ALTERNATIVE METHODS FOR EYE IRRITATION TESTING

Transition in toxicological safety testing through a multi-level perspective

Master thesis (30 ECTS)

MSc. Science & Innovation Management
Utrecht University

October 2011

Author: D. van der Zwan BSc
Cambridgelaan 735
3584DW Utrecht
d.vdzwan@gmail.com
Student number: 0340650

Supervisors: M. Kooijman MSc
Dr. E.H.M. Moors
Prof. Dr. Huub Schellekens

Page intentionally left blank

PREFACE

During the six months (and little bit) that went into preparing this report, I have had the opportunity to become familiar with field of toxicology. Relatively unbeknownst to me beforehand, it has been a journey to discover the ins and outs the safety measures that are in place for products that I use on a daily basis.

I would like to thank the people that helped me with conducting the research for this Master Thesis. First of all, I would like to thank Marlous Kooijman for all her contributing insights, feedback and general support. Her guidance has offered me not only a clear theoretical and methodological direction for the research, but more importantly, motivated me for the topic of safety-testing and stimulated my ambition to contribute to the understanding of the subject. Next I would like to thank my second-first reader, Ellen Moors for her valuable constructive feedback, especially later on during the research when I had a tendency to drown in knowledge. With a particular clarity, she considerable helped to improve the overall quality of the research, for which I am grateful.

A special thanks goes out to all the experts that have freed up valuable time for interviews, providing valuable answers to my questions and being very open about the subject. This always kept me enthusiastic about the subject, and their knowledge delivered valuable information about the complex field of toxicology.

Finally, I would like to thank all my friends and family for their unconditional support.

SUMMARY

Throughout the years, animal models have been used extensively for toxicological safety assessment. Alternative methods that do not require animals have been developed extensively, but are adopted scarcely in spite of scientific, economical and societal deficiencies of animal models. This is believed to be part of a transition problem, which refers to the structural reorientation of an entire sector, in this case the life sciences. One example where technological change did happen is in the field of eye irritation testing. The Draize eye irritation test using live rabbits has been (partially) replaced in OECD guidelines with the Isolated Chicken Eye (ICE) test and the Bovine Corneal Opacity and Permeability (BCOP) test. For this reason, the ICE and BCOP test serve as a case study to how and why they were successful in replacing the Draize test. The main research question is:

Which sociotechnical factors influenced the successful technological development and regulatory validation of the BCOP and ICE test methods for (partially) replacing the Draize eye irritation test in the field of toxicological safety testing?

The multi-level perspective (MLP) on transitions is employed as an analytical framework to analyze the case study according to the different levels (i.e. niche, regime and landscape) of the MLP. Data is obtained through literature review and interviews with twelve international experts in the field of toxicology. The results indicate that a globally operating, multi-actor network forms the basis of the safety-testing regime. It can be concluded that the safety-testing regime has undergone a transformation path: moderate landscape pressure in the form an effective anti-Draize campaign happened at a moment when the ICE and BCOP niche-innovations had not been sufficiently developed to fully substitute the Draize test. The safety-testing regime was disrupted enough for regime actors, industry and public authorities in particular, to pick up on the niche innovation. They responded by modifying the innovation activities that were needed to get organotypic methods validated for use in formal regulatory safety-assessment. From the moment that the first validation studies took place, the ICE and BCOP test have coexisted in symbiotic fashion next to the Draize test. Cumulative adjustments and reorientations in the safety-testing regime had to take place to validate the ICE and BCOP test: legislation was created that forced the development of alternatives methods. Industry in collaboration with public authorities and public research support actors started extensive validation programs, which have improved considerably through learning, by including prediction models, protocols and a tiered testing strategy. Under the landscape pressure of globalization, the EU and US saw increased collaboration that resulted in a retrospective validation study, which paved the way for US federal endorsement and OECD acceptance of the ICE and BCOP test.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AAVS	American Anti-Vivisection Society
AWF	Animal Welfare Foundation
BCOP	Bovine Corneal Opacity and Permeability test
BfR	German Federal Institute for Risk Assessment
BGA/BMBF	German Federal Health Office / Department of Research and Technology
BRD	Background Review Document
CAAT	US John Hopkins Centre for Alternatives to Animal Testing
CEC	Commission of the European Communities
CEET	Chicken Enucleated Eye Test (=ICE)
CM	Cytosensor Microphysiometer assay
COLIPA	European trade association for the cosmetic, toiletry and perfumery industry
CPSC	US Consumer Product Safety Commission
CRO	Contract Research Organization
CTFA	US Cosmetics Toiletry and Fragrance Association (now PCPC)
DG RTD	Directorate General "Research and Technological Development"
EC	European Commission
EC JRC	European Commission Joint Research Commission
EC/HO	European Commission / UK Home Office
ECETOC	European Centre for Ecotoxicology and Toxicology of Chemicals
ECHA	European Chemicals Agency
ECVAM	European Centre for the Validation of Alternative Methods
EET	Enucleated Eye Test
EPA	US Environmental Protection Agency
ERGATT	European Research Group for Alternatives in Toxicity Testing
ESAC	ECVAM's Scientific Advisory Committee
EU	European Union
FDA	US Food and Drug Administration
FHSA	US Federal Hazardous Substances Act
FL	Fluorescein Leakage assay
FRAME	Fund for Replacement of Animals in Medical Experiments
GHS	Globally Harmonized System for hazard classification and labeling
HCE	Human Corneal Epithelium model
HET-CAM	Hen's egg test on the Chorio-Allantoic Membrane assay
ICCVAM	US Interagency Coordinating Committee on Validation of Alternative Methods
ICE	Isolated Chicken Eye test
IIVS	Institute for In Vitro Sciences
IRAG	Interagency Regulatory Alternatives Group
IRE	Isolated Rabbit Eye test
IS	Innovation System
J&J	Johnson and Johnson
(M)MAS	(Modified) Maximum Average Scores
MHW/JCIA	Japanese Ministry of Health and Welfare/ Japanese Cosmetics and Toiletries Association
MLP	Multi-level perspective
NCA	National Competent Authority
NICEATM	US National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods
NIEHS	US National Institute of Environmental Health Sciences
NIH US	National Institutes of Health
NRR	Neutral Red Release assay
NVT	Nederlandse Vereniging voor Toxicologie
OECD	Organisation for Economic Co-operation and Development
PCPC	US Personal Care Products Council
(Q)SAR	(Quantitative) Structure-Activity Relationship
R&D	Research and Development

R36	Irritating to eyes
R41	Risk of serious damage to eyes
RBC	Red Blood Cell haemolysis test
REACH	EU Regulation 1907/2006 on the Registration, Evaluation, Authorisation and restriction of Chemicals
RhT	Reconstructed human Tissues
SAR	Structure-Activity Relationship
SCCS	Scientific Committee on Consumer Safety
SMI	Slug Mucosal Irritation assay
TG	Test Guideline
US	United States
USDA	United States Department of Agriculture
WTO	World Trade Organization

1. INTRODUCTION

Animal models are widely recognized for playing a key role in the safety assessment of drugs and chemicals (Jucker 2010). Virtually all drugs and cosmetics that are currently available have at some point found their way into or onto the body of animal species. Technological development has resulted in alternative methods that do not use living animals, but the adoption, and hence innovation, of alternative methods occurs more slowly (Balls and Fentem 1997). The introduction first outlines the background of animal model replacement and continues with an innovation perspective on the matter.

1.1 BACKGROUND

The practice of using animals for regulatory toxicology flourished in the first half of the 20th century as a result of the increasing role of chemicals in everyday life and the accompanying need to guard public safety from possible adverse effects. It was also at this time that small animal laboratory research established the basis for current safety testing approaches, which has left its imprint on international regulations (Hartung and Daston 2009).

SHORTCOMINGS OF TRADITIONAL ANIMAL MODELS

Today, regulatory testing accounts for over 23% of all animal use, split up between toxicology and safety assessments (8% or roughly 1 million animals) and veterinary and human medicine products safety (15.3% or roughly 1.8 million animals) (EC 2007). And while finding new medicines for human disease and assuring their safety can certainly be a justifiable cause for animal testing, the use of animals in the production and testing of consumer healthcare products is more controversial. In addition, the predictive value of animal tests is increasingly being challenged. Out of the few studies that have been undertaken to establish the predictive value of animal models (Hottendorf 1987; Olson et al. 2000; Schein et al. 1970), and taking into account the flaws of the respective methods (Matthews 2008), only a limited correlation could be established.

Meanwhile the current trends in drug development are towards complex, human specific diseases (type II diabetes, Alzheimer, rheumatoid arthritis, etc.) for which effects are difficult to mimic in animals. Many drugs in development are biopharmaceuticals, lowering the predictive value of animal studies even further because of the species specificity and immunogenicity of such drugs (Bussiere et al. 2009). This poses several problems, most obviously the possibility that a substance turns out to be harmful to humans, even after animal tests indicated otherwise, providing what is in essence a false sense of safety. The other way around it is possible for a substance to be incorrectly dismissed because of apparent toxicity in animals, while it could have been a viable drug candidate in humans.

Practically, animal studies require the sustained attention of specialized personnel. Toxicity is determined by injecting substances into animals, watching to see if they get sick, and then looking at their tissues under a microscope. “That approach is clearly quite expensive, it is time consuming and uses animals in large numbers” (Collins 2008). Besides, there are public and political discussions about the ethics of animal experimentation. Considering these scientific, economic

and societal deficiencies of animal studies, one has to state the question why there seems to be so little innovation in the replacement of animal methods for safety and efficacy testing[1].

Can it be the researcher's lack of committed effort? As early as 1959, Russel and Burch published the book 'The principles of Humane Experimental Technique'. At that time, mainly driven by ethical considerations on the use of animals for experimentation, they introduced the 3R's: replacement, reduction and refinement. They argued for the *replacement* of animals in experiments with non-conscious or non-living alternatives whenever possible. In those cases where animals have to be used, *reduction* of both the amount of animals used and the amount of tests performed should be strived for. Finally, every effort should be made to *refine* experiments to minimize the involved suffering (Russell, Burch, and Hume 1959). Despite its seemingly obvious claims, it was not until the 1980's that the 3R principles were generally acknowledged to be of importance to research methods, and finally to legislation as well. Today, many organizations are involved in the funding and development of alternative methods. In the UK, the Fund for Replacement of Animals in Medical Experiments (FRAME) promotes consideration of the ethical and scientific issues involved in the use of laboratory animals for medical research (FRAME 2010). In the Netherlands, ZonMw provides structural funding for projects that enhance the application of alternative methods (ZonMw 2011). Both the EU and the US founded official bodies (ECVAM and ICVAM respectively) for the coordination and evaluation of alternative testing methods. Companies also increasingly research and develop *in vitro* and *in silico* methods that can provide information about drug safety and efficacy (Bottini and Hartung 2009), but the validation and actual widespread use of such methods, while an integral part of their innovation, remains low.

Transition and the need for innovation

The need for innovation is emphasized in the FDA (2004) report "Innovation or stagnation: challenge and opportunity on the critical path to new medicinal products", stating that an "inability to predict (adverse drug effects) before human testing or early in clinical trials dramatically escalates costs." Consider this in the light of the ongoing sustainability problems in the life sciences sector. Over the last decades, soaring research and development costs (Baxendale et al. 2007) coupled with diminishing R&D returns (Grabowski, Vernon, and DiMasi 2002) have been driving the need for transition in the pharmaceutical industry (Hill and Langvardt 2010). The FDA urges the scientific community to move forward in finding innovative investigational methodologies for drug development (FDA 2004). Certainly, one way to do this is to look at non-animal research techniques which can be extrapolated directly to actual human endpoints, providing a greater predictive power and lowering the overall cost of drug development[2] (Schaeffer 2005).

[1] The term 'alternative methods' in the remainder of this document refers to alternative, non-animal methods for safety and efficacy testing of pharmaceuticals.

[2] This is not necessarily true for all non-animal methods, however developments such as 3D-computer modeling systems and other *in silicio* methods in predictive toxicology have demonstrated such promises in the past (FDA, 2004).

