vrije Universiteit, Amsterdam 2006, p 3.
[5] Jennifer C Hall, *Product Design to Reduce Restricted Substances*, IIIEE Reports 2001:2, Lund, p. 18.
[6] Charles Auer, "U.S. Experience in Applying 'Informed Substitution' as a Component in Risk Reduction and Alternatives Analyses" transcript of an oral presentation given at the Chemicals, Health, and the Environment Conference Ottawa, Ontario, Canada, October 2006.

> "Substitution means the replacement or reduction of hazardous substances in products and processes by less hazardous or non-hazardous substances, or by achieving an equivalent functionality via technological or organisational measures."[7]

> "Substitution of a hazardous substance or product signifies its replacement by less a hazardous substance, product or process."[8]

In our view, it is obvious given the purpose of the substitution principle, that it should be interpreted as promoting the substitution of the use of a hazardous chemical by some (chemical or non-chemical) method that reduces potential for damage to health or the environment. As one example of this, brominated flame retardants in plastic casings for electronic equipment can be replaced by other, less hazardous flame retardants for the same casings. Alternatively, the plastic casings can be replaced by metal casings. [9] The substitution principle should not rule out such, larger technological changes that can solve an environmental problem. In other words, the substitution principle should be understood as referring to *functional* replacement, not chemical replacement.

1.3 Degrees of functionality

Some hazardous substances are difficult to replace because it is difficult to find other materials or methods that are equally functional. Hence, the most efficient pesticide for a certain purpose may have unacceptable effects on the environment or on the health of exposed persons. In such a case it may be necessary to replace it by a (chemical or non-chemical) method that is less efficient. Such a replacement would presumably be seen as an application of the substitution principle. This means that in the application of the principle, the intended function of the original chemical cannot always have absolute priority over protection of health and the environment.

However, a recent influential definition of the substitution principle requires that the substitute satisfies the functional requirements to the same degree as the original chemical:

> "Substitution means the replacement or reduction of hazardous substances in products and processes by less hazardous or non-hazardous substances, or by achieving an equivalent functionality via technological or organisational measures."[10]

According to Lohse et al, who authored this definition, a key element is "functional equivalence, i.e. the achievement of the same functionality by less hazardous means".[11] However, as we see it, the substitution principle has to be interpreted as being part of

---

[7] Joachim Lohse et al, *Substitution of Hazardous Chemicals in Products and Processes, Final Report*. Report compiled for the Directorate General Environment, Nuclear Safety and Civil Protection of the Commission of the European Communities, Hamburg 2003.

[8] Andreas Ahrens et al, *Hazardous Chemicals in Products and Processes. Substitution as an Innovative Process*, Physica Verlag 2006, p. 22.

[9] Beverly Thorpe and Mark Rossi, *The Louisville Charter: Background paper for reform no. 1 of the Louisville charter for safer chemicals*. August 2005, p. 3.

[10] Joachim Lohse et al, *Substitution of Hazardous Chemicals in Products and Processes, Final Report*. Report compiled for the Directorate General Environment, Nuclear Safety and Civil Protection of the Commission of the European Communities, Hamburg 2003.

[11] Joachim Lohse et al, *Substitution of Hazardous Chemicals in Products and Processes, Final Report*. Report compiled for the Directorate General Environment, Nuclear Safety and Civil Protection of the Commission of the European Communities, Hamburg 2003. – Cf: "Functional equivalence is a key element: if the replacement of the chemical leads to lower product quality or to unsurmountable problems with the process, one cannot speak of a (successful) substitution." (Frans Oosterhuis, *Substitution of hazardous substances. A case study in the framework of the project 'Assessing innovation dynamics induced by environmental policy*'. Institute for Environmental Studies, Vrije Universiteit, Amsterdam 2006, p 1.)

environmental policies that do not always give environmental objectives lower priority than the objectives of material production. Therefore, this restriction of the substitution principle must be rejected. Instead, the substitution principle should be seen as a principle that requires the replacement of dangerous substances by less dangerous alternatives, while attempting to (i) reduce the danger as much as possible, (ii) retain (or in principle increase) the functionality that the original substance had as much as possible, and (iii) keep costs as low as possible. The priority between these three objectives is a matter for negotiation and adjustment in each particular case, and cannot be settled beforehand by requiring that one of these objectives always has absolute priority over the others.

1.4 Substance or product

Most of the definitions quoted above refer to the substitution of substances; one of them refers instead to the "substitution of a hazardous substance or product".[12] In practice, substitution can refer either to chemical substances or to chemical products (i.e. commercial mixtures). Good arguments can be given why an efficient substitution policy should operate both on substances and products.

Beginning with substances, most of the known negative effects of chemicals on human health and the environment emanate from a particular substance. For substitution to be efficient in such cases, it must refer to the substance in question, not only to a particular product that contains it. As an example of this, an active substance in pesticides can be subject to substitution.

However, there are cases when a product is not acceptable although none of its components is unacceptable as such. The proportions between ingredients can for instance be suboptimal, so that the product contains too much of a particular substance. In such cases, substitution of the product should be an option.

1.5 Substitution according to hazard or risk

In chemicals regulation, much emphasis is put on the distinction between a hazard and a risk. A hazard can be defined as a potential risk. As applied to chemicals, the hazard is a property of the substance as such, whereas the risk depends on how the substance is used and handled. Hazard is mostly treated as a non-quantitative concept, whereas risk is often quantified in terms of probabilities.

Public authorities have often emphasized the application of the substitution principle to chemical hazards. Substitution based on hazard can be general, i.e. apply to many or all uses of a substance, and does not require detailed assessments of the particular circumstances under which the substance is used in individual companies. In contrast, representatives of the chemical industry have claimed that substitution should be performed on the level of risk, not hazard. The following quotes are from a report by CEFIC:

> "All chemical management decisions should be based on risk."[13]
>
> "Substitution is, however, not necessarily a simple process since it is necessary to ensure that the overall risk is reduced and that a decrease in one risk is not overshadowed by the increase in another. Although it sounds simple to identify a

[12] Andreas Ahrens et al, *Hazardous Chemicals in Products and Processes. Substitution as an Innovative Process*, Physica Verlag 2006, p. 22.
[13] *CEFIC paper on substitution and authorisation under REACH*, 23 May 2005, p. 1.

material's 'hazard' and to then replace it with a substance of lesser hazard, this is too simplistic."[14]

"Substitution is only acceptable provided the suitability of the alternative has been demonstrated through socio-economic costs/benefits analysis and a risk assessment."[15]

(The last of these three quotes introduces a strange asymmetry between the first substance that is used for a particular purpose and its successors. No socio-economic analysis is required for the introduction of the first substance, and presumably CEFIC does not propose its introduction. The one-sided application of such a requirement to a potential substitute, but not to the substance that it is intended to replace, does not seem to be conducive to innovation.)

Good reasons can be given both for applying the substitution principle to risks and for applying it to hazards. Basically, all decisions on substitution should be based on the best possible evidence. Therefore, if we know both the hazard (inherent potential to damage) and the exposure conditions that are associated with the alternatives under consideration, it would be irrational not to make use of all this information when deciding on potential substitutions. Under the conventional assumption that the risk can be calculated from the hazard and the exposure, this can be expressed as follows: When we know the risk, then substitution decisions should be based on the risk. (However, this definition of "knowing the risk" is too simplistic, as we will discuss in chapter 4.)

On the other hand, there are occasions when a substitution based on hazard is equally sensible. Consider two substances A and B that are alternatives for being used as degreasers (or for some other well-defined purpose). They can be used in the same way, and since they are equally technically efficient the same volume will be used irrespective of which substance we use. However, A dominates the market, and very few companies use B. On the other hand, a hazard analysis shows that the hazard associated with B is much smaller than that associated with A. We do not have access to the detailed quantitative information that would be required to perform a risk analysis (in the conventional sense). However, the information that we have is sufficient to show that a substitution of A by B would reduce the risks. Since no risk analysis is available, such a substitution would be counted as hazard-based rather than risk-based. As this example shows, hazard-based substitutions can contribute to reducing risks.

More generally speaking, applications of the substitution principle should be based on the best available evidence. This evidence can be sufficient to warrant a substitution even if quantitative risk estimates cannot be made.

1.6 Summary

In summary, we propose to define the substitution principle as follows:

If risks to the environment and human health and safety can be reduced by replacing a chemical substance or product either by another substance or by some non-chemical technology, then this replacement should take place. All decisions on such substitutions should be based on the best available evidence. This evidence can be sufficient to warrant a substitution even if quantitative risk estimates cannot be made.

The substitution principle should not be understood as an absolute principle but as an efficient first-hand method to achieve the goals of chemicals risk management. The responsibility for achieving these goals, which is often best done by rigorously applying the substitution principle, rests primarily on the companies that produce and use chemical substances.

---

[14] *CEFIC paper on substitution and authorisation under REACH*, 23 May 2005, p. 2.
[15] *CEFIC paper on substitution and authorisation under REACH*, 23 May 2005, p. 3.

2 SUBSTITUTION AND INHERENT SAFETY

In many situations, we have a choice between different ways to deal with a chemical risk. We can for instance replace the dangerous chemical, or we can retain it but reduce the exposure. The substitution principle can be described as *a strategy in safety engineering*, namely the strategy that favours the former of these alternatives. Seen in this way, the substitution principle is an instance or a special case of the principle of inherent safety, that has a strong tradition in the chemical industry's management of its own plants.

2.1 The principles of inherent safety

Inherent safety, also called primary prevention, consists in the elimination of a hazard. It is contrasted with secondary prevention that consists in reducing the risk associated with a hazard.[16] Secondary prevention can consist in reducing either the probability or the consequences of an adverse event, such as an accident in which the hazard is realized. For an example, consider a process in which inflammable materials are used. Inherent safety would consist in replacing them by non-inflammable materials. Secondary prevention would consist in removing or isolating sources of ignition and/or installing fire-extinguishing equipment. As this example shows, secondary prevention usually involves the use of add-on safety equipment.

Traditionally, four types of safety measures are recommended in inherently safer design of plants:

> *minimize* (*intensify*): use smaller quantities of hazardous materials
>
> *substitute*: replace a hazardous material by a less hazardous one
>
> *attenuate* (*moderate*): use the hazardous material in a less hazardous form
>
> *simplify*: avoid unnecessary complexity in facilities and processes, in order to make operating errors less likely.[17]

Although inherent safety is the most common term, some authors prefer "inherently safer design". The motivation for this is that (full) safety can never be achieved. An inherently safe plant does not exist. It is therefore more constructive to think in terms of inherently safer processes.[18] However, as others have pointed out, although there is no technology that is safe,

---

[16] It is also common to distinguish between four types of risk reduction strategies: inherent safety, passive (engineered) safety, active (engineered) safety, and procedural safety. The last three of these can all be subsumed under the concept of secondary prevention.