Despite the endorsement, alternative methods for regulatory safety testing are adopted scarcely. Reasons as to why this is the case range from technological immaturity (Worth and Balls 2002), to a culture that puts too much emphasis on 'safety first' out of precaution and liability concerns (Bottini and Hartung 2009), to difficulties with the validation of alternative methods (Bruner et al. 1996). As part of a larger research project that investigates the transition to sustainable drug development, the objective of this project is obtain insight in the cases where technological change did happen and alternative methods have been successfully developed and validated. In order to do this, we will take a side step to the general area of toxicology and safety assessment of chemicals[3]. Reason for this is the lack of examples of successful replacements of animal models that are used specifically for drug development (Coleman 2011). Instead, the field of toxicological safety-assessment has seen such an example[4] with the recent partial replacement of the Draize eye irritation test with two *ex vivo* methods: the Bovine Corneal Opacity and Permeability assay (BCOP) and the Isolated Chicken Eye (ICE) test (See Box 1 for a short background). The area of eye irritation testing can be considered a pioneer in the development and validation[5] of alternative[6], *in vitro* methods to replace an animal model (Eskes 2010). It was one of the first areas that saw dedicated effort in the development of alternative methods, and one of the few for which OECD acceptance of alternatives is reached, the highest formal degree of acceptance possible. For these reasons, it eye

BOX 1: BACKGROUND OF THE DRAIZE TEST

The Draize Eye irritation test is named after the American toxicologist John Henry Draize, who developed the method for determining whether a substance is capable of producing ocular irritation or severe eye irritation. The original test consisted of nine rabbits, the treated eyes of two groups of three rabbits each were washed with 20ml water 2 or 4 seconds after instillation of a substance. In a third group of rabbits, the treated eyes remained unwashed. The Food and Drug Administration (FDA) incorporated the method in 1964 for regulations dealing with pharmaceuticals and cosmetics. In 1972 the test was also introduced in the guidelines of the Consumer Product and Safety Commission (CPSC) for the testing of household products. Eventually the testing procedure was reduced to three rabbits.

Eye irritation has been a pioneer in the field for the development, evaluation and validation of *in vitro* methods to replace the Draize test. Major multi-laboratory research efforts where undertaken as early as in the 1990's, resulting in around 30 *in vitro* alternative test methods. It took until 2004 before a thorough review could be carried out to advance the validation of alternatives, and the most promising alternatives to replace the animal test were identified. Today, two test have been approved to (partially) replace the Draize test: the Bovine Corneal Opacity and Permeability assay (BCOP) and the Isolated Chicked Eye (ICE) test method (OECD, 2009). Two additional cell-function based *in vitro* assays have been recently endorsed as scientifically valid by the ECVAM's Scientific Advisory Commission (ESAC).

[3] Including medicinal products but also cosmetics and industrial chemicals

[4] And even in this example, replacement is only partial, demonstrating the difficulties in replacing animal models.

[5] (Technological) development and regulatory validation are considered to be synonymous with 'replacement', because once a method for safety testing has been validated by regulatory authorities, it is more or less automatically adopted by companies and research units.

[6] The terminology used to describe methods that do not require the use of living animals varies widely between publication. 'Alternative methods' is used often by people in the field, but can be vague on its own, when used out of context. *In vitro* tests/methods is a common descriptor and provides a better technical meaning, however the definition of *in vitro* formally does not include tests that use functional organs that have been removed from the intact organism, such as the ICE and BCOP test, which are technically *ex vivo* methods. For a similar reason, 'non-animal methods' could also be confusing. For readability purposes, this report adheres to the use of 'alternative methods' as much as possible.

irritation testing serves as a case study in this research, unraveling how and why partial replacement of the Draize test was successful, paving the way for innovations in other alternative methods.

1.2 Aim and research questions

The central aim of this research is to understand the cases of successful substitution of animal models by other alternative means of toxicological safety-assessment, in particular the replacement of the Draize test with the ICE and BCOP test. From an innovation studies point of view, the (lack of) adoption of alternative methods in drug development can be understood as a broader transition problem. Transition refers to a structural reorientation of an entire sector, in this case the life-sciences. It entails a complex, large scale and long-term transformation that encompasses simultaneous change at different levels (i.e. the micro and macro level of the economy) (Rotmans 2007). Consequently, transitions face many challenges as the structural reorientation is by definition multi-actor, multi-dimensional and multi-level (Elzen and Wieczorek 2005). In the emergence and progression of transition processes, innovation is considered a *critical factor* (Coenen and Lopez 2008; Gerlach 2001; Hekkert et al. 2007). These authors all emphasize the need for innovation that goes beyond changes in technology and draws upon sociotechnical change, involving substitution of a technology, as well as changes in other societal elements. In safety-assessment for instance, regulations, user practices, scientific inquiry, and safety expectations, are all aligned to the currently established practice and new technologies (i.e. alternative methods) often face a miss-match (Freeman and Perez 2000), requiring a structural reconfiguration to become successful.

Theoretical contributions in the field of transitions build upon two major approaches that both consist of detailed analyses of innovation processes: the innovation system approach (see for example (Edquist and Lundvall 1993; Jacobsson 2002)) and the multi-level perspective (see for example (Geels 2004; Rotmans 2007). Both approaches show similarities in theoretical grounds. They highlight the importance of actors, activities, networks and learning processes together with the crucial role of institutions for successful innovation processes. Also they acknowledge phenomena such as path dependency, lock-in, interdependence, non-linearity and coupled dynamics (Markard and Truffer 2008).

Innovation systems (IS) are composed of networks of actors and institutions that develop, diffuse and use innovations. Different strands have emerged in the literature e.g. sectoral IS (Malerba 2002), national IS (Lundvall 1992); and technological IS (Carlsson and Stankiewicz 1991; Hekkert et al. 2007)). Innovation systems can be compared with regard to the functions they fulfill (Bergek et al. 2008; Hekkert et al. 2007). In the context of replacing animal models however, application of an IS approach would be problematic because the functions that provide the IS approach with its analytical power seem to be based on the assumption of (free) market forces and would lose some of their strength in analyzing a heavily regulated toxicology system where market forces are weak (Nichols et al. 2004).

The multi-level perspective (MLP) has been developed as an appreciative theory to understand the complex dynamics of sociotechnical change in transitions and integrates findings from different strands of literature, mainly evolutionary economics and technology studies (Kemp 1994; Schot, Hoogma, and Elzen 1994). It provides analytical and heuristic concepts of sociotechnical regimes,

niches and landscapes that can be used to view sociotechnical change as a multi-level process. The strength of the MLP is that innovation and transition processes can be explained by the interplay of developments at the regime level, and destabilizing landscape pressures combined with the emergence of innovation at the niche level (Markard and Truffer 2008). The regime level for instance, comprises[7] a policy, science and technology dimension, the interplay between which arguably all have been important in replacing the Draize test (Eskes 2010). An innovation systems approach would regard this success as the consequence of the performance of the corresponding innovation system according to the systems functions[8], which significantly narrows the analysis and leaves very little room for external influences that for example hinder the innovation process (which simply would be treated as blocking mechanisms, while they may have been much more than that, e.g. the result of strategic intervention of incumbent actors (Markard and Truffer 2008)). In understanding how the Draize test got replaced, the MLP as a framework leaves room for process explanations of such influences in terms of interrelatedness of dimensions and therefore is more suitable to analyze the adoption of the ICE and BCOP test in the context of a transition in toxicological safety-assessment.

To summarize, this research considers the ongoing desire to use alternative methods as part of a transition in toxicological safety-assessment and employs the multi-level perspective on transitions to analyze the pioneering case of the Draize eye irritation test. The following main and sub-research questions are formulated:

Which sociotechnical factors influenced the successful technological development and regulatory validation of the BCOP and ICE test methods for (partially) replacing the Draize eye irritation test in the field of toxicological safety testing?

- What technologies are used in the BCOP and ICE test methods?
- How were the BCOP and ICE test methods developed and validated?
- Which barriers had to be overcome for the partial replacement of the Draize test?
- Which insights can be derived from this case that could facilitate the transition towards alternative methods through implementation of other alternative methods?

1.3 Scope and delineation

First of all, this research is concerned with the replacement of the Draize test in the context of its application in toxicology testing for the safety-assessment of substances. The test is mainly used for safety-assessment of chemical substances (e.g. for use in cosmetics). Its pharmaceutical appliances are limited to the in-house screening of intermediate compounds used in the production process of drugs, in order to establish safety measures for 'occupational hazard' (i.e. the protection of employees). Geographically, the research is delineated to the EU and the US, which are the two regions that are most influential in the OECD decision process and most of the validation efforts are undertaken here (Eskes 2010). However, the OECD includes other member states and when relevant to the case study, developments in other geographical areas will be mentioned.

[7] But is not limited to
[8] Functions of which the applicability by themselves in this case is already questionable, as explained before.

Second, transitions involve wide-ranging sectorial changes. The transition in toxicology testing is broader than the replacement of the Draize test with the ICE and BCOP tests: it includes the development for other areas of toxicology testing (i.e. skin irritation, reprotoxicity, etc.). For the scope of this research however, the ICE and BCOP test serve as the unit of analysis, because more insight is desired in why and how these tests in particular have been accepted. As a practical delineation, the transition in toxicology testing will be discussed to the extent that it is has influenced, or has been influenced by, developments for alternative methods for eye irritation testing. Other areas of toxicity testing lie beyond the scope of this research.

Societal relevance

The ongoing scientific discussion about the predictive value of animal studies (Matthews 2008), together with the ongoing public debate about the ethics and need for animal studies (Fraser 1999) and the joint efforts of private firms and regulatory bodies to find alternative approaches to safety testing, all indicate a desire to improve the current practice of safety evaluation. This is all the more illustrated by recent funding that has been supplied by Dutch ministry of Health, Welfare and Sport as part of a plan to reduce animal use over the period 2011-2021 (VWS 2011). Despite these motivations and efforts, the sociotechnical change in this field has progressed slowly over the past decades. This research hopes to contribute to this change for the better, by analyzing a success story and providing 'lessons learned' for the adoption of future alternative methods.

Scientific relevance

The scientific approach of this research is novel in the sense that the multi-level perspective – and transition theories in general – are hardly ever applied to problems in the life sciences sector. In doing so, the empirical evidence on the multi-level perspective will be tested in a new field of application. The vast majority of publications that use transition theories are concerned with environmental and transportation problems (Geels 2006a; Geels 2006b; Geels 2006c; Genus and Coles 2008). Several characteristics of the life sciences sector – and toxicology testing in particular, differ considerably from those in the energy and transport sectors, which could have implications on a theoretical level. For instance, toxicology testing for safety-assessment is a branch that largely exists because of regulatory requirements for product safety. As a result, the demand for toxicology tests is created by –and subject to- strong regulations, more so than in other 'traditional' sectors with a free market (Bottini and Hartung 2009). In addition, the market for toxicology tests is rather 'artificial': there is no natural demand and the tests are developed and produced by the same companies that (are obligated to) use them. Finally, safety-assessment regulations are in place to safeguard the wellbeing of humans, resulting in a tendency (of legislators) to act on the basis of a precautionary principle[9] when decisions have to be made (Hansson and Rudén 2003).

[9] Meaning that, if an action or policy has a suspected risk of causing harm to the public, in the absence of scientific consensus on the harmfulness, the burden of proof that it is not harmful lies with those taking the action and the 'default' operation has to presume harmfulness.

1.4 Instructions for further reading

This report is structures as follows. Chapter two explains in detail the theory of the multi-level perspective (MLP) that was shortly explained in the introduction. Theoretical concepts of the MLP are operationalized in chapter three on methodology. This chapter further includes the research design and protocol that have been used to construct the case study. Chapter four provides a detailed, historical case description of the replacement of the Draize test with the ICE and BCOP test. The focus in the case description lies on accurate factual representation. Analysis of the case is divided between chapter five and six. Chapter five provides an analysis of the multi-actor network as proposed by the MLP (figure 5.1 and 2.2 respectively). Chapter six draws on the case description and data from expert interviews to analyze the replacement of the Draize test through the three levels from the MLP: niche, socio-technical regime and landscape. Figure 6.3 provides an overview of the main findings. Conclusions are drawn in chapter 7, which also answers the main research questions. Implications from theory and recommendations are provided in the discussion.