[17] Faisal I Khan and SA Abbasi, "Inherently safer design based on rapid risk analysis", *Journal of Loss Prevention in the Process Industries* 11:361-372, 1998. RE Bollinger, DG Clark, AM Dowell III, RM Ewbank, DC Hendershot, WK Lutz, SI Meszaros, DE Park and ED Wixom, *Inherently Safer Chemical Processes – A Life Cycle Approach*, Center for Chemical Process Safety of the American Institute of Chemical Engineers, New York 1996. – Another way to systematize inherent safety was provided on p 364 in Faisal I Khan and SA Abbasi, "Inherently safer design based on rapid risk analysis", *Journal of Loss Prevention in the Process Industries* 11:361-372, 1998.
"1. intensification–using less of a hazardous material;
2. attenuation – using a hazardous material in a less hazardous form;
3. substitution – using a safer material;
4. limitation – minimizing the effect of an accident;
5. simplification – reducing the opportunities for error and malfunction."

[18] Faisal I Khan and Paul R Amyotte, "Integrated Inherent Safety Index (I2SI): A Tool for Inherent Safety Evaluation", *Process Safety Process* 23:136-148, 2004.

some technologies are practically speaking absolutely safe along certain dimensions. In particular, some chemicals are not flammable, or not explosive.[19]

2.2 Inherent safety and the basic principles of safety engineering

Since the 19th century, engineers have specialized in worker's safety and other safety-related tasks. With the development of technological science, the ideas behind safety engineering have been subject to academic treatments. There are now many ways to systematize the practices of safety engineering. However, none of them has gained general acceptance. A major reason for this is that the discussion of safety engineering is rather fragmented between discussions related to different areas of technology. A large number of safety principles have been proposed. Many of them overlap, or are just applications of the same basic ideas in different areas of engineering. In addition to inherent safety, the following three principles can be mentioned as particularly important in several areas of engineering safety:

Safety factors. In engineering design, constructions are usually made strong enough to resist loads and disturbances exceeding those that are intended. A common way to obtain such safety reserves is to employ explicitly chosen, numerical safety factors. Hence, if a safety factor of 2 is employed when building a bridge, then the bridge is calculated to resist twice the maximal load to which it will in practice be exposed. Safety factors have been used in structural mechanics since the 1860s. In regulatory toxicology, safety factors have been used since the 1940s. Here, a safety factor is defined as the ratio between an experimentally determined dose and a dose that is accepted in humans in a particular regulatory context.[20]

Negative feedback. Negative feedback mechanisms are introduced to achieve a self-shutdown in case of device failure or when the operator loses control. Two classical examples are the safety-valve that lets out steam when the pressure becomes too high in a steam-boiler and the dead man's handle that stops the train when the driver falls asleep.

Multiple independent safety barriers. Safety barriers are arranged in chains. The aim is to make each barrier independent of its predecessors so that if the first fails, then the second is still intact, etc. Typically the first barriers are measures to prevent an accident, after which follow barriers that limit the consequences of an accident, and finally rescue services as the last resort. One of the major lessons from the Titanic disaster is that an improvement of the early barriers (in this case: a hull divided into watertight compartments) is no excuse for reducing the later barriers (in this case: lifeboats).

The major principles of safety engineering all have one important trait in common: they aim at protecting us not only against risks (in the technical sense) but also against hazards that cannot be assigned meaningful probability estimates, such as the possibility that some unforeseen event triggers a hazard that is seemingly under control.

Inherent safety is one of the safety principles with general applicability that are based on this type of thinking about uncertainty. Proponents of inherent safety have shown that other things being equal, elimination of a hazard is a better option than to retain it and construct safeguards against it. The major reason for this is that as long as the hazard still exists, it can be realized by some unanticipated triggering event. Even with the best of control measures, some unforeseen chain of events can give rise to an unforeseen exposure. Even the best add-on safety technology can fail, or be destroyed in the course of an accident.

---

[19] Nicholas Askounes Ashford and Gerard Zwetsloot, "Encouraging inherently safer production in European firms: a report from the field", *Journal of Hazardous Materials* 78:123-144, 2000.
[20] Jonas Clausen, Sven Ove Hansson and Fred Nilsson, "Generalizing the Safety Factor Approach", *Reliability Engineering and System Safety* 91:964-973, 2006.

2.3 Inherent safety and the chemical industry

Although safety measures that we would today call inherent safety have a long history, explicit discussions of inherent safety have their background in a tragic event a third of a century ago. In June 1974 a chemical plant in Flixborough, exploded in an accident that killed 28 persons and seriously injured 36. This disaster gave rise to extensive discussions on how safety could be improved in the chemical industry. In these discussions Trevor Kletz, a chemist employed by Imperial Chemical Industries (ICI), put focus on several factors that had aggravated the accident. Perhaps most notably, large quantities of inflammable chemicals had been stored close to occupied buildings. The accident would not have reached its catastrophic proportions if simple measures had been taken to reduce hazards. Based on these tragic experiences, Kletz proposed that whenever possible, the chemical industry should eliminate hazards rather than just strive to manage them. He originally used the term "intrinsic safety" for this concept, but it was soon replaced by "inherent safety".[21]

Since the 1970s, inherent safety has been further developed and operationalized by major chemical companies. Employees of Rohm and Haas, Exxon Chemicals, Union Carbide, Sandoz, Dow and others have contributed to the development of methodology for inherently safer design.

Substitution is indeed a key component in the chemical industry's own work to achieve inherent safety and reduce hazards in their own plants. Hence, in an excellent paper, Tim Overton and George M King of Dow Chemical Company exemplify how their company has used "substitution, a strategy to replace one material with a less hazardous substance if the chemistry of the process will allow." Dow has for instance replaced benzene with a less hazardous substance, it has replaced a highly inflammable solvent with another, less hazardous solvent, and it has replaced chlorine gas with sodium hypochlorite as a water purification chemical.[22] Dennis Hendershot of Rohm and Haas Company mentions an important example: solvent-based paints can for many purpose be substituted by aqueous latex paints that are inherently safer. (However, as he also points out, the substitution is not universal since there are "applications where the increased performance of solvent based paints justifies their use, with the appropriate layers of protection.")[23]

It has been important for the development of inherently safer design that it is often economically preferable to other approaches, in particular if one looks beyond the original investment. Hence, inherently safer constructions are often associated with greater reliability of production, which can be crucial for production economy. Add-on safety is maintenance intensive and therefore often costly in the long run.[24] As was noted by Tim Overton and George M King of Dow, inherently safer technology "can result in lower capital cost in new plant design, and typically produces lower operating costs, greater reliability, and quicker

---

[21] T Kletz, "What You Don't Have, Can't Leak", *Chemistry and Industry*, May 1978, 287-292.
[22] Tim Overton and George M King, "Inherently Safer Technology: An Evolutionary Approach", *Process Safety Progress* 25:116-119, 2006.
[23] Dennis C. Hendershot, "Inherently safer chemical process design", *Journal of Loss Prevention in the Process Industries* 10:151-157, 1997.
[24] Nicholas Askounes Ashford and Gerard Zwetsloot, "Encouraging inherently safer production in European firms: a report from the field", Journal of Hazardous Materials 78:123-144, 2000. Faisal I Khan and SA Abbasi, "Inherently safer design based on rapid risk analysis", Journal of Loss Prevention in the Process Industries 11:361-372, 1998. Dennis C. Hendershot, "Inherently safer chemical process design", Journal of Loss Prevention in the Process Industries 10:151-157, 1997.

start-up times. Plants with inherently safer technologies tend to be simpler in design, easier and more friendly to operate, and more tolerant of errors."[25]

Inherent safety also contributes to the reduction of insurance costs. This applies in particular to facility siting and the choice of distances in a plant between hazardous processes and places of work. The reduction of hazards through siting and spacing is much furthered by property insurance requirements.[26]

However, some of the economic gains from inherent safety may go unobserved, which may lead to higher costs and less safety than what would have been achievable. As was observed by Brian Moore of Eli Lilly and Company: "However, the costs associated with an accident or chemical release are typically not included in the economic analysis during the project's design stage. This may result in the analysis providing an incorrect or incomplete result."[27]

Hazard indices are important tools in the work for inherent safety in the chemical industry. Several hazard indices have been constructed for judging the inherent safety of a chemical plant. The most widely used of these is the Dow Fire and Explosion Index. These indices serve a useful purpose, but they are highly simplified and there is a large potential for improvement. Hence, all these indices are so simplified that inventories of chemicals are judged by their size, so that a large inventory of dangerous or harmless chemicals are judged the same way.[28] The indices are, generally speaking, less reliable for toxicity assessment than for the assessment of fire and explosion hazards.[29] In one study it was shown that the various indices varied so much that there was no similarity in the indices for toxicity, since the toxicity assessment method differed drastically between the models used in the different indices.[30]

Some commentators have complained that progress in the actual implementation of inherent safety is too slow.[31] According to Kletz, progress has been real, but the concept has not been adopted nearly as rapidly as quantitative risk assessment, that was introduced into the chemical industry only a few years earlier.[32] One reason for this may be that inherent safety is often difficult (but far from impossible[33]) to implement in existing plants. However, worries have also been expressed about new plants. Currently, bulk chemicals production is