2. THEORY

This chapter elaborates on the innovation theory used in the research. The multi-level perspective by Geels and Schot (2007) is used to analyze the uptake of the ICE and BCOP test into the safety-testing regime. It serves as the basis for conceptualizing technological innovation in terms of socio-technical transition, thereby acknowledging its multi-facetted, multi-actor nature in a societal context (Geels 2004). Further operationalization of the concepts mentioned in this chapter is part of the methodology chapter.

2.1 THE MULTI-LEVEL PERSPECTIVE ON SOCIO-TECHNICAL TRANSITIONS

The term 'multi-level perspective' refers to three different interplaying levels it distinguishes in order to explain technological transitions: the niche level (micro), the socio-technical level (meso) and the landscape level (macro)(Geels 2005). These levels are not ontological descriptions of 'reality', but analytical and heuristic concepts to understand the complex dynamics of socio-technical change.

At the core of the framework lies the socio-technical regime, a coherent, highly interrelated and stable structure at the meso-level, characterized by established products and technologies, user practices, expectations, stocks of knowledge, norms, regulations, etc. (Markard and Truffer 2008). In evolutionary terms, this regime represents the 'selection environment' for technological change in a technological field, and provides barriers for the diffusion of an innovation if it is not aligned with the regime. However, the niche level is an environment outside of the main regime and thereby provides protection from early selection pressures (Schot and Geels 2008). This enables the development of new (technological) variety that can lead to innovation. The macro-level, or so-called landscape, then forms the background against which the micro- and meso-developments occur. It includes factors that are hard to change and are outside the influence of regime actors, but at the same time are capable of influencing a regime when a change does occur (Geels 2005). Such forces then destabilize an existing regime, creating the opportunity for niche innovations to break through. Such sociotechnical change is called a transition and involves changes in technologies and technical artifacts, as well as in user practices, markets, policies, industrial structures and supporting infrastructures(Geels 2002). The next section examines the various aspects of the MLP as shown in figure 2.1.

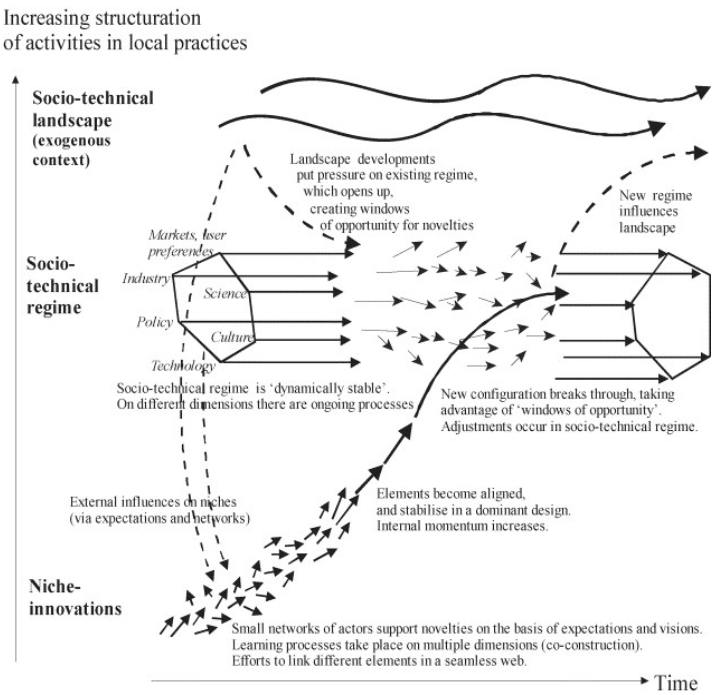

FIGURE 2.1: OVERVIEW OF THE MULTI-LEVEL PERSPECTIVE (SOURCE: GEELS AND SCHOT 2007)

THE MACRO LEVEL: A SOCIO-TECHNICAL LANDSCAPE

The macro-level is formed by the sociotechnical landscape. Here the focus lies on dynamic aspects of the wider exogenous environment, which affect socio-technical development. The metaphor 'landscape' is used because of the literal connotation of relative 'hardness ' and to include the material aspects of society. In the literature, the landscape has been defined as a: "*set of heterogeneous factors, such as oil prices, economic growth, wars, emigration, broad political coalitions, cultural and normative values, environmental problems*" (Geels 2002).In a more general way, Markard and Truffer (2008) regard the landscape as a set of residual factors that have an impact on innovation and production processes without being influenced by the outcome of innovation processes on a short to mid-term basis. Landscapes are beyond the direct influence of actors and cannot be changed at will. Empirically this means that broad, structural developments such as recessions and general societal norms and values regarding the welfare of animals can be taken into account.

THE MESO LEVEL: A SOCIO-TECHNICAL REGIME

Socio-technical regimes as used within the MLP are an elaboration on the 'technological regime' as proposed by Nelson and Winter (1982). Rip and Kemp widened the technological regime concept, defining it with the social category of rules: '*the rule-set or grammar embedded in a complex of engineering practices, production process, technologies, product characteristics, skills and procedures, ways of handling relevant artifacts and persons, ways of defining problems - all of them embedded in institutions and infrastructures*' (1998). The same definition is used in this research. The sociotechnical regime concept explicates the broader influence of social processes on technological change. It refers to a set of rules that are carried and shared by different actors that fulfill a societal function (Geels 2005). The sociotechnical regime consists of the (i) multi-actor network, (ii) regime dimensions[10] and (iii) their rules.

Figure 2.2 gives an overview of the multi-actor network in a socio-technical regime. These, in combination with the interaction and alignment activities the actors have amongst each other are the socio-institutional dynamics.

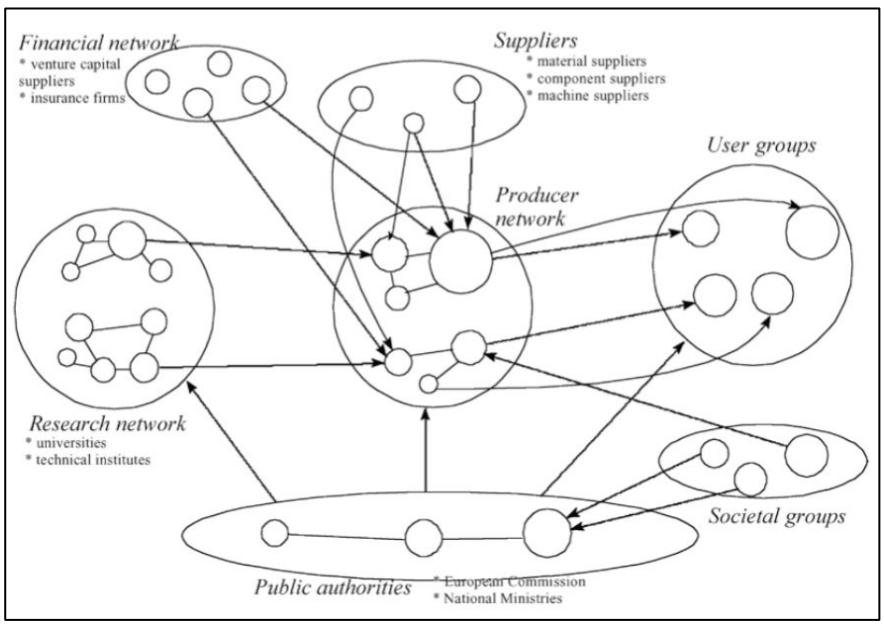

FIGURE 2.2: THE MULTI-ACTOR NETWORK (SOURCE: GEELS AND SCHOT 2007)

The socio-technical regime's multi-dimensionality can be characterized by the alignment of ongoing co-constructive and co-

[10] Each dimension can in fact be considered a different regime (Geels, 2004). The socio-technical regime is therefore appropriately called 'a patchwork of regimes'. However, for clarity purposes this research refers to the sub-regimes as dimensions.

evolutionary processes on multiple dimensions. The dimensions are technology, science, policy, industry, socio-cultural and markets/users (Wanningen 2009). In terms of the transition in toxicological safety testing, this implies that the regime of animal-based tests can be explained in terms of interrelatedness and dynamics between these dimensions. The uptake of alternative methods is viewed as the destabilization of these dynamics, that results in sociotechnical change (i.e. the relevant changes in the dimensions of technology, science, policy, culture and markets/users).

Rules are both the linkages within regimes, and can explain the alignment of activities of different groups (Geels 2007). In this sense, they do not differ from 'institutions' in other innovation approaches (Pierick and van Mil 2009). Analytically, rules can be grouped in three types of dimensions: formal, normative and cognitive (Geels 2007). The formal dimension refers to explicit rules, which constrain behavior. Normative rules are often highlighted by traditional sociologists (Durkheim 1997; Parsons 1949) and include values, norms, and expectations. Cognitive rules constitute the nature of reality and the frames through which meaning or sense is made (Geels 2007).

The micro level: technological niches

The micro-level is formed by niches, which form the locus for variations that can lead to innovations. Because the performance of radical novelties is initially low, they emerge in 'protective spaces', which shield them from mainstream market selection. Niches thus act as 'incubation rooms' for novelties (Geels 2004). "*A niche can be defined as a discrete application domain (habitat) where actors are prepared to work with specific functionalities, accept such teething problems as higher costs, and are willing to invest in improvements of new technology and the development of new markets*"(Hoogma, Kemp, and Schot 2002). While novel technologies or products tend to be the focus in innovation study context, it must be noted that niches may also refer to old technologies which have existed for quite some time, and have settled in a relatively stable niche environment (Markard and Truffer 2008). A distinction can be made between two particular kinds of niches depending on *how its particular selection environment comes about*: market niches and technological niches. Market niches are the more naturally occurring phenomenon, they emerge due to particular application contexts or consumer preferences that significantly deviate from the usual ones. Technological niches are created more deliberately by actors and are supported by specific institutions, including entrepreneurs, policy-makers and outsiders of a regime. The potential advantages of regimes are not shared among the regime actors, because they are still uncertain (Hoogma, Kemp, and Schot 2002).

As far as structure and texture is concerned, niches do show similarities with regimes. With regard to the aggregation level and stability however, they are very different. According to Geels and Schot (2007 p.402): "*. . . technological niches and sociotechnical regimes are similar kinds of structures, although different in size and stability . . . Both niches and regimes have the character of organizational fields (community of interacting groups). For regimes, these communities are large and stable, while for niches they are small and unstable. Both niche and regime communities share certain rules that coordinate action. For regimes these rules are stable and well-articulated; for niche innovations, they are unstable and 'in the making'.*"

The origins of the BCOP and ICE test methods for eye irritation ultimately can be traced back to a niche, this forms the starting point for our empirical case study. Analysis of this niche provides part of the explanation for the subsequent uptake in the socio-technical regime.

2.2 Interplay between the macro, meso and micro level

According to Geels (Geels 2005), the logic of the three levels is that they provide different kinds of coordination and structuralization of activities in local practices. The key point here, is that transitions come about through the *interplay* between the *dynamics* at *multiple levels*. The meso level of socio-technical regimes embody the aspects of how to use, produce and regulate incumbent technologies. In niches, problems or shortcomings of the existing regime are circumvented by novelties and learning takes place of *how* to use, produce and regulate *new* technologies. Niches are supported in the hope that their novelties will eventually be used in a regime, or replace it. This is difficult because the existing regime is embedded (institutionally, economically, culturally, organizationally, etc.). The socio-technical landscape sustains the current regime and niches, but can exert pressure on the existing regime, resulting in a window of opportunity for niches to break through. Geels and Schot (2005) identified four transition pathways, based on empirical research. These differ in the kind and timing of multi-level interactions: transformation, de-alignment and re-alignment, technological substitution and reconfiguration. These pathways are ideal types and combinations of pathways can be identified in a single transition. Table 2.1 gives an overview of pathways. The next chapter explains how the theoretical concepts from this chapter are used in the research.

Table 2.1: Typology of transition pathways (Source: Adapted from Geels and Schot 2007)

Transformation	Moderate landscape pressure leads to a need for change in the regime. Because niche technologies are in early stage of development, they cannot take full advantage of landscape developments yet. Nonetheless, regime actors are able to address the pressure by adopting symbiotic elements of niche technologies.
De-alignment and re-alignment	Landscape pressure is sudden and diverse. Regime actors lose faith in current technology, but no niche technology has developed well enough to be a replacement. A prolonged period of competition between niche technologies emerges, after which one gains momentum and replaces the regime technology.
Technological substitution	Niche technology has developed fully, but is kept at bay by a powerful regime. A sudden shock of landscape pressure destabilizes the regime, after which the niche technology breaks through.
Reconfiguration	As in the transformation pathway, symbiotic elements of niche technologies are adopted to address moderate landscape pressure. However, in this case, subsequent adoptions lead to changes in the basic architecture of the system. Besides technical changes, this may also lead to changes in user practices, perceptions, and search heuristics (i.e. cognitive, normative, and formal rules).

3. METHODOLOGY

This chapter explains the methods used in this research. Let us first put the multi-level perspective into perspective. As becomes clear from the chapter on theory, the MLP encompasses a vast amount of concepts and insights, mostly rather abstract generalizations from case studies. It does not pretend to provide an analytical blueprint for successful transitions or predictions of reality (Geels 2005). Instead it lies down a framework to describe transitions in terms of interrelated, complex processes on multiple levels. This has as implication that the MLP is mostly suited for ex-post description and analysis (Pierick and van Mil 2009), and therefore, that is how it is used in this research.

3.1 RESEARCH DESIGN

For the transition towards alternative methods in toxicity testing, this research concentrates on the case of the Draize eye irritation test, which has been partially substituted by the Bovine Corneal Opacity and Permeability assay (BCOP) and the Isolated Chicken Eye (ICE) test method. This implies a unit of analysis at the technological niche level (namely the ICE and BCOP tests) that can be traced all the way into the uptake of the regime.

An explorative, single case study methodology is used. A case study protocol is warranted when researching a broad topic, with the inclusion of contextual conditions and the reliance on multiple sources of evidence (Yin 1993). The nature of sociotechnical transitions according to the MLP; unpredictable and influenced by complex processes on multiple levels, requires research that includes the criteria proposed by Yin.

This research provides an in-depth, longitudinal examination of the technological development and regulatory validation of the ICE and BCOP test for eye irritation as a single case study. Even though the main research question is empirical in nature, it is not known beforehand exactly what information is required and inquiry can only take place along the concepts as defined in the MLP (and as described in the research section). Ex-post description of the development and validation of the BCOP and ICE tests then allows us to generate new insights on the transition towards alternative methods. This also warrants an explorative approach, as not to exclude potentially relevant developments in the history of these tests. The theoretical framework requires mainly qualitative information on the development and validation on the aforementioned tests (i.e. how and why did it happen).

Data comes from literature sources, attendance of a symposium from the Dutch Toxicology Association (NVT) and interviews with eleven experts in the field of toxicology testing. The results are analyzed qualitatively by structuring the information from the interviews to the concepts of the MLP. This forms the basis for a detailed case description and successive analysis along the MLP's multi-actor network with niche and regime dynamics.

Figure 3.1 displays a schematic overview of the research steps that are taken to form conclusions.

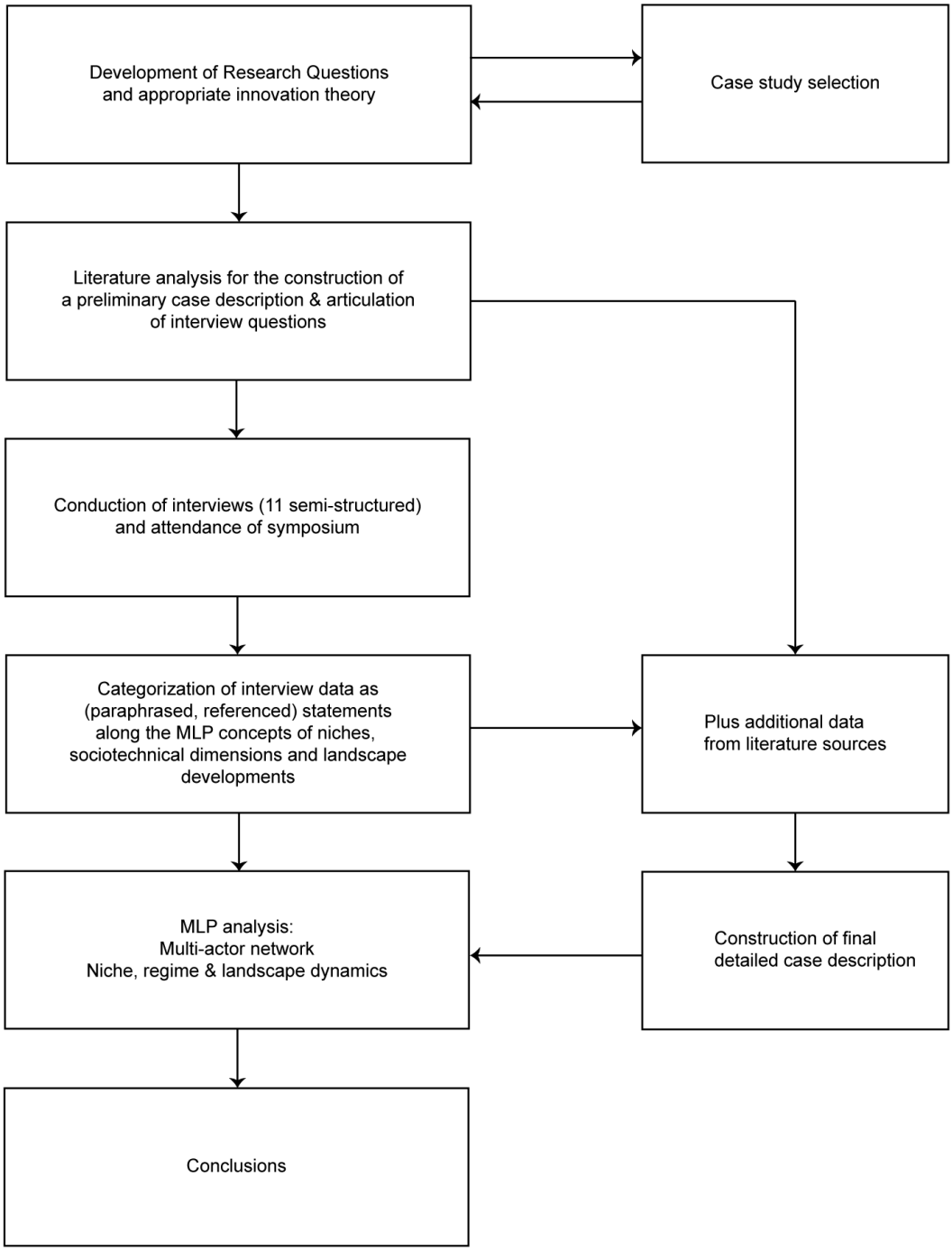

FIGURE 3.1: RESEARCH STEPS UNDERTAKEN FOR THIS PROJECT

3.2 Operationalization

Throughout their work, Geels and Schot (2007; 2008) emphasize that the MLP does not provide a systemic way of how an analyst should go about making the transition from the conceptual to the empirical level, i.e. the case study at hand. The demarcation of the empirical level and subsequent operationalization are dependent of the subject in question. In terms of the main concepts that are used theoretically, this requires to ask the question 'what data is necessary to sufficiently cover the main concepts of the MLP for a satisfactory analysis?' To this end, the operationalization serves as a guideline for the formulation of interview questions and as a handle for literature analysis, but not as an exhaustive list of every characteristic that is encompassed by the MLP.

The sociotechnical regime

As discussed in the theory chapter, a sociotechnical regime consists of *actors* and *rules*. For this research, all *relevant actor-groups in the multi-actor network* are identified and positioned with notion to the *rules*[11] that are responsible for the coordination and structuration of activities. The regime in question is named the 'safety-testing regime', which in the remainder of this research refers to all actor groups and rules pertaining to the development, production, validation, use and regulation of tests for the assessment of possible adverse effects of chemicals on living organisms.

The general structure of the multi-actor network has already been depicted in figure 2.2 above, and serves as a template for the identification of actors that are involved in the science, industry, policy, market, or technology-related activities regarding toxicological safety testing for eye irritation. Actors can then be grouped on the basis of their general roles and activities (i.e. performing eye irritation tests, legislating toxicity requirements, etc.) to shape the multi-actor network for toxicity testing.

The relevant actors in case of toxicity testing have been identified alongside this general depiction, and section 5.1 analyzes the multi-actor network.

The sociotechnical regime further consists of six dimensions (Geels 2010) of the socio-technical regime (technology, science, industry, policy, socio-cultural, and market/user preferences). Regimes themselves are made up of rules that feature three dimensions (formal, normative and cognitive). By crossing the dimensions of the socio-technical regime with the dimensions of rules, we get an analytical tool to describe the different regimes.

Again, the rules mentioned in table 1 do not provide an exhaustive list of all possible rules, but do provide enough input for semi-structured interviews, where possible different rules can be identified through open-ended questions, once more stressing the explorative character of this research. The interview questions are operationalized on the basis[12] of table 3.1, and are added as appendix A.

[11] To conceptualize coordination between actors, the term 'institutions' is often used in innovation research, however these can be confused with the (non-market) organizations, hence the term 'rules'. With regard their description of coordination between actors, the two terms are the same.

[12] The explicit distinction between formal, normative and cognitive rules is discarded because it serves no further analytical purpose after 'complete' data collection, this is explained further in the discussion.

TABLE 3.1: EXAMPLES OF RULES PER DIMENSION (SOURCE: ADAPTED FROM GEELS, 2004)

	FORMAL	NORMATIVE	COGNITIVE
TECHNOLOGY	Technical standards, product specifications, functional requirements (articulated by customers or marketing departments), accounting rules to establish profitability for R&D research, expected capital return rate for investments, R&D subsidies	Companies own sense of itself, authority structures in technical communities or firms, testing procedures.	Search heuristics, routines, guiding principles, technological guideposts, problem solving strategies
SCIENCE	Formal research programs (in research groups, government), professional boundaries, rules for government subsidies	Review procedure for publications, norms for citation, academic values and norms	Paradigms, exemplars, criteria and methods of knowledge production
INDUSTRY	Formal administration practices, product specifications, formal mission and vision statements	Companies sense of self, safety procedures, product testing practices	Routines, problem solving strategies, beliefs about the markets
POLICY	Administrative regulations and procedures which structure the legislative process, formal regulations of technology, subsidy programs, procurement programs	Policy goals, interaction patterns between industry and government, institutional commitment to existing systems.	Ideas about the effectiveness of instruments, guiding principles (e.g. liberalisation), problem agendas.
SOCIO-CULTURAL	Rules which structure the spread of information production of cultural symbols (e.g. media laws)	Cultural values in society or sectors, ways in which users interact with firms.	Symbolic meanings of technologies, ideas about impacts, cultural categories.
USERS, MARKETS AND DISTRIBUTIONS NETWORKS	Construction of markets through laws and rules, property rights, product quality laws, liability rules, market subsidies, tax credits to users, competition rules, safety requirements.	Interlocking role relationships between users and firms, mutual perceptions and expectations.	User practices, user preferences, user competencies, interpretation of functionalities of technologies, beliefs about the efficiency of (free) markets, perceptions of what 'the market' wants (i.e. selection criteria, user preferences).