---

[25] Tim Overton and George M King, "Inherently Safer Technology: An Evolutionary Approach", *Process Safety Progress* 25:116-119, 2006.
[26] Roy E Sanders, "Designs that lacked inherent safety: case histories", *Journal of Hazardous Materials* 104:149-161, 2003.
[27] Brian Moore, "Pharma and Semiconductor – Inherent Safety", *Process Safety Progress* 25:266, 2006.
[28] Mostafizur Rahman, Anna-Mari Heikkilä, and Markku Hurme "Comparison of inherent safety indices in process concept evaluation", *Journal of Loss Prevention in the Process Industries* 18:327-334, 2005.
[29] Mostafizur Rahman, Anna-Mari Heikkilä, and Markku Hurme, "Comparison of inherent safety indices in process concept evaluation", *Journal of Loss Prevention in the Process Industries* 18:327-334, 2005.
[30] Mostafizur Rahman, Anna-Mari Heikkilä, and Markku Hurme "Comparison of inherent safety indices in process concept evaluation", *Journal of Loss Prevention in the Process Industries* 18:327-334, 2005.
[31] Chidambaram Palaniappan, Rajagopalan Srinivasan, and Reginald Tan, "Expert System for the Design of Inherently Safer Processes. 1. Route Selection Stage", *Industrial and Engineering Chemistry Research* 41:6698-6710, 2002.
[32] Trevor A. Kletz, "Inherently safer design: The growth of an idea", *Process Safety Progress* 15:5-8, 2004.
[33] Nicholas Askounes Ashford and Gerard Zwetsloot, "Encouraging inherently safer production in European firms: a report from the field", *Journal of Hazardous Materials* 78:123-144, 2000. Dennis C. Hendershot, Jonathan A. Sussman, Gerald E. Winkler, and G. Lee Dill, "Implementing inherently safer design in an existing plant", *Process Safety Progress* 25:52-57, 2006.

transferred to low-cost countries. It has been argued that some of the new plants will be less inherently safe than corresponding plants in Europe and Northern America.[34]

It has repeatedly been proposed that inherent safety should be exported to other industries, including mining, construction, and transportation.[35] However, the only other industry that has adopted the concept of inherent safety is the nuclear industry. Much effort has been spent on developing nuclear reactors that are inherently safer than those currently in use. By this is meant that even in the case of failure of all active cooling systems and complete loss of coolant, fuel element temperatures will not exceed the limits at which most radioactive fission products remain confined within the fuel elements.[36]

In the recent discussions on the new European chemicals legislation (REACH), the substitution principle was one of the key conflicts, up to the last phase of the negotiations. It was introduced into the legislative process in the European Commission's 2001 White Paper that recommended "the substitution of dangerous by less dangerous substances where suitable alternatives are available".[37] This is a formulation that could have been taken from a paper by the experts on inherent safety of any of the major chemical industries. However, in spite of this, requirements of substitution became one of the major areas of conflict between the Commission and representatives of the chemical industry. This can be seen very clearly in a press release in December 2006 from the European Chemical Industry Council (CEFIC) and three other major business organizations. Here, the "balanced and open approach" of the final legislation was acknowledged. However, one major remaining complaint was expressed, namely the requirement of "a substitution plan for all the substances where a suitable alternative exists, even if they are adequately controlled". This requirement was said to generate a burden for both chemical producers and downstream users, without any benefit for the end consumer. According to this statement, substitution "does not automatically represent the best option in terms of safety, functionality or overall environment performance of a product".[38] Conspicuously absent from this text is the statement, so often made by safety experts in the chemical industry, that whenever possible one should replace a hazardous substance by a less hazardous one, rather than employ add-on safety equipment to deal with the hazard.

Hence, the European Commission adopted a safety principle that has a long history in the engineering practice of the chemical industry, but neither the Commission nor the chemical industry recognized the strong association between this principle and the chemical industry. Instead, the chemical industry strongly opposed it, and described its implementation as a major threat to their business interests.

2.4 The principle of using Best Available Technology (BAT)

The BAT principle is a well established risk management principle that is closely related to the principle of inherent safety and substitution principles.

---

[34] David W Edwards, "Export inherent safety NOT risk", *Journal of Loss Prevention in the Process Industries* 18:254-260, 2005.
[35] JP Gupta and David W Edwards, "A simple graphical method for measuring inherent safety", *Journal of Hazardous Materials* 104:15-30, 2003.
[36] G Brinkmann, J Pirson, S Ehster, MT Dominguez, L Mansani, I Coe, R Moormann, and W Van der Mheen, "Important viewpoints proposed for a safety approach of HTGR reactors in Europe. Final results of the EC-funded HTR-L project", *Nuclear Engineering and Design* 236:463-474, 2006.
[37] *Strategy for a future Chemicals Policy*, White Paper, European Commission 27.2.2001. http://europa.eu.int/eur-lex/en/com/wpr/2001/com2001_0088en01.pdf.
[38] *Implementation of REACH: A demanding challenge for industry*, Joint Press Release, 6 December 2006, http://212.3.246.117/Common/GetFileURL.asp?FileURL=F_1.

BAT is included in the EU rules for permitting and controlling industrial installations (Directive 96/61/EC) concerning integrated pollution prevention and control (the IPPC Directive). According to this directive, the BAT principle requires the use of "the most effective and advanced stage in the development of activities and their methods of operation", taking into account e.g. the use of less hazardous substances, the furthering of recovery and recycling of substances, technological advances and changes in scientific knowledge and understanding, the need to prevent or minimize the overall impact of the emissions, and the need to prevent accidents (European directive 96/61EC September 24th, 1996).

BAT is furthermore included in HELCOM (Convention on the Protection of the Marine Environment of the Baltic Sea Area), and the OSPAR convention (Convention for the Protection of the Marine Environment of the North-East Atlantic), and it has also been implemented in the U.S. e.g. in the New Source Review programme as part of the 1977 Clean Air Act Amendments.

2.5 Conclusion

We find it both enlightening and constructive to view the substitution principle as an application of the more general safety approach of inherent safety. The goals of inherent safety in the chemical industry have been adequately summarized by Faisal Khan and co-workers:

> "The salient features of an inherently *safe* plant are:
>
> - it uses harmless materials,
>
> - it contains small inventories of hazardous materials insufficient to cause significant harm even if released, and
>
> - the hazardous material is held in forms or under conditions that render them effectively harmless (diluted, at ambient temperature and pressure, etc.)"[39]

Clearly, "downstream" users of chemicals, i.e., the customers of the chemical industry, have much to learn from the principles of inherent safety that have been developed in the chemical industry. The same basic principles for inherent safety that are applied in the chemical industry can also be applied when other industries, farmers, consumers etc., use chemical products. One of these principles is that, other things being equal, the replacement of a hazardous product by a less hazardous one is always a worthwhile improvement. This applies also in cases when the hazard is believed to be "adequately controlled". As safety professionals of the chemical industry have repeatedly made clear, absolute safety in the control of a chemical hazard is not achievable. This is a reason for vigilance, and also for the elimination and reduction in use of hazardous substances.

In order to apply the substitution principle, knowledge and competences are needed that the chemical industry has to a much higher degree than most of their customers. The toxicological, chemical, and industrial considerations that are involved in successful applications of the principle can therefore to a large extent be provided as services by a competent supplier of chemicals to its customers. A chemical company can offer its customers less hazardous products for the chosen function or purpose, and it can in many cases offer to deliver a product in a less hazardous form. There does not see to be any conflict between the substitution principle and the business interests of a chemical company that possesses the core competences of a supplier of chemical products.

---

[39] Faisal I Khan, Rehan Sadiq, and Paul R Amyotte, "Evaluation of Available Indices for Inherently Safer Design Options", *Process Safety Progress* 22:83-97, 2003, p. 84.

3 THE SUBSTITUTION PRINCIPLE AND PRECAUTION

In informal discussions about the substitution principle we have found that its relation to precaution and to the precautionary principle are considered by many to be unclear. In this chapter we will investigate the relation of the substitution principle to cautiousness in general (Subsection 3.1) and to the precautionary principle (Subsection 3.2)

3.1 Cautiousness

To begin with, we need to clarify what is meant by being cautious (as distinct from applying the precautionary principle). A fairly operative definition can be obtained from mainstream decision theory. It defines cautiousness in relation to expected utility maximization.[40]

Expected utility maximization is the dominating approach to decision-making with known probabilities (decision-making under risk). It is "the major paradigm in decision making since the Second World War"[41] and most of modern risk analysis is based on it. It could, more precisely, be called "maximization of probability-weighted utility", since its central idea is to weigh the utilities (measures of usefulness) of outcomes according to their probabilities. For an example, let us consider the dilemma of a regulator who does not know whether or not a certain substance is a carcinogen. For simplicity, we may assume that she has only two options, namely either to prohibit the substance or not to do so. We can present this decision in the standard format of decision theory, that of a decision matrix. In a decision matrix, the alternatives open to the decision-maker are tabulated against the possible states of nature. The alternatives are represented by the rows of the matrix, and the states of nature by its columns. Each cell in the matrix contains a representation of the outcome that will ensue in the respective case:

	Carcinogen	**Not a carcinogen**
Prohibition	Costs, no fatalities	Costs, no fatalities
No prohibition	No costs, fatalities	No costs, no fatalities

In order to use a matrix to analyze a decision, we also need an evaluation of the outcomes. The most common way to represent the values of outcomes is to assign to each of them a real number that represents its utility. Utilities come in many variants. One common approach is to convert everything into money, and put a price on all factors that may influence how we evaluate an outcome (including human lives, lost species, etc.).[42] Another approach, favoured by utilitarian ethicists, is measure utility in an arbitrary unit. Suppose, for instance, that you regard the preservation of the giant panda to have twice as much value as the preservation of the Egyptian pyramids. Then it does not matter for the analysis if you assign to them the values 2 and 1 or the values 5000 and 2500. It is only the proportions that matter.

---

[40] The decision-theoretical jargon is unfortunately quite imprecise. Often, "pessimism", "cautiousness", and "risk-aversion" are taken to be synonyms. This is a much too crude approach. A pessimist is a person who assigns a higher probability than others to the most unwanted outcomes. A cautious person is one who takes more care than others to prepare for these outcomes.

[41] Schoemaker, P.J.H. 1982. The Expected Utility Model: Its Variants, Purposes, Evidence and Limitations. *Journal of Economic Literature* 20, 529-563, p. 529.

[42] On cost-benefit analysis, see Sven Ove Hansson, "Philosophical Problems in Cost-Benefit Analysis", *Economics and Philosophy*, in press.

When utilities are available, the standard approach is to enter them into the matrix, so that they replace the verbal descriptions of outcomes. In this way, the matrix is transformed into a *utility matrix*. In our example, negative numbers are appropriate:

	Carcinogen	**Not a carcinogen**
Prohibition	–5	–5
No prohibition	–20	0

In order to make a decision, we need at least one more piece of information that is not represented in the utility matrix: We need information about how plausible the various states of nature are. This information may vary widely in precision, ranging from exact probability assignments to vague statements about what is more or less plausible. In order to maximize expected utility we need exact probability assignments.

The expected utility of each alternative is the probability-weighted average of its utility values under different states of nature. In our example, suppose that the probability that the substance is a carcinogen is 0.2. Then the expected utility of the prohibition alternative is

$$0.2 \times (–5) + 0.8 \times (–5) = –5,$$

and that of the no-prohibition alternative is

$$0.2 \times (–20) + 0.8 \times 0 = –4.$$

Hence, under the given assumptions the rule of maximizing expected utility recommends us not to prohibit the substance.

Expected utility maximization is generally taken to be the *risk-neutral* decision rule. Decision rules that put more emphasis on avoiding the worst possible outcomes are called risk-averse. Decision rules that deviate in the opposite direction may be called risk-seeking or risk-taking. The most risk-averse decision rule is the maximin rule. It requires us to choose the option with the highest security level (= highest utility in the worst possible case).

Expected utility theory puts heavy demands on the information input. In order to apply it, we must assign exact numerical values to both outcomes and probabilities. In most cases of decision-making on chemicals risks, at least one of these is lacking, and consequently, the risk-neutral point cannot be precisely defined. As a consequence of this, cautiousness is not either precisely definable.

As we showed in Chapter 2, the substitution principle is part of a general safety engineering tradition that strives to avoid uncertainty. As we noted there, this tradition requires that even apparently well-controlled hazards should be removed or reduced if possible. The reason for this is that uncertainty can never be abolished, in other words that absolute safety in the control of a chemical hazard is not achievable. If the substitution principle is interpreted in this way, it includes an application of cautious decision-making.

3.2 The precautionary principle

Although the precautionary principle can be described as expressing a form of cautiousness, it should not be equated with cautiousness in general. The precautionary principle is much more specific than that.

In the Rio Declaration (Principle 15), the Precautionary Principle was expressed as follows: "Where there are threats of serious or irreversible damage, lack of full scientific certainty shall not be used as a reason for postponing cost-effective measures to prevent environmental degradation." At an international conference in Wingspread, Wisconsin, in January 1998, experts from Europe and Northern America agreed on the following definition: "When an activity raises threats to the environment or human health, precautionary measures should be taken, even if some cause-and-effect relationships are not fully established scientifically. In this context, the proponent of an activity, rather than the public, should bear the burden of proof."[43] These are among the most authoritative explications of the Precautionary Principle. They both combine two basic intuitions about what the Precautionary Principle aims at. First, this is a principle that requires us to be cautious, risk-averse, and on the safe side with respect to possible dangers. Secondly, it requires that we take action even in the absence of full scientific evidence.

In accordance with this tradition, we take the precautionary principle to be a principle for decision-making under scientific uncertainty. Since it concerns the relationship between scientific knowledge and regulatory decision-making, it is best understood against the background of an account of the scientific knowledge process and how it can be adjusted to the needs of practical decision-making.[44]

Scientific knowledge begins with data from experiments and other observations. Through a process of critical assessment, these data give rise to the scientific corpus. (See the diagram.) The corpus can, roughly, be described as consisting of those statements that could, at the time being, legitimately be made without reservation in a (sufficiently detailed) textbook. Alternatively it can be described as consisting in that which is taken for given by the collective of researchers in our continued research, and thus not questioned unless new data give us reason to do so.

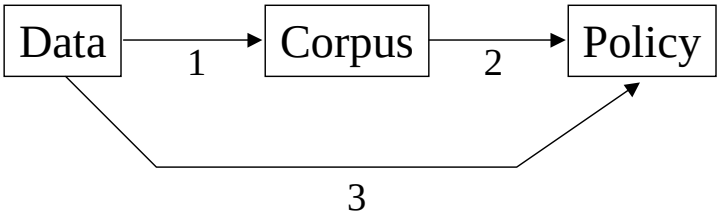
The corpus is constantly updated. In the discussions leading up modifications of the corpus, those who put forward a new scientific hypothesis or claim to have discovered a new phenomenon have the burden of proof. Fairly strict standards of evidence are applied when evaluating their argumentation. In other words, the corpus has high entry requirements. This has to be so in order to prevent science from erroneous conclusions.

But scientific knowledge is not only developed for its own sake. It is also used to guide decisions. Whenever it does, we must be careful to distinguish between the practical decision to be guided by science and the scientific process itself. Should we consider it an established fact that substance A causes developmental neurobehavioral defects in humans? This is a scientific issue, to be determined with the criteria of evidence that we have developed for the internal dealings of science. Should we prevent human exposure to this same substance in order to avoid potential neurotoxic effects? This is distinctly different issue, although the same scientific data that guides the scientific decision should be used here as well.

---

[43] Hileman, B. 1998. Precautionary Principle, *Chemical and Engineering News*, February 9, pp. 16-18.
[44] For a more detailed account, see Sven Ove Hansson, "Values in Pure and Applied Science", *Foundations of Science*, in press.

The most obvious way to use science for policy purposes is to employ information from the corpus (arrow 2 in the diagram).



For this the corpus is well suited in one important respect: The high entrance requirements guarantee that the information in the corpus is reliable enough to be trusted in almost all practical contexts. But from another point of view, the corpus is insufficient for many practical decisions: Due to the same high entry requirements the corpus will not contain all the information that may be useful for the practical decision. Information that did not make it into the corpus may nevertheless have sufficient evidential weight to legitimately influence some practical decisions. To exemplify this, suppose that we discover that a certain toxic substance leaks from feeding bottles for babies. Furthermore, suppose that there is weak but relevant evidence that this particular substance may be toxic to humans, and that most experts consider it equally plausible that there is a toxic effect in humans and that there is not. Given what is at stake in this situation, it would be perfectly rational for the company that produces the bottles, or for a government agency, to decide on this basis to remove the substance from the production of new bottles, and perhaps to take some measures concerning the ones already in use. Such a decision would have to be based on scientific information that did not satisfy the criteria for corpus entry. In other words, a direct road from data to policy is required (arrow 3 in the diagram).

This bypass route for scientific information is practically important in chemicals regulation. We typically wish to protect ourselves against suspected health hazards even if the evidence is weaker than what is required for full scientific proof. Many of us have similar attitudes to protection of the natural environment. This leads us to take the bypass route, i.e. to base practical decisions on less than full scientific proof. Unfortunately, this is a difficult road to take. There is a risk that scientific data will be judged according to whether they suit preconceived policy ideas rather than according to their scientific value. When conflicting parties in a decision use science in this way, the result can be a “science charade”[45] in which policy disagreements are camouflaged as disagreements on scientific detail. In order to avoid this, the following three simple principles should be applied:

1. The same type of evidence should be taken into account in the policy process as in the formation of the scientific corpus. Suppose, for instance, that we want to decide whether or not to restrict the use of a substance that is suspected to be a reproductive toxicant. Then the same type of toxicity studies should be used as a bases for the policy-specific process (arrow 3) as in ordinary science (arrow 1). Policy decisions are not served by the use of irrelevant data.

[45] Wendy E Wagner Wendy E Wagner "The Science Charade in Toxic Risk Regulation", *Columbia Law Review*, 1995.

2. The assessment of how strong the evidence is should be the same in the two processes. If there is stronger scientific evidence that exposure to substance A leads to toxic accumulation in biota than that exposure to substance B does so, then this evidence should be counted as stronger in policy discussions as well.

3. The two processes may however differ in the *required* level of evidence. It is a policy issue how much evidence is needed for a decision to restrict the use of a substance. The chosen level of evidence is a matter of practical, not theoretical rationality. This means that non-scientific criteria, such as how we appraise the severity of the possible danger, have a legitimate role.[46]

These three principles summarize what we have elsewhere called *science-based precaution*.[47] The precautionary principle thus interpreted, is not synonymous with the substitution principle. Most basically, the precautionary principle concerns the criteria for when to take protective action whereas the substitution principle primarily concerns preference for one particular form of such protective action. However, the two principles are part of the same type of cautious thinking. In actual applications they tend to support and strengthen each other.

To exemplify how the two principles can be combined in a proactive policy, consider two substances A and B. We know that substance A is transformed to CO_2 fairly fast whereas B is a vPvB substance, i.e. it is very persistent and very bioaccumulative. Nothing is known about the potential ecotoxicity of the two substances. Hence, we do not know if any of them damages the environment. Should they then be treated equally? A strong case can be made that they should not, for the simple reason that if substance B turns out to be for instance a reproductive toxicant, then the effects will be much more severe than if substance A turns out to have such an effect. This can be used as an argument for substituting substances like B by less hazardous substances, even if we do not substitute substances like A. This is a plausible conclusion that can be reached by the combined application of the substitution and the precautionary principles.

[46] This is reflected in some cases by the use of assessment factors to compensate for uncertainties and data gaps.
[47] Hansson, S.O. *Philosophical Perspectives on Risk*. Keynote address, Research in Ethics and Engineering, Delft , April 25-27 2002. Available at http://www.infra.kth.se/~soh/downloads.htm. Hansson, S.O. and Rudén, C. (Eds.) *BETTER CHEMICALS CONTROL WITHIN REACH* Printed by US-AB, 2004 Stockholm, Sweden. ISBN 91-7283-704-7.

4 SUBSTITUTION AND RISK ANALYSIS

Both the substitution principle and the precautionary principle have often been compared to risk analysis. Claims have been made that risk analysis represents a more adequate way of thinking about chemical risks, and that the substitution and precautionary principles stand in the way of rational decision-making.

In the same way that we distinguished in Chapter 1 between substitution and the substitution principle, it is important to distinguish between on one hand risk analysis as a methodology and on the other hand the thesis or principle that decisions on risk should comply with the recommendations of risk analysis. The latter standpoint does not have an established name; we propose to call it the *RA postulate*. In subsection 4.1 we will introduce risk analysis and in subsection 4.2 the RA postulate. In subsections 4.3 and 4.4 we discuss the problem of unknown risks that is a central problem for the RA postulate, and in subsection 4.5 we discuss how the substitution principle, the precautionary principle and the RA postulate relate to each other.

4.1 Risk analysis

The discipline of risk analysis grew out in the late 1960's and the 1970's, largely as a response to growing public opposition to new technologies. From the beginning the discipline has had a strong focus of attention on risks that are subject to public opposition, such as nuclear energy and chemicals. Risk analysis is a multi-disciplinary field that uses methods from the natural, behavioural, and social sciences. Major areas of study include risk perception, risk communication, risk-benefit analysis, and various industrial applications such as fault tree analysis. In toxicological applications, risk analysis is often integrated with epidemiology.