Niche-innovations

Niches are structurally similar to the socio-technical regime, featuring a (small) actor-network that is bound by rules. Methods such as the BCOP and ICE test are technologically categorized as 'organotypic methods' (i.e. using organs of deceased animals to perform a test) and this niche serves as our unit of analysis. Niches provide protective spaces that guard the technology from selection criteria in the regime. Niche innovations thrust on the basis of expectations and visions, these need to be identified, along with possible learning processes that enhance the technology or facilitate its uptake into the regime. Questions on expectations and visions are part of the interview protocol in Appendix A.

The sociotechnical landscape

According to the definition of in section 2.1, the sociotechnical landscape includes a set of 'heterogeneous factors' ranging from economic growth to environmental problems. Analyzing the factors mentioned in interviews and literature and comparing them to the rules in table 3.1 achieves this. When a factor is influential, but cannot be identified as part of the regime, it is considered a heterogeneous factor in the sociotechnical landscape. The most important criterion for this decision is landscape factors cannot directly by influenced by regime actors.

3.3 Data gathering

The central question as posed in the introduction is best answered by empirical research and data analysis. The results of this research are based upon literature study and findings from semi-structured interviews.

First, a preliminary case description has been constructed through literature study of scientific articles, government reports, regulatory guidelines and legislation. This provided the background information that was necessary for formulation of the interview questions and provided a timeline for the sequence of events that occurred because of – and in order to, develop and validate alternative methods for eye irritation. The following search string was used in Google Scholar to identify relevant articles: ("eye irritation") AND (BCOP OR ICE OR "bovine corneal" OR "isolated chicken") AND (history OR development OR evaluation OR review). However, yielding 1000 articles, the results were narrowed further by only including articles that fulfilled at least one of the following criteria:

- Provided background information on *in vitro* alternatives for eye irritation testing as main objective of the article.
- Described the development of either the BCOP or ICE test.
- Documented a validation study in which either of the two tests was included

This condensed the amount of articles to just fewer than 60. The resulting preliminary case description identified the expected actor groups, important events and regulations, and served as a basis to define the roles of the actor groups to provide input for the Traide model.

Second, the data from the literature analysis was used in the interviews for (1) filling in the gaps of knowledge, especially relating to interactions between actors and 'why' and 'how' questions

relating to development; and (2) validating the original findings. The interview protocol was semi-structured and designed in such a way that it could identify relevant rules per dimension of the sociotechnical framework, and provided the opportunity to confirm them when the literature study suggested relevance. This is achieved by posing open-ended questions (What, in your opinion, has been important for X? > Is it correct that Y has played a role in X and if so, how?) The interview protocol is added in appendix A.

Interview subjects have been selected on the basis of actor groups, with the inclusion of at least 1 interviewee per actor group. Several subjects have worked at different organizations throughout the years, and thus were able to comment from different perspectives. A list of interviewees is added in appendix C.

3.4 Data analysis

Since the interviews were explorative, they delivered a wide variety of data, calling for a comprehensive analysis before any conclusions can be drawn. Interview recordings were played for a second time, while formatting a comprehensive table, resembling table 3.1, with statements from the experts that related to niche or the regime dimensions. This increased the structuring of the interview results considerably, the table that includes the results is added as appendix B. Simultaneously an event database was updated in Excel, whenever (1) the interviewee mentioned an event that in his/her opinion was relevant to the research or the question posed and (2) could be linked to a specific date and (3) could be considered a creation or change in terms of the *rules per dimensions* as discussed in the operationalization.

Analysis was performed interactively with the interview results and existing data, i.e. the preliminary case description was constantly updated to reflect new insights from interviews, and later interviews included questions to confirm statements from earlier interviews. This interactive data analysis led to a story line that serves as an historical case description of the ICE and BCOP tests, complemented with an analysis in terms of the MLP, explaining the multi-actor network and sociotechnical factors that have influenced the eventual OECD acceptance of the ICE and BCOP test.

3.4 Quality criteria

Several quality aspects of the case study should be kept in mind in performing the research. These are the construct validity, the internal validity, the external validity, and the reliability of the results. The first test, the construct validity, is the most problematic in case study research. This test is about the correctness of the operational measures for the concepts being studied (Yin, 1993). The use of invalid or incomplete concepts can lead to the acquisition of incorrect data. Using multiple sources of evidence minimizes this problem. Conducting multiple interviews and using documentation as means for data collection maximizes the variety of sources. Furthermore, the construct validity could be increased by a review of the draft data by the experts (Yin, 1993), which for instance was done for an overview of regulations.

The internal validity tests the correctness of the supposed causal relations (Yin 1993). The MLP does not predict any causal relations, but provides process explanations that link sequences of

events to explain phenomena. Still, the effects of event sequences will have to be verified by triangulation to be considered robust.

The third test, the external validity, deals with the generalization of the study's findings beyond the immediate case study. Case studies rely on analytical generalization; this has the objective to generalize a particular set of results, from the case study, to some broader theory (Yin, 1993). It is not the main aim of this research to generalize findings to the MLP framework, however it might provide some specific findings as to the applicability of the MLP for transitions in life sciences. For further certainty, multiple case studies would have to be performed, that falls outside the scope of this research.

The final criterion, the reliability, minimizes the errors and biases in a study. The objective of this test is to be sure that when the same case study would be repeated, following the same procedures as described in the research, the same findings would be obtained and the same conclusions would be drawn (Yin, 1993). In order to make the repetition of the case study possible, the procedures are documented in this methodology chapter and the interview protocol is added in appendix A.

4. Case description: Replacement of the Draize test

Dozens of potential alternative methods have been invented for the Draize test throughout the years. This chapter contains a rich case description on the history of eye irritation testing, the development of possible replacements with a focus on the ICE and BCOP tests, and the dedicated efforts that eventually led to regulatory acceptance on an OECD level. The emphasis here lies on readability and correct factual presentation that creates an understanding of the events that had to take place to get where we are right now. Further theoretical analysis of the case study is part of the next chapter.

4.1 History of eye irritation testing [~ 1944]

Testing for eye irritation developed during the 20th century as the pharmaceutical industry grew, and drugs were found to affect the eye. Experimental methods of Claude Bernard (1813-1878) (Miller 1935) initiated the use of animals for understanding the eye's responses to external substances. The incentive for testing was not to screen medications for toxicity, but to identify chemicals that could harm and blind. The first signs of institutionalized testing arose from another phenomenon: chemical warfare. World war I saw the first large-scale release of poisonous gas. Dichloroethylsulfide, also known as Kampfstoffe in Germany or mustard gas in the UK and US, is one of warfare gasses most destructive to the eyes. Within minutes to hours after exposure, it causes epithelial sloughing of the conjunctiva and epidermal blistering, temporarily disabling the eye. Several countries set up research laboratories for testing the eye irritation potential of chemicals early in the 20th century, out of fears that mustard gas and other chemical warfare substances were going to be used in future wars (Wilhelmus 2001).

The use of animals, usually rabbits, in studying eye irritation was led by British and American military researchers in the period between WWI and WWII. Ida Mann (1893-1983) was assigned this task in Britain in 1939, in the hope that a rabbit model would unravel the mechanisms of eye irritation and lead to a possible remedy (Potter 1989). However, her data lacked a quantitative method that would allow for systematic comparisons. The U.S. military accelerated the research into chemical warfare during the 1940's and funded a large study to assess the effects of warfare gasses and other chemicals on the eye (Dahl et al. 1985). There, Jonas Friedenwald (1897-1955) developed quantitative scoring by grading the severity of different aspects of eye irritation on a numerical scale. Corneal edema, for instance, was graded from 0 to 4. The interobserver agreement of this system was shown to be reasonable (Knapp 1949), providing a way to compare different substances and allow statistical description.

John Draize (1900-1992), a pharmacologist at the FDA, adapted Friedenwald's procedure to test the safety of products intended for topical application. After obtaining his PhD in pharmacology and working for the University of Wyoming, he was recruited in 1935 by the U.S. Army to join a research team in Maryland, where the Chemical Warfare Service tested mustard gas and other chemical warfare agents (Graham, 2000). It was during this time that Lash-Lure, an eyelash dye containing severely irritating ingredients, caused its first victims. Photos of ocular complications were publicly displayed at an exposition dubbed 'Chamber of Horrors' by journalists (Lamb, 1936), which became partly responsible for the Federal Food, Drug and Cosmetic Act of 1938, passed by

the U.S. Congress. This was the first legislation that required manufacturers to assess their product safety before marketing. John Draize, by now familiar with harmful chemicals, joined the FDA's 'Division of Pharmacology' in 1939 (Draize and Kelley 1959). Here he was assigned to develop methods for testing the side effects of cosmetic products (Clark 1958). As head of the Dermal and Ocular Toxicity Branch, his studies on beauty products led to a system for quantitatively assessing the toxicity of topical substances. In the 1944 publication 'Method for the study of irritation and toxicity of substances applied topically to the skin and mucous membranes', Draize and his colleagues described how to asses acute, intermediate and chronic exposures by applying compounds the skin, penis and eyes of rabbits (Draize, Woodard, and Calvery 1944). Little did they know, this paper would become one of the most cited publications in toxicology, averaging over one journal citation per week during the last two decades of the 20th century (Wilhelmus 2001).

4.2 The rise of the Draize test [1945 – 1970]

Following the initial report, the FDA continued the use of Draize's techniques for safety evaluation of substances such as sunscreens, antiseptics and insecticides. More laboratories soon adopted it and the range of screened compounds increased considerably. Draize and other FDA researchers refined the method in the next years through experimental results, recommending the use of magnification devices to make observations (Draize and Kelley 1959). One of his papers (1955) concluded "Assessment of data, a consideration of species differences, and the allowance for a margin of safety, should make possible an estimation, in the case of a given compound, of the amounts, concentration, and frequency of application that may be tolerated by man". In the early 1960's scientists simply began referring to 'the Draize technique' or 'Draize test' for the descriptive scoring of biological effects on the eyes of laboratory animals (Wilhelmus 2001).

A method based on (and closely resembling) the original method as described Draize became an authorative testing process. It was incorporated in 1964 into the FDA regulations dealing with pharmaceuticals and cosmetics, and in 1972 also into the guidelines of the Consumer Product Safety Commission (CPSC) dealing with household products (Eskes 2010). These implementations mainly resulted in increased animal use for scientific testing during the 1960s and 1970s (Rowan A.N., Loew F.M., and J.C. Weer 1995). The cosmetics industry was forced to test for safety of their products, and in the case of eye irritation the Draize test remained the only alternative for a long while.

Meanwhile the publication of '*principles of humane experimental technique*' by Russell and Burch in 1959, while not passing by completely unnoticed in scientific circles, had little obvious impact on thinking or practice by academia and industry. A significant event did take place in 1969 with the founding of the Fund for the Replacement of Animals in Medical Experiments (FRAME) by Dorothy Hegarty, with the specific aim to advance Russell and Burch's vision that systemic application of the 3R approach would lead to humanitarian and scientific benefits. FRAME focused it activities mainly on *replacement* as the ultimate long-term goal, and saw *reduction* and *refinement* as achievable in the short term (FRAME 2010).

4.3 PUBLIC AWARENESS AND LEGISLATION [1971 – 1989]

The early part of the decade saw a substantial increase in laboratory animal use (Balls et al. 1999). This did not go unnoticed by animal welfare organizations and the media, fueling great public concern in Britain that eventually led to the Animal Welfare Year campaign of 1976/77 (Zurlo, Rudacille, and Goldberg 1994). In turn, the Committee for the Reform of Animal Experimentation (CRAE) was established to reform the 1876 'Cruelty to Animals Act'. In the Netherlands, the 'Act on Animal Experimentation' was adopted in 1977 and while not specifically referring to the 3Rs, did prohibit the use of animals for purposes that could be achieved by *in vitro* methods.