All the various approaches to risk analysis are dominated by a particular definition of risk, namely as the statistical expectation value of the unwanted effects under investigation. In other words, risk analysis operates as a branch of expected utility analysis. (On expected utility, see subsection 3.1.) The procedure is to multiply the probability of a risk with its severity, and to call the expectation value obtained in this way "the risk". This definition of risk does not coincide with common usage of the term, and it is in fact a problem for risk analysis (and in particular risk communication) that the key term of the discourse has so different meanings in everyday and specialized language. Although expectation values have been calculated since the 17th century, the use of the term "risk" in this sense is relatively new. (It was introduced in the influential Reactor Safety Study (WASH-1400, the Rasmussen report) from 1975.[48]) Today it is claimed by many risk analysts to be *the* meaning of the term "risk" – a linguistic policy that is often integrated with the RA postulate.

4.2 The RA postulate

The methodology of risk analysis, including the calculation of expectation values, can be seen as a decision aid, the outcome of which should be taken into account by decision-makers along with various other considerations. However, many proponents of risk analysis make a much stronger claim than that. They claim that a rational decision-maker is bound to judge risk issues in accordance with these expectation values ("risks"), so that an outcome with a smaller expectation value ("risk") is always preferred to an outcome with a larger expectation value. The following quotations illustrate this way of thinking:

---

[48] Rechard, RP "Historical Relationship Between Performance Assessment for Radioactive Waste Disposal and Other Types of Risk Assessment", *Risk Analysis* 19(5): 763-807, 1999, p. 776.

> "The only meaningful way to evaluate the riskiness of a technology is through probabilistic risk analysis (PRA). A PRA gives an estimate of the number of expected health impacts – e.g., the number of induced deaths – of the technology, which then allows comparisons to be made with the health impacts of competing technologies so a rational judgment can be made of their relative acceptability. Not only is that procedure attractive from the standpoint of scientific logic, but it is easily understood by the public."[49]
>
> "The precautionary principle is threatening to take the place of risk analysis as the basis for regulatory decision making in a number of places, particularly in Europe."[50]
>
> "All chemical management decisions should be based on risk."… "Substitution is only acceptable provided the suitability of the alternative has been demonstrated through socio-economic costs/benefits analysis and a risk assessment."[51]

The last two of these quotes also illustrate how the RA postulate is contrasted with the precautionary principle and the substitution principle.

The RA postulate has several problematic features. It excludes considerations that may be relevant in the assessment of risk, in particular ethical consideration[52], and its basic assumptions about decision-theoretical rationality may also be put into question. Here, we will focus on a problem for the RA postulate that is particularly important in the context of chemical hazards, namely how it deals with unknown and unknowable risks.

4.3 Known and unknown risks

As should be well-known, information about the toxicity of numerous chemical substances is insufficient and in many cases non-existent. For most industrial substances on the market, no toxicological data are at all publicly available, and with few exceptions even when data are available they are much less complete than what a risk assessor would wish.

In addition to the lack of data, a more fundamental problem complicates risk assessments of chemicals: Since many effects are stochastic they may be undetectable for statistical reasons.[53] We can use two simple examples to illustrate the problems involved in determining the presence or absence of lethal effects through direct studies of exposed humans.

For the first example, suppose that 1000 persons are all subject to a chemical exposure that gives rise to hepatic angiosarcoma among 0.5 % of the exposed. Among unexposed individuals, the frequency of this disease is very close to zero. If a proper investigation is made, chances are very high that the overrepresentation of this disease among the exposed population will be discovered.

Next, suppose that another group of 1000 persons are subject to an industrial exposure that increases the incidence of lung cancer from 10.0 to 10.5 %. The expected number of additional cancer cases is the same as in the previous case. However, as can easily be shown with probability calculus, the difference between 10.0 and 10.5 % is in this case

---

[49] Cohen, Bernard L. "Probabilistic Risk Analysis for a High-Level Radioactive Waste Repository", *Risk Analysis* 23:909-915, 2003, p. 909.
[50] Charnley, G. "President's Message", *RISK newsletter*, 19(2), 1999, p. 2.
[51] *CEFIC paper on substitution and authorisation under REACH*, 23 May 2005, pp. 1 and 3.
[52] Sven Ove Hansson, "Ethical criteria of risk acceptance", *Erkenntnis*, 59:291-309, 2003. Sven Ove Hansson, "Philosophical Problems in Cost-Benefit Analysis", *Economics and Philosophy*, in press. Hélène Hermansson and Sven Ove Hansson "A Three Party Model Tool for Ethical Risk Analysis", *Risk Management*, in press.
[53] For a more detailed discussion, see Sven Ove Hansson, "The Moral Significance of Indetectable Effects", *Risk* 10:101-108, 1999.

indistinguishable from random variations. Hence, the effects of this substance cannot be detected in a study of the exposed population.

As a rough rule of thumb, epidemiological studies can reliably detect excess relative risks only if they are about 10 % or greater. For the more common types of lethal diseases, such as coronary disease and lung cancer, lifetime risks are of the order of magnitude of 10 %. Therefore, even in the most sensitive studies, an increase in lifetime risk of the size 10^{-2} (10 % of 10 %) or smaller may be undetectable (i.e. indistinguishable from random variations).

Animal experimentation has similar statistical problems. It is easy to show with examples that an effect has to be quite large in order in order to be discovered in animal groups of the sizes conventionally used in experiments (seldom more than 100–200). However, in animal experiments another option is available. The sensitivity of animal tests can be increased by increasing the dose to levels much higher than those that are relevant for human exposure. As an example, suppose that a cancer test is performed on an animal strain whose normal incidence of female infertility is 10 %. If the substance increases this incidence to 11 % at the doses humans are exposed to, then this increase will be hidden by chance effects unless a very large number of animals is used. Now suppose that the administered dose is increased so that it is high enough to yield an increase of the incidence by 10 instead of 1 per cent units, so that 20 % of the animals will be affected. Then the effect can be detected in an experiment of reasonable size.

When high-dose tests are used for predictive purposes, it is assumed that effects at high dose levels are indicators of corresponding effects at lower levels. In addition to this qualitative assumption, a quantitative assumption is often made to the effect that the size of the effect at low levels can be inferred from that at high levels, usually on a simple proportionate basis (the ”linear model”). In other words, both qualitative and quantitative *dose extrapolations* from higher to lower doses are made.

Extrapolations from high to low doses are of course fraught with uncertainties, and so are extrapolations from one species to another. This has led some risk managers to base their decisions primarily on human data, and to consider much animal data as more or less irrelevant for decision-making. However, it has often been overlooked in discussions on animal data that when making ourselves more dependent on human data we do not get rid of uncertainty. Instead, we trade uncertainties in dose and species extrapolation for the severe (but less understood) statistical uncertainties in studies on humans.

A comment is in place about the ethical aspects of these uncertainties. There is of course no objective answer to the question how small health effects should be of concern to us. However, many attempts have been made to set a limit of concern, expressed either as “acceptable risk” or “de minimis risk”. Most people seem to agree that if a human population is exposed to a risk factor that will, statistically, kill one person out of 10^9, then that risk will not be an issue of high priority. Arguably, it is no big problem that our risk assessment methods are insufficient to discover risks of that order of magnitude. On the other hand, most of us would consider it a serious problem if a risk factor that kills one person out of 100 or 1000 cannot be detected. The most common proposals for limits of concern for lethal risks are 1 in 100,000 and 1 in 1000,000. It is difficult to find proposals above 1 in 10,000. These values are of course not objective or scientific limits; they belong to the ethical realm. However, it is important to note the presence of an *ethical gap*, a gap between those risk levels that are scientifically detectable and those that are commonly regarded to be ethically acceptable or at least of minor concern. This gap seems to have a breadth of 2–4 orders of magnitude. Hence, even if no adverse effects have been found in exposed populations, there

may still be effects that are at least 100 to 10000 times larger than commonly proposed levels of concern or acceptability.

There is an obvious regulatory solution to this ethical gap, namely the introduction of sufficiently large uncertainty factors (safety factors) to bridge the gap. If the dose-response relationship is linear, then the safety factor will have to be equal to the ethical gap. Hence, if the largest effects that we cannot discover are 100 times bigger than the largest effects that we would accept, then (under the linearity assumption) an uncertainty factor of 100 is needed for protection.

4.4 Risk analysis and unknown risks

As was mentioned in subsection 3.1, expected utility maximization is generally taken to be the *risk-neutral* decision rule. It is also generally assumed that probabilistic risk (and risk-benefit) analysis in its common form is risk-neutral since it employs expected utility maximization. However, this is in general not so, due to inadequate representation of the options.

In order to see this we need to make the distinction between *endodoxastic* and *metadoxastic* uncertainty.[54] Endodoxastic uncertainty is the uncertainty expressed or inherent in an assessment. It can be communicated by the assessor either in probabilistic or non-probabilistic terms. Metadoxastic uncertainty is uncertainty about whether or not an assessment is correct, or about the choice between alternative assessments.

A study of actual practices in risk analysis, for instance as they are presented in various risk assessment reports, will confirm that expected utility maximization is the standard approach to endodoxastic uncertainty. In contrast, metadoxastic uncertainty is mostly dealt with by selecting or developing the most credible assessment, and then relying on that assessment in the analysis. (More sophisticated approaches, that take into account the uncertainty deriving from the choice between different models and assessments, can be found in risk assessments performed in the nuclear sector, but more seldom in other areas such as the risk assessment of chemicals.[55]) As a consequence of this practice, other possible assessments with lower but non-negligible credibility will have no influence on the calculations.[56]

For an example of this, consider a decision of what risk management measures to take with respect to the potential acute effects of a particular substance. This substance, we may assume, has not been investigated with respect to acute toxicity, so there is no substance-specific data on which a risk assessment can be based. In order to perform a risk analysis we

---

[54] For details, see Sven Ove Hansson, "Economic (ir)rationality in risk analysis", *Economics and Philosophy*, 22:231-241, 2006.

[55] Rechard, RP "Historical Relationship Between Performance Assessment for Radioactive Waste Disposal and Other Types of Risk Assessment", *Risk Analysis* 19(5): 763-807, 1999, pp. 766 and 781.