Several particularly important changes began to take place in the early 1980s. Henri Spira (1927-1998) launched a campaign to abolish the Draize eye test in the US. After learning that cosmetics manufacturer Revlon used 2000 animals in 1978, Spira began writing letters and questioning the company's chairman at stockholder meetings (Rowan A.N., Loew F.M., and J.C. Weer 1995). Under the 'Coalition to Stop Draize Rabbit Blinding Tests', he brought together over 400 animal organizations, and outlined the coalition's goals of how to eliminate the Draize test. 15 April 1980, Spira ran a full-page advertisement in the New York Times, seeking to influence the public opinion. A second advertisement was published on 7 October 1980. Protesters handed out leaflets and stickers in department stores over the world, and newspaper articles and television attention spurred the public awareness. Revlon's chairman, by December 1980, announced the company would offer $250.000 a year for three years for research at the Rockefeller University to research alternative safety tests (Rowan A.N., Loew F.M., and J.C. Weer 1995). After that, it was a simple step to persuade Avon (the industry leader), Bristol-Meyer and other cosmetics companies to contribute a 1 million dollar start-up grant through the Cosmetic, Toiletry and Fragrance Association (CFTA) to establish the John Hopkins Centre for Alternatives to Animal Testing (CAAT) (CAAT 2011).

Despite the commotion around cosmetics products, the chemical industry was bound by OECD guideline 405 (OECD 2002), first adopted in 1981, to use the Draize test for safety assessment of acute eye irritation of chemicals.

Meanwhile in Europe, discussions began that would later lead to the 'Council of Europe Convention for the Protection of Vertebrate Animals Used for Experimental and Other Scientific Purposes' and 'Council Directive 86/609/EEC of 24 November 1986 on the approximation of laws, regulations and administrative provisions of the Member States regarding the protection of animals used for experimental and other scientific purposes', encouraging 'research into the development and validation of alternative techniques' and implementing 3R principles (EC 1986). That same year, the European Research Group for Alternatives in Toxicity Testing (ERGATT) was established.

After the anti Draize campaigns in the U.S., simultaneous attempts to pass legislation focused industrial and congressional attention on alternative methods. The Animal Welfare Act and the Public Health Service Policy on the Humane Care and Use of Laboratory Animals were both revised under public pressure, incorporating the requirement that the 3Rs have to be considered before the start of research involving animals (Balls et al. 1999). In 1986, the Health Research Extension Act gave legislative force to the aforementioned Public Health Service Policy. So by the end of the 1980's, several new laws were in place in various countries over the world, while the

Draize test was still being used under the Food, Drug and Cosmetic act in the U.S. and the Cosmetics Directive in the E.U.

THE FIRST ORGANOTYPIC METHODS

Ever since the publication of the Draize test, the subjective component in the scoring of eye irritation was of concern to researchers . This first led to slight optimizations in the protocol, but it also caused a desire for objective parameters for better comparability. As early as 1971, Burton showed that measurement of corneal thickness in the rabbit *in vivo* correlated well with the eye lesions observed (Burton 1971). Ten years later he introduced the Enucleated Eye Test (EET) with isolated eyes of deceased rabbits, originally as prescreen for the *ex vivo* assessment of severe eye irritants (Burton, York, and Lawrence 1981). The parameters to measure eye irritation were expanded to include corneal opacity and fluorescein retention besides corneal thickness. The EET was innovative for being the first 'organotypic method' with potential to replace the Draize test. Organotypic methods use isolated organs of (dead) animals in toxicity tests (Eskes et al. 2005). The rationale for their use is that isolated organs maintain their normal physiological and biochemical functions, at least in the short term, to serve as a suitable model for the otherwise *in vivo* organ (Chamberlain et al. 1997b).

ICE TEST DEVELOPMENT

The EET was utilized at TNO Toxicology and Nutrition in the Netherlands and Shell Research Ltd. in the UK[a, c]. Both parties ran in-house validation studies against the live Draize test and both concluded that 'the method can be used for a reliable prediction of the potential of chemicals to cause eye injury' (Koeter and Prinsen 1985; Price and Andrews 1985). The main drivers for further use and development of the EET method at this point were personal convictions of the involved researchers, Herman Koeter and Menk Prinsen[a, c, g]. The large amount of animals used for contract research within TNO did not pass them unnoticed and it was felt that similar results could be obtained with alternative methods. Public awareness of the availability of alternative methods was very low in the Netherlands at this point, the early 1980s[c]. More importantly even in professional circles the significance of alternative methods was marginal at best, with many executives including the board at TNO holding the view that animal tests were necessary and could not be replaced, lacking any impulse to do so[a,c, g].

In 1988-1989, the Commission of the European Communities sponsored an initial study on the evaluation of alternative methods for the eye irritation test, including the EET method, and concluded that the test correctly predicted the in vivo grade of the compounds (CEC 1991). However, experience at TNO identified practical problems with the use of rabbit eyes in the EET[a, c]. First, the relative scarceness of rabbit slaughterhouses could give proximity problems for the adoption of the test by other laboratories. Second, even when a slaughterhouse was near, the process for slaughtering the rabbits made it difficult to obtain undamaged eyes (Prinsen and Koeter 1993). The use of chickens circumvented this problem, because they are decapitated first, leaving the eyes intact for the researcher (instead of slaughterhouse personnel) to remove them. Research that led to chickens as the optimal choice was not funded by TNO but by the 'Dierenbescherming' instead (Dutch welfare organization for the protection of animals), something unequalled at the time, and indeed a first for a Dutch welfare organization to financially support an organization that also performed animal tests[c, g]. This demonstrated the beginning of a shift in the general mentality of welfare organizations, now actively participating in

the replacing of animal tests instead of passively criticizing the practice[g]. The result was the successively named Chicken Enucleated Eye Test (CEET), which for harmonization purposes would later be renamed the Isolated Chicken Eye method (ICE). Despite the results from TNO that the ICE "can be considered a sensitive, but not over-sensitive, means of predicting the eye irritancy potential of all types of compounds" (Prinsen and Koeter 1993 p.75), the method at this point had only been tested with 21 reference compounds in-house, and formal validation would still need to be undertaken. At this point, the expectation was that within a timeframe of 5-10 years, a complete validation was possible, in particular for the acute toxicity tests[a, c, g].

BCOP TEST DEVELOPMENT

Meanwhile, different alternative methods besides the EET and ICE test were being developed in different parts of the world during the 1980s. Favored by the spectacular advances in cell culture technology at that time[c], the great majority of these alternatives were cytotoxicity-based methods. This technology consists of a process in which a target cell, generally of non-ocular origin is exposed to a substance and some endpoint is measured (Borenfreund and Borrero 1984). Frazier (1987) reviewed an extensive list of over 30 essays, but put forward one of the main criticisms on cytotoxicity tests: they do not address the major clinical sign of eye irritation, corneal opacity. Precisely because of this reason, Muir (1984; 1985) had begun to develop an organotypic method that used bovine cornea (again, because they were 'freely available from the local slaughterhouse') and assessed eye irritation through 'corneal opacity development' with a specially designed device now known as an opacitometer[13].

Severeal years later, around 1988[h], the method as described by Muir gained the attention of Merck[14] in a research program to develop alternatives for eye irritation testing (Gautheron et al. 1992). Originally, the method was to be used internally to guard personnel safety by predicting the irritant potential of process intermediates and compounds in the production of pharmaceuticals[e, j, h]. Many of such compounds were solids, proving difficult to test with available cytotoxicity methods[h]. Also it was sensible to measure corneal opacity, as it was still the most heavily weighed component in *in vivo* scoring systems. The technique was substantially modified and extended from Muir's version(Gautheron et al. 1994). Besides opacity, dye penetration was added as another endpoint, determined by amount of fluorescent dye that penetrates through the cornea(Gautheron et al. 1992). This resulted in "an assay that could accurately and sensitively predict the ocular irritation potential of chemicals from a variety of classes, based on measurement of opacity and dye penetration" (Gautheron et al. 1992 p.443). The method was adequately named the Bovine Corneal Opacity and Permeability (BCOP) assay.

Even though the method was originally developed for internal purposes[h], it would become clear in consecutive years that other laboratories started to use the BCOP assay to predict the irritancy of

---

[13] The opacitometer allows a beam of white light to be directed through the cornea in its holder to a photocell. With an empty holder in the light beam, the voltage is set at 2.5 V before each reading by adjusting the variable resistor if required. All surfaces of the opacitometer are matt black. Dark current is small with this system, a voltage of below 0.005 being recorded with the lamp off. With a cornea in the holder, the light beam that passes through measures the 'development of opacity', indicated by a drop in the voltage measured (Muir, 1984).

[14] Being one of the largest pharmaceutical companies in the world, the company name has seen changed over the years. Formally the company is registered as 'Merck & Co., Inc' and outside the US also known as Merck Sharp & Dohme (MSD). The remainder of this report simply uses 'Merck' for clarity purposes.

cosmetics, pharmaceuticals and chemicals (Rougier et al. 1992; Vanparys et al. 1993) after publication of data confirming its usefulness[h]. While tests that are used internally by pharmaceutical companies for occupational hazard screening are hardly subjected to regulation[15], wider use of the method required extensive validation. In particular, external reproducibility and technical feasibility have to be established, and a greater variety of substances and chemical structures have to be used. Between 1991-1992, Pierre Gautheron at MSD led an interlaboratory study with 12 public and private laboratories from seven European countries to get more information on the potential of the essay to replace the Draize test, in one of the first major validation works for the BCOP. Financed by the European Commission, data indicated that the BCOP correctly predicted (irritant/non-irritant) in 85% of the substances tested (Gautheron et al. 1992).

4.3 Multi-laboratory undertakings and commitment to change [1990 – 1999]

Publication of the alternative methods and results of initial small or in-house validation studies in the late eighties led to an extensive body of literature on potential replacements for the Draize test. By now many of the alternatives were used in-house or as a pre-screen for an original Draize test[c,h,j], but not in legislation. A number of governmental bodies, including the European Commission (EC), the British Home Office (HO), the European Centre for Ecotoxicology of Chemicals (ECETOC), as well as various industrial companies and industry associations, scientists and animal welfare groups felt that reliable and relevant alternatives should replace the Draize test as soon as possible (Balls 1995). The aforementioned stakeholders arranged a meeting in November 1991 to review the recent progress and concluded that current methods were mature enough for a formal validation study that would later become known as the EC/HO validation study. Nine tests were included on the basis of an EC-funded pilot study. At this point, expectations from the ones involved in development were high; this was going to be the 'ultimate' validation study that would decide upon a replacement for the Draize test[b,c,g,h,k,l]. However, analysis showed that combinations of data from essays explained more variability of the data than any single test, and none of the tests used alone was sufficiently predictive of in vivo eye irritancy for the full set of test chemicals (Balls et al. 1995). The most valuable contributions included the importance of optimizing protocols, and the refinement of the prediction models *before* entering alternatives in a large-scale study (Balls et al. 1999).

The COLIPA study was designed with the results of the EC/HO study in mind, trying to incorporate the 'lessons learned', but results were essentially similar: the combinations of endpoints resulted in better prediction models and no single *in vitro* method could be validated firmly (Brantom et al. 1997). In total, six major validation studies were undertaken between 1990 and 1999 (including the timeframe for publication of the results). Table 4.1 provides an overview of the assays and chemicals included in each study, as well as an overview of the main conclusions and remarks. By now, around 30 methods and hundreds of test materials had been evaluated by an extensive number of laboratories.

[15] Regulation with regard to requirements for the test: formal validation was and is not required for such tests. However, the screening for occupational hazard by companies is required though OSHA regulations in the US[f].

In retrospect, several factors can be identified that could account for the low predictions in the EC/HO study[a,b,c,h,l,l] (Balls et al. 1999; Eskes 2010): a) the choice of test chemicals; b) the fact that *in vitro* tests only partially model the complex in vivo eye irritation response; c) the variability of the in vivo eye irritation responses linked to the subjectivity of scoring, uncontrolled exposure conditions and variability of animal responses; d) the use of the MMAS[16] as the in vivo endpoint; e) the choice of statistical methods (correlation analysis and linear regression) and; f) the protocols and prediction models which might have been insufficiently developed at that time.

Despite the fact that still no single method could replace the Draize test, many of the lessons above improved the practice of validating alternative methods[b,d,g,i,l] (also in applications other than eye irritation[g,i,l]). The extensive evaluation of the alternative methods demonstrated their usefulness to industry and CRO's for specific and limited purposes.

[16] Modified Maximum Average Score, a scoring system used in the original Draize test.

TABLE 4.1: OVERVIEW OF MULTI-LABORATORY VALIDATION STUDIES

STUDY	TIMEFRAME	TEST METHODS	DATA/SUBSTANCES	CONCLUSIONS	REMARKS
EC/HO (BALLS ET AL. 1995)	1991-1994	RBC haemolysis,NRU, FL, SM, IRE, ICE, BCOP HET-CAM , EYETEX	Test set of 60 single chemicals	- good reproducibility between laboratories - none of the 9 tests sufficiently predictive of in vivo eye irritancy - combinations of data from essays (FL, ICE, BCOP, NRU) explain more of the variability in the data than any single test.	- variability of *in vivo* data may be responsible for low precision of results - importance of optimizing protocols and refinement of PM's before entering validation studies
COLIPA (BRANTOM ET AL. 1997; PAPE ET AL. 1999)	1993-1996	CAMVA, EYETEX, FL, HET-CAM, NRU, PTG, Predi-Safe, RBC, SM, TEA	55 test substances, of which 23 were cosmetic ingredients and 32 were formulations.	- by the predefined criteria of reliability, none of the methods could be confirmed as valid replacement for Draize - FL, RBC and TEA essay satisfied 1 criterion of reliability, but no firm conclusions regarding their validity. - FL had close fit in 1 lab, other labs overpredicted eye irritation for FL.	- multivariate analysis of combined test data also resulted in better predictions - moderately irritating substances underrepresented, while difficult/crucial to identify
BGA/BMBF (SPIELMANN, LIEBSCH, AND KALWEIT 1996)	1988-1994	3T3 NRU, HET-CAM	Phase 1: 34 test chemicals Phase II: 166 industrial chemicals, representative for pharmaceutical and chemical industry	- 57 chemicals were excluded because of poor in vivo or *in vitro* data - chemicals can be classified as severe irritants (R41) with sufficient reliability by the combined use of HET-CAM and 3T3 NRU test. German authorities accept data for R41 classification.	- earlier project had indicated NRU and HET-CAM as promising - multivariate statistics in the development of PM's was fruitful.
CFTA (GETTINGS ET AL. 1996)	1990-1996	24 tests, excluding variations: ADM, AMA, BCOP, CAM, CAMVA, DMEM, FBS, FHSA, HBSS, HET-CAM, IS, IT, LDE, LDH, MDC,	Phase I: hydro-alcoholic formulations (10 materials); Phase II: oil-water emulsions (18 materials); Phase III: surfactant-	- variability in Draize data striking, even though animal tests were carried out in single lab - predictivity of each *in vitro* method was shown to vary according to the type of material being investigated. - In Phases I and II, variability was smallest for the	- novel inclusion of a 95% prediction interval reflected the variability of both the *in vitro* and the in vivo test

		MDCK, MEM, NR, PBS, PGE2, RBC, RMA, UMA	based formulations (25 materials).	least irritating chemicals, and increased as the irritancy increased. In contrast, variability in Phase III was greatest in the middle of the irritancy range and smallest at the two ends of the scale. The Draize scores were confined to the lower end of the Draize scale, which is the most relevant range for cosmetic formulations.	
IRAG (CHAMBERLAIN ET AL. 1997A)	1991-1994	29 different methods	Based on available *in vitro* and animal data; over 60 data sets from 41 laboratories	- results revealed differences in predictivity between test methods for the same types of chemicals, and between chemical types for the same test method - none of the tests showed a satisfactory performance across all chemical groups and no combination of tests could completely replace the animal test - the ability to obtain strong *in vitro*-in vivo correlations was compromised by the variable nature of the animal test.	- set of guidelines was developed to standardize data for submission and review - 'alternatives to the Draize test are currently being used by industry as screens in the risk assessment process for product development' - first retrospective study
MHW/JCIA (OHNO ET AL. 1994)	1991-1994	HET-CAM, CAM-TB, HD, RBC haemolysis, HD, SKIN2 (ZK1100 model) and MATREX; CornePack, SIRC-CVS, SIRC-NRU, HeLa-MTT, CHL-CVS and EYTEX.	38 cosmetics ingredients in three phases	- none of the alternative methods could be used to test all types of test substances. - a battery of tests would be needed to optimize the ability to predict eye irritancy.	- in vivo data were found variable, particularly MMAS values in the range 15-50, which is important for the evaluation of cosmetic ingredients.

Changes in legislation

Several influential decisions have been made on a policy level. In 1991 the European Centre for the Validation of Alternative Methods (ECVAM) was established as part of a requirement in *Directive 86/609/EEC* on the protection of animals used for experimental and other scientific purposes, pointing out that the 'Commission and the Member States should actively support the development, validation and acceptance of methods which could reduce, refine or replace the use of laboratory animals' (ECVAM, 2011). While not involved directly in the six major validation studies on alternatives for eye irritation, ECVAM has been the main EU body for the coordination of scientific and regulatory acceptance of alternative methods ever since their establishment[b,l].

In 1993 two other major decisions were made in Europe. Resolution decision 2179/98/EC implemented a 50% reduction objective for the number of vertebrate animals used for experimental research before the year 2000 (EC 1993b). That same year, the 6th amendment to the cosmetics directive (93/35/EEC) was accepted, introducing a ban on the marketing of cosmetic products containing ingredients, or combinations of ingredients that have been tested on animals as per 1998 (EC 1993a). However, the 6th amendment also stated that if there had not been sufficient developments in alternative methods, the Commission would be able to postpone the ban as long as they did so before Jan 1st 1997. The report for 1996 concluded with regard to eye irritation that "Currently there are no validated alternative methods capable of replacing the OECD-405 in vivo eye irritancy test" (EC 1997a) and the ban was postponed to June 30th 2000 with directive 97/18/EC (EC 1997b). The date of the ban was postponed again, against the will of the European Parliament, to June 30th 2002 with Directive 2000/41/EC due to 'insufficient progress in satisfactory methods to replace animal testing' (EC 2000). Regardless of the repeated postponement, the message of the marketing ban was clear and provided a major stimulus for the development of alternatives[b,d,f,g,h,l,jl], both for eye irritation tests and in general.

The Unites States saw the NIH Revitalization Act of 1993, charging the National Institutes of Health with developing research methods that do not require animals or reduce the number of animals used (NIH 1993). The National Institute of Environmental Health Sciences (NIEHS) established an Applied Toxicology Program that would eventually become the Interagency Coordinating Committee for the Validation of Alternative Methods (ICCVAM) in 1994. Created as an ad-hoc group of (at the start) 15 representatives, ICCVAM's first task was to develop criteria for the validation and regulatory acceptance of alternative methods and recommend a process through which scientifically validated alternative methods can be accepted for regulatory use (ICCVAM 1997). The report was published in 1997, but ICCVAM did not get actively involved in validating the organotypic eye irritation tests until after 2000.

ECVAM organized a workshop named 'Eye Irritation: The Way Forward' (Balls et al. 1999) in 1999, marking the end of the century with a thorough review of the problems associated with the validation efforts so far. Results from the workshop were the proposition of Reference Standards (RS) to circumvent the Draize test variability and again emphasizing the need for validated alternatives (Zuang 2002), especially in the light of the demands set by the Cosmetics Directive[b]. Both points would require further investigation[b,l].

4.4 TWO TESTS ARE BETTER THAN ONE: TIERED TESTING & DECISIVE VALIDATION [2000 ~]

The new millennium coincided with a Reference Standard study organized by ECVAM during 1999-2000 as a follow-up to the Eye Irritation workshop, to assess the feasibility of using RS for validating five alternative tests . Conclusions were that scientific validation could be possible with a RS approach, but classifications of chemicals in different classes needed to be redefined first because of a false assumption that an RS approach would allow the testing of broad classes of chemicals (Zuang 2002). Furthermore, the EU carried out a survey amongst Member States to establish the uses and acceptance of the ICE, IRE, HET-CAM and BCOP test methods. Data from all tests appeared to be accepted already by several national authorities on a case-by-case basis to identify severe irritants. On that basis, it was suggested to member states that they should harmonize their position on the acceptance of the tests through the EC (Zuang et al. 2007).

The previously postponed marketing ban on cosmetics was modified in the 7th amendment to the Cosmetics Directive 2003/15/EC and now consists of two distinct bans (EC 2003):

- The use of animals for testing of cosmetics or cosmetics ingredients is banned in the EU, and effective from March 11th 2009.
- A ban on the marketing of products (containing ingredients) tested on animals as imposed from March 11th 2009, with the exception of tests for repeated dose toxicity, which should be banned from March 11th 2011 on the condition that alternative methods are available.

Partly because of these decisions, a thorough review on the status of the most promising alternatives for eye irritation was carried out on behalf of the European Commission (Eskes 2010). The idea for using tiered testing strategies had been around in ECVAM since the early 2000's[b], because of the consensus on the idea that the range of eye irritation potential covered by the Draize test was unlikely to be reproduced by a single alternative method[b,g,l]. This is represented in figure 4.1, where the optimal test range for eye irritancy of the BCOP test and EpiOcular[17] is compared to the range covered by the Draize test, in relation to different types of substances. A solution to the narrow range of individual alternative methods would be to make use of combined testing strategies that utilize strengths of individual methods to assess a specific range of chemical classes or irritation potential. On the basis of this idea and the pressure from the Cosmetics Directive, it was decided to start a retrospective validation study on the basis of available data in 2004[b,l]. ECVAM and ICCVAM cooperation had only increased since their respective establishments[b,l], and ICCVAM at this point in 2004 was already in the process of reviewing several organotypic methods[b]. As such, ICCVAM became responsible for performing the retrospective study on four organotypic methods, while ECVAM would perform the retrospective study for cytotoxicity-based methods[b] (Zuang et al. 2007). Table 4.2 provides an overview of the included methods. In 2005, an ECVAM-organized workshop formally adopted the tiered-testing strategy as the most promising option to replace the Draize test (Scott et al. 2010). It suggested 2 approaches: Bottom-Up (begin with using test methods that can accurately identify non-irritants) or Top-Down (begin with using test methods that can accurately identify severe irritants) approach. This eventually formed the basis for the retrospective validation studies[b,d,l].

[17] An alternative method for eye irritation testing based on a Reconstructed Human Tissue model

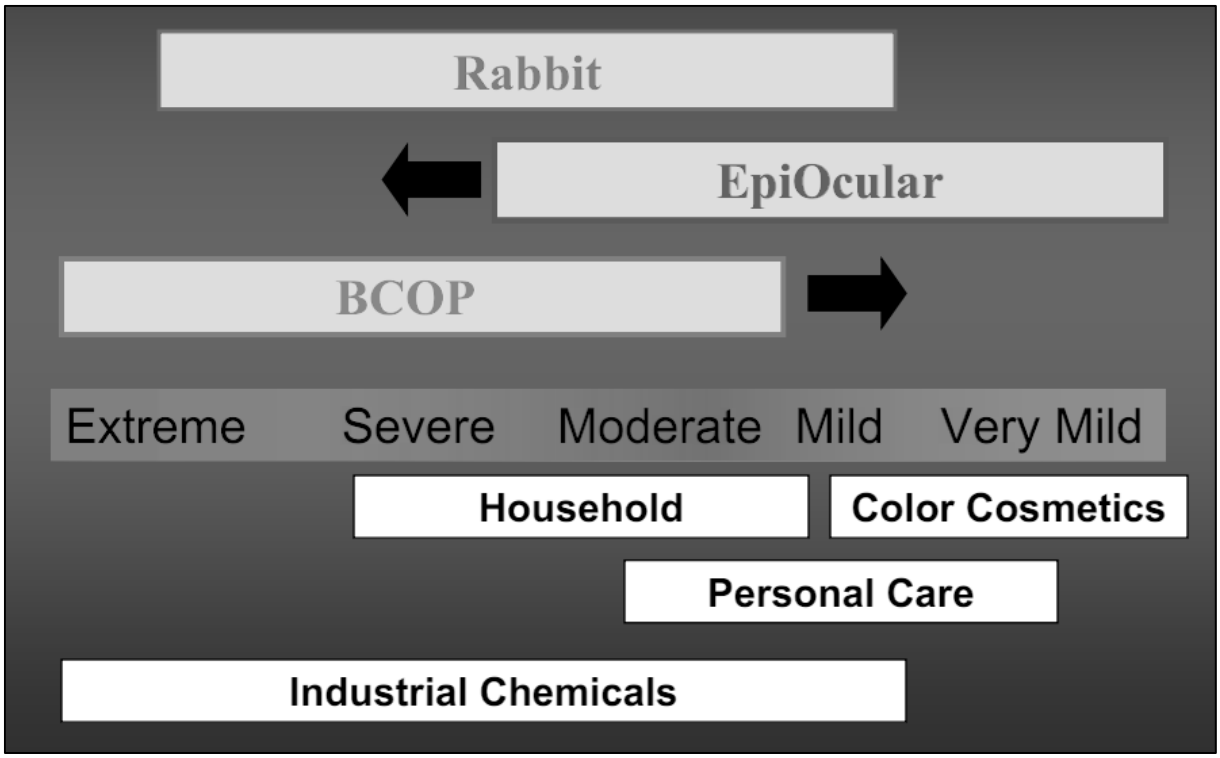

FIGURE 4.1: THE OPTIMAL IRRITANCY RANGE COVERED BY THE BCOP AND EPIOCULAR TEST

Table 4.2: overview of validation efforts on Draize alternatives since 2000 (Source: Adapted from Eskes 2005)