[56] Another way to express this is that metadoxastic uncertainty is often *de facto* dealt with in accordance with the maxiprobability rule, which consists in disregarding all alternatives except the one that has the highest probability. This is not a decision rule that has been proposed by decision-theorists, but rather a method that decision-makers can be observed making use of. (Sven Ove Hansson and Mikael Johannesson, "Decision-Theoretic Approaches to Global Climate Change", pp. 153-178 in Gunnar Fermann (ed.) *International Politics of Climate Change*, Scandinavian University Press 1997. Sven Ove Hansson, "Adjusting Scientific Practices to the Precautionary Principle" *Human and Ecological Risk Assessment*, 5:909-921, 1999.) The epistemic requirements of the maxiprobability rule are less demanding than those of expected utility maximization. Thus, in order to apply the maxiprobability rule to a case of metadoxastic uncertainty between different sources of information we only need to know which is the most reliable source, and use the information obtained from it. In order to maximize expected utility in such a case, we will have to integrate the information received from the different sources, assigning to each of them weight according to the estimated probability that the information it provides is correct.

need value assignments to the various outcomes. Let us assume that the costs of handling the substance in the same way as known acutely toxic substances is estimated at €5000. Furthermore, let us assume that the expected negative effects of treating an acutely toxic substance as a non-toxic substance is €50000 (for instance costs associated with disease).

The standard approach in chemical risk assessment to this principle is very simple. There "is no risk" (i.e. no known risk), so the situation is perceived as in the following matrix:

Standard risk analysis

Treat as toxic	–€5000
Treat as non-toxic	€0

Hence, standard risk analysis will recommend us to treat this substance in the same way as substances that we have goods reasons to treat as non-toxic.

Next, let us apply expected utility theory properly to this decision problem. Then we need to take into account the possibility that the substance may be toxic, and furthermore we have to assign a probability to the state of affairs that this is the case. The best estimate that we are aware of is 20%. (20% of the substances regulated as new chemicals in the pre-REACH European legislation satisfied the criteria for acute toxicity.[57]) Using this estimate we obtain the following, somewhat more complex decision matrix:

Correct application of expected utility

	Substance is non-toxic (p = .8)	Substance is toxic (p = .2)
Treat as toxic	–€5000	–€5000
Treat as non-toxic	€0	–€50000

It then emerges that the expected utility of treating the substance as non-toxic is –€10000, whereas that of treating it as toxic is –€5000. In other words, expected utility analysis will recommend us to treat the substance as if it were toxic.

(Like all numerical examples, this one relies on a number of simplifying assumptions. One of these is particularly important from a regulatory perspective, namely that there may be other, perhaps intermediate alternatives for the treatment of a substance with unknown properties than treating it either as if it were known to be non-toxic or as if it were known to be toxic.)

Metadoxastic uncertainty is always present in chemical risk analysis, but it is seldom explicitly discussed. Instead, there is traditionally a strong focus on the attainment of consensus and on basing decisions exclusively on the best available expert assessment. This

[57] Nordberg, A., Rudén, C., and Hansson, S.O. "Towards more efficient testing strategies - analyzing the efficiency of toxicity data requirements in relation to the criteria for classification and labelling", submitted manuscript

.

approach is clearly problematic from a decision-theoretical point of view. Although expected utility maximization is controversial it has strong theoretical arguments in its favour. The support that risk analysis can have in more basic normative theory can only be drawn from the normative reasonableness of expected utility maximization. To the extent that risk analysis deviates from the principles of expected utility, it loses its normative credibility.

4.5 Risk analysis, precaution and substitution

It is instructive to compare in the context the standard approach in chemical risk analysis to applications of the precautionary principle. Whereas standard risk analysis disregards metadoxastic uncertainty, the precautionary principle urges us to focus on it and let it influence decision-making. One possible version of the precautionary principle could be to assign weights to unproven but scientifically credible dangers, and adjust these weights so that they accord with the estimated probabilities of these dangers. Such a variant of the precautionary principle would be much closer than standard risk analysis to a risk-neutral treatment of metadoxastic uncertainty.

The following example illustrates the interplay between expected utility, precaution and substitution. Consider three substances A, B, and C that are alternatives for being used in an application where they will be spread into the aquatic environment. B is persistent and bioaccumulative, whereas both A and C are readily degraded in the environment. A and C have both been extensively tested for ecotoxicity, and we have very good reasons to assume that A is ecotoxic and C non-ecotoxic. B has not been tested for ecotoxicity. The best possible estimate (based on structural analogies) is a 5% probability that B is ecotoxic. In order to apply expected utility analysis we need to assign tentative numerical values to possible outcomes. We can assume that the (dis)value of spreading a toxic substance that is persistent and bioaccumulating (PB) is –500 units whereas that of spreading a toxic substance that is non-PB is –10 units. Applying standard risk analysis, we obtain the following matrix:

Standard risk analysis

A (non-PB, ecotoxic)	–10
B (PB, ecotoxicity unknown)	0
C (non-PB, not ecotoxic)	0

Hence, standard risk analysis would support the substitutions A→C (and indeed the substitution A→B) but it would not support the substitution B→C. This is clearly in conflict with the precautionary principle, that would support the substitution B→C.

Next, let us apply expected utility theory in the proper way to this problem. We then obtain the following matrix:

Correct application of expected utility

	B is not ecotoxic (95%)	B is ecotoxic (5%)
A (non-PB, ecotoxic)	–10	–10
B (PB, ecotoxicity unknown)	0	–500
C (non-PB, not ecotoxic)	0	0

Here, the expected utility of using substance A is –10, whereas the corresponding value for substance B is –25, and for substance C it is 0. Hence, this analysis supports not only the substitution A→C but also the substitution B→C that is supported by the precautionary principle. (It also supports the substitution B→A, but that is inconsequential when C is available.)

Hence, if the RA postulate is interpreted in accordance with prevailing traditions in chemical risk analysis, then it conflicts with the precautionary principle and with a precautionary application of the substitution principle. If, instead, the RA postulate is interpreted so that it takes metadoxastic uncertainty into account, then its application will be accord with precautionary and substitution principles. As already mentioned, it is only the latter interpretation of the RA postulate that can draw support from expected utility theory.

5 COMPARATIVE RISK ASSESSMENT

As we said in chapter 1, substitution involves comparing two or more alternatives (chemical – chemical, or chemical – non-chemical) using a set of criteria or objectives. The purpose of substitution is to reduce environmental, safety, and/or health risks while maintaining an acceptable functionality of the process at a reasonable cost. Substitution can thus go wrong in at least three ways; The substitute entails a higher (or even equal) risk compared to the original substance, the substitute does not fulfil the required level of functionality, or the cost will be significantly higher than what is expected and motivated. The uncertainties in the different aspects underlying a substitution decision can be more or less pronounced. In some cases the benefits of a substitution may be obvious, e.g. there is no trade-off between functionality, cost, and risk reduction. In other cases there might be large uncertainties about one or more aspects, and the outcome of a substitution is far from easily estimated. In these cases substitution may not be the best choice of method for risk management.

In this chapter we will focus on the hazard/risk aspect. Chemical risk is determined by the exposure in combination with the hazard, i.e. the inherent properties of the chemical to cause harm in a biological system. The hazard is defined by its severity (type of adverse effect or endpoint) and the potency (the doses/exposures needed for the adverse effect to materialize). Two chemicals used for the same purpose may give rise to different exposure patterns, due for instance to differences in volatility or in the volumes needed to achieve functionality. To be able to compare two alternative chemicals with respect to their risk, both exposure and hazard information is needed.

For some groups of chemicals we have access to a reasonable amount of toxicity and exposure data, and that data have in many cases been generated for regulatory purposes using reliable and standardized methods. Pesticides is one well-known such group of chemicals. On the other hand, for the large number of general, industrial chemicals, lack of data is a major obstacle in comparing risks or hazards among chemicals.

A large number of methods/models have been proposed for evaluating various technological processes and their (potential) impact on health and/or the environment (e.g. Rapid Risk Analysis, Inherent Safety Indices, Life Cycle Analysis, Programmatic comparative risk analysis, Comparative risk analysis of alternatives, Risk Trade Off analysis, Health-health analysis, Benefit-cost analysis, Cost-effectiveness analysis, Environmental Impact Assessment, Ecological Risk Assessment, Material Flow Analysis, HAZOP, Dow Fire and Explosion Index, METRIK). Together they are sometimes referred to as “environmental systems analysis tools”.

A sub-set of these approaches focus on the relative evaluation (ranking or categorization) of chemicals. About 150 chemical ranking and categorizing systems have been proposed, and fifty-one such systems for chemicals are reviewed by Davies *et al*.[58]

5.1 Toxicity and classification

For estimating the potential hazard associated with the use of a particular chemical substance, these systems include criteria for different types of experimental data to evaluate adverse effects in the environment or to human health. Some systems include bioaccumulation and/or persistence in the environmental effects, others include these parameters as exposure

[58] Davis, G.A. and Swanson, M. “Comparative Evaluation of Chemical Ranking and Scoring Methodologies”. EPA Order No. 3N-3545-NAEX University of Tennessee Center for Clean Products and Clean Technologies, April 7, 1994.

variables. The manner in which systems evaluate toxicity varies significantly among the many different systems and the specific endpoints selected to represent health effects are numerous.[59] Sometimes the toxicity is aggregated to one single parameter "general toxicity" including both environmental and health effects, sometimes the approach is more elaborate.

Potency is sometimes taken into account, typically for acute toxicity endpoints such as lethality, for which the dose required to elicit the effect (i.e. potency) is commonly reported (e.g. LC50). For chronic effects and carcinogenicity a weight-of-evidence approach is more common. There are good reasons to also consider the potency of a chemical also for other types of effects.[60] The severity of an effect is a function of both the nature and size of the effect. Many systems do not include measures of severity at all.

The classification and labelling system can be seen as a way to group chemicals according to the adverse effects that they can cause and according to the severity of that effect. This is represented by the different danger classes, categories, warning symbols, and risk phrases that can be used for comparative purposes. Through the definition of the different classes, the system contains an implicit comparison of different types of effects. The system has major advantages such as being well-established and experience-based, well-defined in terms of standardized tests, and closely connected to regulatory measures. Its major draw-back is that untested chemicals and endpoints will remain unclassified, so a lack of a classification cannot be taken to imply lack of danger. An untested substance that may or may not have serious toxicity comes out in the same way in this system as a substance that has been thoroughly tested without any sign of toxicity. This, however, is a drawback that the classification system shares with the other assessment systems for toxicity that we are aware of.

5.2 Exposure data

For a detailed exposure assessment, an internal dose in the critical organ/target is ideally wanted, and measured or modelled concentrations can be used. However, exposure assessment for ranking of chemicals is usually not performed at the degree of precision that is needed to make such comparisons possible. Alternative, and simpler, approaches draw on production or use volume as a major input. Furthermore a multimedia fate and transport model to estimate the distribution of chemicals in the environment based on release data can be added. Further input data are related to the properties of the chemical determining its bioavailability, fate and behaviour in the environment (use patterns, degradation/metabolism, persistence/reactivity, lipophilicity, bioconcentration, flammability, etcetera.)

For human exposure assessment some intermediate level of data precision should be considered. For chemicals already in use, measurements from actual exposures at workplaces and elsewhere can be used.

5.3 Approaches to incomparability

In evaluating alternative chemicals, there are at least two types of problematic comparisons.

The first type consists in comparing different effect types. This problem arises when we have a similar dataset for all alternatives, but the results of the different tests differ so that you need to weigh one type of effect against another (a chemical with high acute toxicity vs. a low potent carcinogen for instance). We will call this the *evaluation problem*.

---

[59] Davis, G.A. and Swanson, M. "Comparative Evaluation of Chemical Ranking and Scoring Methodologies". EPA Order No. 3N-3545-NAEX University of Tennessee Center for Clean Products and Clean Technologies, April 7, 1994.

[60] Sven Ove Hansson, "Choosing Priority-Setting Criteria for Carcinogens", *Human and Ecological Risk Assessment* 7:475-491, 2001.

The other type consists in comparing data to lack of data. This problem arises when there are different datasets for the different alternatives, so that we have to compare something that is un-known for one chemical to some (more or less) known property of the alternative, e.g. a tested and classified sensitizer vs. a chemical untested for sensitization. We will call this the *data gap problem*.

As previously mentioned, for large groups of chemicals the available data necessary for complete assessment of hazards and risks are limited. Therefore, the second type of problem is a practical reality in virtually all applications of the substitution problem.

In this section we will discuss some major approaches to comparative risk assessment that try to deal with one or both of the above-mentioned problems. We begin with four general approaches that were identified by Davies *et al.*[61]

1. Assign one endpoint per criteria being assessed and estimate missing data on the basis of the available data E.g. use the rat oral LD_{50} for (acute) human toxicity criteria. This can include estimating the toxicity of one compound based on the toxicity of other, structurally similar, compounds i.e. group assessment. OECD is currently developing a "toolbox" with the purpose to facilitate such extrapolation of data within groups of chemicals which may be useful in such an approach.

This is an attempt to solve the data gap problem. It can in many cases be a valuable contribution to risk assessment, in particular since it helps identify substances for of potential concern

2. Choose data from a hierarchy of endpoints, listed in order of preference, based on data quality, appropriateness of test, etc. Preference is given to high quality data. Data of lower quality may be used in the absence of high quality data. If multiple data are available within the high quality classification, then the data yielding the highest score (i.e. most conservative or health-protective) is used.

This seems to primarily a method to solve the evaluation problem, and as such it is useful. However, it requires the availability of a substantial data-set and does not contribute to solving the data gap problem. (Data gaps are treated as indicating no toxicity).

3. Choose the most conservative value from a pool of different endpoints. In systems where a group of endpoints could be considered for a given criterion, then the data yielding the most conservative (health-protective) result are chosen. If multiple data are available for a particular criterion (e.g. acute mammalian toxicity), then the data resulting in the highest estimated risk are used.

This, too, is primarily a method to solve the evaluation problem. It corresponds to regulatory praxis in many areas. One drawback that should be noted for such a system is that it does not create incentives for further testing since further data are unlikely to reduce risk estimates. Furthermore, it promotes the use of untested chemicals instead of tested ones, which should not be an inherent preference of the system.

4. Assign cut-off (or trigger) values to a large number of criteria and select a chemical if one or a specified number of the criteria are met. Several systems designed to select chemicals for regulatory action or for further study do not provide an overall score for a

[61] Davis, G.A. and Swanson, M. "Comparative Evaluation of Chemical Ranking and Scoring Methodologies". EPA Order No. 3N-3545-NAEX University of Tennessee Center for Clean Products and Clean Technologies, April 7, 1994.

chemical. Rather, a chemical is selected if it meets certain criteria. An example of this type of system is cut-off values to determine if a substance is to be considered PBT.

This is also a method so solve the evaluation problem. It does not deal with the data gap problem. (Here as well, data gaps are treated as indicating no toxicity).

In addition to these four methods, we would like to mention three other methods (all of which have been investigated in the NewS research programme).

5. SARs and QSARs are can be used to estimate chemical properties for which data are lacking. This approach is widely used in the U.S. for managing data gaps in assessing the risks with general industrial chemicals. Their (regulatory) use in Europe has been limited so far, but an increased regulatory acceptance is foreseen within the REACH system.

This methodology can be used to deal with the data gap problem. Just like method 1 above, it can in many cases be a valuable contribution to risk assessment, in particular since it helps identify substances of potential concern

6. Expert judgment is used in a few systems to give qualitative or semi-quantitative scores to chemicals (e.g. high, medium, low). Extensive use of qualitative judgement may lead to results not being scientifically reproducible.[62]

7. Finally, assigning **default values** is another approach to filling data gaps [63]A default value is an assumption that is used in the absence of adequate information and that should be replaced when such information is obtained.

For intra-scientific purposes, if we are not reasonably certain of, for instance, the fate of a substance in the environment or of its effects on fish, then we say that it has an unknown fate and that its effects on fish are unknown. For decision-making purposes, this is not sufficient. We have to treat the substance in one way or the other while waiting for more evidence. Hence, for regulatory purposes, a substance with unknown toxicity will have to be treated for instance *as if* it were severely toxic, *as if* it were moderately toxic, *as if* it were non-toxic, etc. It is in practice unavoidable that a regulatory decision (or non-decision) on a substance with unknown properties will have the effect of treating it in the same way as if its properties were known. Rules for risk management decisions under lack of information can therefore be described as *(regulatory) defaults*.

In our view, a combination of these methods will be needed. In order to solve the data gap problem, a combination of methods 1, 5 and 7 would seem reasonable. The latter method needs some specifications, to which we will now turn.

5.4 Risk-neutral defaults

The standard procedure in many branches of risk management is still to treat the unknown as non-hazardous. This means in practice that a default of zero toxicity is used if data is lacking. As we have explained in sections 3.1 and 4.4 - 4.5, this approach is not supported by standard

---

[62] See e.g. Morgenstern, R.D., Shih, J-S., Sessions, S.L. “Comparative risk assessment: an international comparison of methodologies and results” *J Haz Mat* 78:19-39.

[63] S Sandin, P., and Hansson, S.O. (2002) "The Default Value Approach to the Precautionary Principle" *Journal of Human and Ecological Risk Assessment* (HERA) 8:463-471. Per Sandin, Bengt-Erik Bengtsson, Åke Bergman, Ingvar Brandt, Lennart Dencker, Per Eriksson, Lars Förlin, Per Larsson, Agneta Oskarsson, Christina Rudén, Anders Södergren, Per Woin, and Sven Ove Hansson, “Precautionary Defaults – A New Strategy for Chemical Risk Management”, *Human and Ecological Risk Assessment*, 10(1):1-18, 2004; Hansen , B.G. *et al* “Priority setting for existing chemicals: European union risk ranking method” *Environmental Toxicology and Chemistry* 18(4):772-779.

decision-theory. To the contrary, treating a substance with unknown properties in the same way as a substance known to be innocuous is an irrational behaviour from a decision-theoretical point of view. (The term “risk-based” is sometimes used for this method, but that does not make it rational.)

The standard view of decision-theoretical rationality, expected utility maximization, requires instead that we treat such a substance as if it had the statistically expected degree of toxicity. Expected utility maximization is generally recognized as the “risk-neutral” decision method, i.e. it is neither risk-averse (“precautionary”) nor risk-seeking.

Mathematically, let *s* be a variable representing the degrees of toxicity of a substance, and let *p* represent our best probability estimate. Then a substance with unknown toxicity should, according to this principle, be treated in the same way as a substance with the severity

$$\int_{s=0}^{\infty} p(s) \times s$$

We can call this a *risk-neutral default* for toxicity. The difficulties in applying it in practice are of course to find an adequate variable *s* and an adequate probability distribution for it.

For simplicity, we will discuss this problem with respect to acute toxicity. There is a well-established potency measure for acute toxicity, namely the LD50 value.

Consider two substances A and B such that A has a 10 times higher LD50 than B. Then roughly speaking it is (ceteribus paribus) ten time worse to be exposed to B than to A. We can say that B is ten times more acutely toxic than A. Generalizing the argument, the inverse of the LD50 value can be used as a potency measure.

Practically speaking, in order to find the risk neutral default for acute toxicity, we should therefore find a set of substances that we believe to be representative for substances with unknown toxicity. The harmonic mean of the LD50 values of these substances can then be used as a risk-neutral default value for new substances. (The harmonic mean is the inverse of the arithmetic means of the inverses.) It is important to observe that due to the large effect of highly toxic substances this value will be higher than the median.

In discussions of substitution, this approach to toxicity of unknown should of course be applied equally to substances in use and substitution candidates not yet in use. We propose therefore that “average toxicities” for the universe of industrial chemicals be calculated, for use as default values in situations where no data is available. When data is available these defaults can be adjusted or replace accordingly (for instance with methods 1 and 5 as mentioned in section 5.3.

6 IMPLEMENTATION

Substitution should not be seen as a single decision but as a continuous development towards safer processes. Decisions to substitute can be taken at different levels of organization from the primary producer of a product, deciding on which products to manufacture and on the production processes and composition of those products, to the users of the products that may substitute one product or process for another. The manufacturers obviously have the expertise in the production processes and should also be have sufficient knowledge about the properties of their products and the chemicals that they consist of. The end-users usually have detailed information about the actual use of the product and the corresponding exposures. Agencies, on the other hand usually do not have the detailed information required to take individual substitution decisions. Their role is rather to create a regulatory framework that promotes the implementation of the substitution principle both by manufacturing industry and by other users of chemicals.

Many methods have been used to encourage or enforce substitution. In this chapter we will first provide a list of some major such methods, and then discuss the choice of methodology in substitution work.

6.1 Some major methods to encourage substitution.

1. Increasing the availability of toxicity data. The lack of data for chemical substances is the major obstacle to substitution. If companies have access to reliable and comparable data for alternative chemical substances (and other alternative technological methods), then they are in a much better position to choose the environmentally better alternative. As we have already mentioned, REACH takes a step in that direction, but more data is needed than what has up to now been decided in the REACH regulation.[64]

2. Increasing availability of data on the chemical composition of products. A major problem for substitution work, and for chemical risk management generally, is that many companies do not have access to adequate information about the chemical contents of the products that they use. In a Swedish study of environmental aspects of product design it was shown that in the automobile industry, neither the vehicle producers nor suppliers of components such as electronics and textiles know which if any BFR:s are used in the components. Some plastics suppliers are unwilling to disclose information on additives, such as flame retardants. They may be willing to guarantee what a product does not contain (such as a specific BFR), but not to tell what it actually contains.[65] Hence those who buy the products, perhaps to use as parts in a larger product, do not know what the product contains, only what it does not contain. This makes it difficult, to say the least, for them to work systematically with substitution. Measures that provide users of products with adequate information about their contents can therefore contribute to improved substitution work in many companies.

3. Increasing the availability of information about technical functionality. As was noted by Lohse and co-workers, "the most significant barrier for substitution appears to be the guarantee of technical properties of a product… Therefore, both manufacturers and, to a smaller extent, users of substituting chemicals tend to be conservative and 'anti-

[64] Christina Rudén and Sven Ove Hansson, "Improving REACH", *Regulatory Toxicology and Pharmacology* 44:33-42, 2006.
[65] Jennifer C Hall, *Product Design to Reduce Restricted Substances*, IIIEE Reports 2001:2, Lund, pp. 48-49 and 51.

innovative'."[66] This is of course a particularly pressing problem in cases when the technical functionality of a chemical can only be seen after some time, such as marine anti-fouling coatings, wood preservation, etc. Information about successful substitution projects can be an important way to stimulate substitution.

Several such efforts have been made. As one example, Danish authorities have created a webpage with more than 200 examples of chemical substitutions, that is continuously updated. At present, funding is sought for making a version in English.[67]

4. Helpdesk functions. The Toxic Use Reduction Act of Massachusetts has provided companies with technical help supplied by university and government experts. This is reported to have led to substantial decreases in the use of hazardous substances.[68]

5. Lists of unwanted substances. Setting up lists of hazardous substances is one of the most common strategies to encourage substitution.[69] In Toxic Use programmes in the US this method has been used to identify substances that should be avoided.[70] The Danish Environmental Agency has published a list of undesirable substances, about 1400 substances, in 2000. KemI published its Observation List in 1998, with about 250 substances. Lists prepared by citizen's groups rather than by public agencies have also had an influence in many industries, and some larger companies have produced lists of their own.[71]

Løkke favours the method used in Scandinavia of issuing lists of unwanted substances, saying: "Information on which substances the authorities regard as problematic is laid out publicly so that downstream users may try to avoid products containing these substances and push for substitution, and so that, in time, producers can work on substitution upfront. The efficiency of the list is highly dependent on the response of producers believing that listed substances may possibly be regulated, in case the voluntary phase out languishes."[72] However, as was noted by Jennifer Hall, the method also has clear limitations: "The problem with restricted substance use is not confined to a fixed number of chemicals with a negative impact. The problem is the use of a large number of poorly understood substances. Thus, substitution of one substance with another may not actually constitute a solution."[73]

6. Ban of dangerous substances. Prohibiting a substance is of course the clearest way in which a jurisdiction can clarify that the substance has to be replaced by less dangerous alternatives. However, for practical reasons this is a method that can only be used in relatively few cases.

---

[66] Joachim Lohse et al, *Substitution of Hazardous Chemicals in Products and Processes, Final Report.* Report compiled for the Directorate General Environment, Nuclear Safety and Civil Protection of the Commission of the European Communities, Hamburg 2003, p. 88.
[67] http://www.catsub.dk. See also http://www.cleanersolutions.org.
[68] Beverly Thorpe and Mark Rossi, *The Louisville Charter: Background paper for reform no. 1 of the Louisville charter for safer chemicals*. August 2005. See also www.turi.org.
[69] Joachim Lohse et al, *Substitution of Hazardous Chemicals in Products and Processes, Final Report.* Report compiled for the Directorate General Environment, Nuclear Safety and Civil Protection of the Commission of the European Communities, Hamburg 2003, p. 20.
[70] Ken Geiser, "Cleaner Production perspectives 2: integrating CP into sustainability strategies", *UNEP Industry and Environment* January 2001, pp 33-36.
[71] Joachim Lohse et al, *Substitution of Hazardous Chemicals in Products and Processes, Final Report.* Report compiled for the Directorate General Environment, Nuclear Safety and Civil Protection of the Commission of the European Communities, Hamburg 2003, p. 22. Peter Montague and Maria B Pellerano "Toxicology and environmental digital resources from and for citizen groups," *Toxicology* 157:77-88 2001.
[72] Søren Løkke, "The Precautionary Principle and Chemicals Regulation", Environ Sci Pollut Res 13:342-349, 2006, p. 348.
[73] Jennifer C Hall, "Product Design to Reduce Restricted Substances", IIIEE Reports 2001:2, Lund, p. v.

Comparative studies have been made of the general ban on trichloroethylene that was introduced in Sweden 1996 and a German regulation that instead imposed strict technical standards for equipment and emissions. One evaluation indicates that due to the large number of exceptions granted in Sweden, the reduction of emissions of trichloroethylene was much larger in Germany than in Sweden.[74] Another well-known example of substitution by prohibition is the phase-out of ozone-depleting CFCs. The success of this regulation depended on one of the major companies (DuPont) developing substitutes.[75]

7. Required substitution plans. The Massachusetts Toxic Use Reduction Act requires companies using more than 10000 pounds of listed substances to develop a plan that evaluates possible options for reductions in use. The act does not require these plans to be implemented, but the preparation of plans has nevertheless resulted in many substitutions and process changes[76].

8. Economic incentives. In some cases, a relatively small change in the structure of economic incentives can make a large difference for product choice. "In the case of flame retardants… a halogen-free circuit board may cost a few Cents more than a conventional one, while the price of the final product may be several hundred Euro. Nevertheless, due to the pressure on margins in the electronic hardware sector it is extremely difficult to implement substitution in this case."[77] The use of economic incentives to disencourage the use of hazardous or untested substances is therefore an interesting option that should be further investigated. An interesting example is the Norwegian environmental tax for pesticides, calculated so that the "basic fee" for each pesticide product is multiplied with a factor between 1 and 8 depending on the health and environmental risk.[78]

6.2 The choice of methodology

Experience shows that the market will usually not automatically ensure that substitution takes place. In general older chemicals are cheaper, since there are investment costs that have to be taken initially for the new ones.[79] There is also some open resistance to the substitution principle. Hence, the European Crop Protection Association (ECPA) has explicitly stated that it "does not support regulatory substitution".[80] In a much discussed "non-paper" that was

---

[74] Florian Birkenfeld et al, *Product ban versus risk management by setting emission and technology requirements. The effect of regulatory schemes taking the use of trichloroethylene in Sweden and Germany as an example*, Passauer Diskussionspapiere, Volkswirtschaftliche Reihe, Diskussionsbeitrag V-37-05, October 2005. See also D Slunge and T Sterner "Implementation of policy instruments for chlorinated solvents. A comparison of design standards, bans and taxes to phase out trichloroethylene", *European Environment* 11:281-296, 2001.

[75] Frans Oosterhuis, "Substitution of hazardous substances. A case study in the framework of the project 'Assessing innovation dynamics induced by environmental policy". Institute for Environmental Studies, Vrije Universiteit, Amsterdam 2006, p. 4.

[76] Brian Hyndman, *Strategies for the Reduction and Control of Environmental Carcinogens in Canada: What's Happening? What's Missing*, Prepared for Cancer Care Ontario, The Alder Group, September 28 2005, pp. 13-14.

[77] Joachim Lohse et al, *Substitution of Hazardous Chemicals in Products and Processes, Final Report.* Report compiled for the Directorate General Environment, Nuclear Safety and Civil Protection of the Commission of the European Communities, Hamburg 2003, p. 87.

[78] Joachim Lohse et al, *Substitution of Hazardous Chemicals in Products and Processes, Final Report.* Report compiled for the Directorate General Environment, Nuclear Safety and Civil Protection of the Commission of the European Communities, Hamburg 2003, p. 25.

[79] Friends of the Earth, *'Sustainable Production and Use of Chemicals' Consultation Response*, Friends of the Earth, London 1998, p. 13.

[80] European Crop Protection Association, *Revision of Directive 91/414/EEC – ECPA comments on various issues discussed at the stakeholder work-shop held on 30 January, Brussels 18 March 2004*, p. 5. http://www.ecpa.be/files/documentslive/9/13228_PositionPaper-revision-General-issues-FINAL.pdf.

assumed to express the US position in 2001, European legislative proposals were criticized and the substitution principle was called "arbitrary discrimination".[81]

AH Vorschoor and L Reijnders investigated five cases in which companies reduced their use of toxic chemicals. All five companies mentioned environmental legislation as a reason for doing this.[82] Another study comprising ten cases of substitution, showed a strong similarity between different substitution cases, namely that (actual or expected) legislation was a powerful driver.[83]

We can conclude that for the substitution principle to be efficiently implemented, regulators and public authorities have to take the lead. However, it is impossible for them to decide on each particular substitution. Public decisions on specific substitutions have to be reserved for special cases. For the vast majority of cases, the role of regulators and authorities should instead be to create incentives for substitution. We believe that most of the methods listed in section 6.1 are useful for that purpose. It is important to conduct this work in such manner that it is accessible to outcome evaluation. By systematically evaluating the effects of substitution-promoting measures, we can learn which methods are the most effective ones.[84]

[81] Joseph DiGangi, *US Intervention in EU Chemical Policy*, Environmental Health Fund, September 2003, pp. 10-11.
[82] AH Vorschoor and L Reijnders, "Toxic reduction in processes. Some practical examples", *Journal of Cleaner Production* 9:277-286, 2001.
[83] Joachim Lohse et al, *Substitution of Hazardous Chemicals in Products and Processes, Final Report*. Report compiled for the Directorate General Environment, Nuclear Safety and Civil Protection of the Commission of the European Communities, Hamburg 2003, p. ii.
[84] Sven Ove Hansson, "Praxis Relevance in Science", *Foundations of Science*, in press.

KEMI
Swedish Chemicals Agency

SWEDISH CHEMICALS AGENCY • P.O. Box 2 • 172 13 SUNDBYBERG
PHONE +46 8 519 41 100 • FAX +46 8 735 76 98 • www.kemi.se • e-mail kemi@kemi.se