Organotypic Methods (ICCVAM Retrospective 2003-2006)	**Cytotoxicity-based methods (ECVAM Retrospective 2005-2008)**	**Reconstructed human tissue models (ECVAM Prospective 2007 - ongoing)**
Bovine Corneal Opacity and Permeability Test (BCOP)	Neutral Red Release assay (NRR)	EpiOcularTM
Isolated Chicken Eye test (ICE)	Red blood cell (RBC) haemolysis test	SkinEthicTM reconstituted Human Corneal Epithelium (HCE) model
Isolated Rabbit Eye (IRE)	Fluorescein leakage (FL)	
Hen`s egg test on the Chorio-Allantoic Membrane (HET-CAM Assay)	Cytosensor Microphysiometer (CM)	

As part of the retrospective validation study, the NTP Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) compiled extensive Background Review Documents (BRD's) on four methods (ICE, BCOP, IRE, HET-CAM) in 2004 (ICCVAM 2006b). NICEATM organized the 24-member international panel in collaboration with the ICCVAM and ECVAM. The conclusions were published in March 2006 in an expert panel report, which formed the basis for recommendation to and acceptance from Federal authorities in the United States. The ESAC also based their recommendations on this review. In 2007 preparations started to adapt the BRD format and data submission to the OECD, which finally adopted the BCOP and ICE methods in Test Guideline 437 and 438 in September 2009 (OECD 2009a; OECD 2009b).

The BRD's also included two recommendations to improve the sensitivity of the BCOP test for the lower eye irritancy ranges. By adding histology as an endpoint, and using a more sensitive

opacitometer, it should become possible to use the BCOP test for assessing mild to moderate irritants (ICCVAM 2006a), a range for which it currently still is not validated. On the basis of these recommendations, Johnson & Johnson have developed a laser-based opacitometer[e,j] that can more accurately establish corneal opacity (Van Goethem et al. 2010).

With validated methods for severe eye irritants, the tiered-testing strategy still requires an alternative method for the assessment of mild to moderate irritants. Reconstructed Human Tissue (RhT) models had been found to be most promising for this purpose as early as 2004[b,l], and ECVAM started a prospective validation study for 2 RhT models in 2007, which is currently still ongoing[b].

In summary, the OECD acceptance of the ICE and BCOP proved to be a milestone for innovation in alternative methods for eye irritation testing. A trajectory of over 20 years of technological development of methods, and scientific progress in the performance of validation studies resulted in the global endorsement of 2 organotypic methods. The next chapter focuses on the multi-actor network responsible for this achievement, and the dynamics that have occurred to overcome the necessary barriers.

5. Socio-technical analysis I: multi-actor network

This chapter uses the framework of the MLP for further analysis of the case description. The multi-actor network of the safety-testing regime is presented and discussed. This lays the basis for the analyzing the dynamics in the safety-testing in the next chapter. Over the years, many different actors have been involved with eye irritation testing. The MLP deals with actors as actor groups, which share common interests or perform similar functions. The focus when discussing the actors is on their roles relating to either development or validation of alternative methods for eye irritation testing.

Figure 5.1 depicts the different actor groups involved in the socio-technical regime of safety testing. Companies or institutions named in the figure serve as examples, not as an exhaustive list of all possible actors in the actor group (i.e. not all cosmetics companies are named in the cosmetics industry). However, the actor groups together do provide a complete picture.

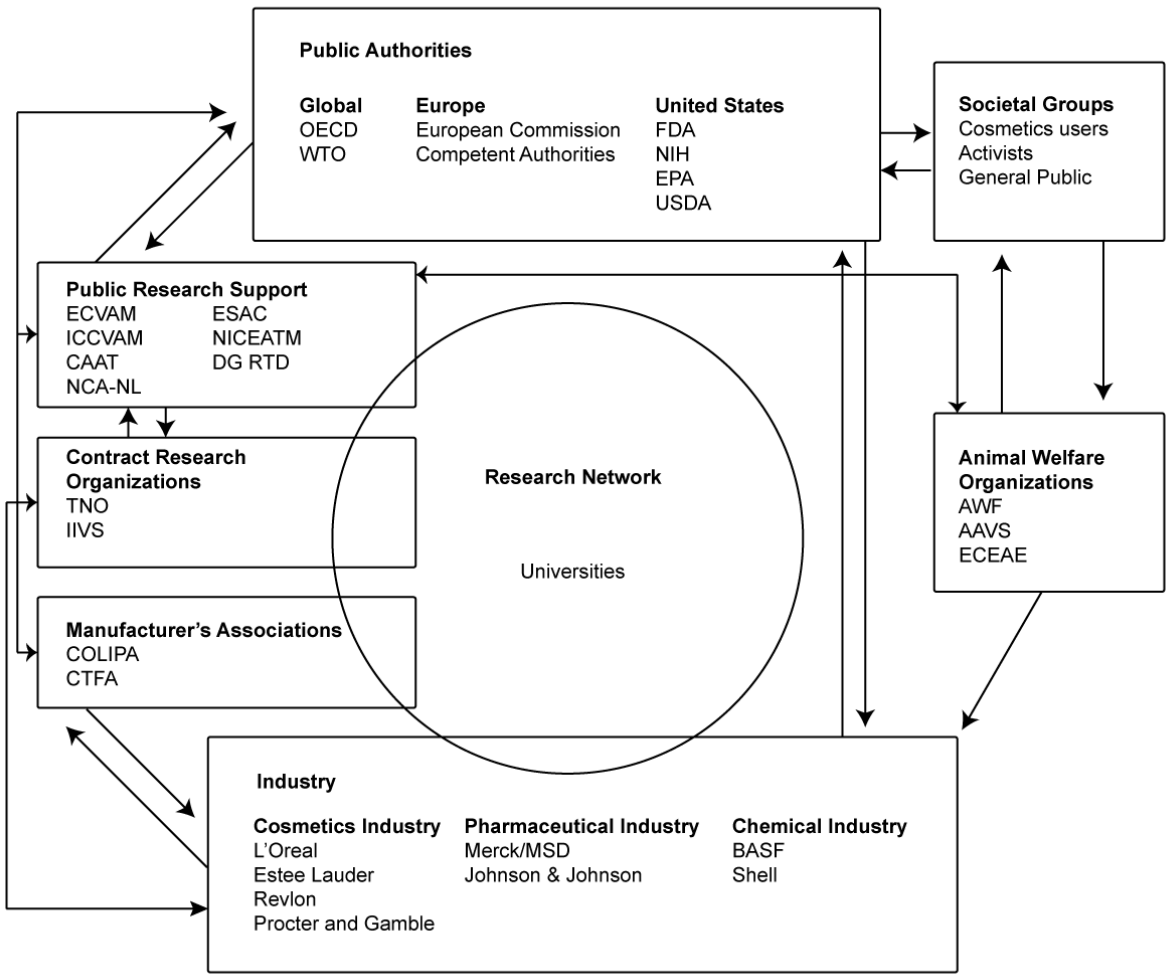

Figure 5.1: the multi-actor network of the safety-testing regime

5.1 Industry

Cosmetics industry

The most prominent actor group in the case of eye irritation testing is the cosmetics industry. It is made up of companies from all different sizes, operating either nationally or internationally, which are bound by local regulations and internal company standards to foresee in a certain degree of safety of their products. To this end, the Draize test has been used extensively[b,d,g,l,l]. The drive of cosmetics companies to develop new products requires them to assess the irritancy potential of many different substances before they appear in a market-ready formulation[i]. This results in the in-house development of screening techniques for safety assessment. On the one hand, such development can boost innovation because the techniques are usually *in vitro.* Animal experiments are relatively costly and time-intensive, *in vitro* methods are (usually) cheaper and easier to perform[c]. On the other hand, innovative *in vitro* methods can provide a competitive advantage over competitors when translated into reduced development costs or improved screening capability (Hartung 2008a). This is beneficial for the company itself, but it also results in the innovative methods having a confidential character and not being shared with other actors openly. The main exception to this is when a company would want to use the method in a formal regulation process, for which a formal validation procedure is also required[h].

In validation studies, cosmetics companies are usually involved from the beginning of a study. Their function is either to provide the test method to other labs, or to perform a part of the tests necessary for the study in their own laboratory. The need for cosmetics companies to comply with the Cosmetics Directive resulted in a situation where the cosmetics industry is the major contributor to the advancement of alternative methods[b,e,f,g,l,j,l].

Pharmaceutical industry

Similar to the cosmetics industry, the pharmaceutical industry is represented by companies differing in size and geographical presence. Their prominence in the case of eye irritation testing is much lower, because the Draize test is not formally required for medicinal products to enter the market[e,j]. Eye irritations tests are usually performed by pharmaceutical companies to test irritancy potential of intermediate compounds used in the production of drugs[b,e,f,j]. Results are used to maintain adequate safety procedures for employees that work with large volumes of such compounds during production; volumes that consumers or the general public would never be exposed to. This allows for different demands to safety margins when compared to consumer product safety testing[e,j]. Throughout Europe and the United States, pharmaceutical companies themselves are responsible to provide reasonable degrees of safety to employees[h]. As a result, they are not bound by formally validated methods and can also use tests that are developed in-house[h,j].

In the case of eye irritation, this did lead to the development of the BCOP test by Merck/MSD. Only later would the BCOP enter formal validation programs because of its potential to replace the required Draize test in formal regulations. Several multinational pharmaceutical companies also produce regular consumer products, such as toothpastes (GlaxoSmithKline), topical ointments (Merck) and skin and hair care products (Johnson and Johnson). These are of course bound by regulations for cosmetics and can lead pharmaceutical companies to also pursue validation efforts, as can be seen in the example of the BCOP test[h].

Chemical industry

The chemical industry sits in between the cosmetics and pharmaceutical industry as far as volume usage of eye irritation tests is concerned[b]. Almost all chemical substances require some degree of safety testing, but before the introduction of REACH it was very dependent upon the type of substance and its perceived uses, which safety data would be required (Lilienblum et al. 2008). The responsibility for safeguarding laid with the companies itself and the burden of proof for determining (lack of) safety of a substance was on the side of the regulators. Despite this, many companies did test for eye irritation potential and Shell was one of the first companies to use the (rabbit) Enucleated Eye Test for in-house testing[a]. Also as a result of the burden of proof issue, was the relatively quick adoption of alternative methods: there was no need for strict, formal validation studies. As long as the company had to decide upon which test would be adequate for the provision of a reasonable safety level, and alternative methods would comply with those (internal) standards based on available experience, it would make sense to use them. Another contributing fact was that it was clear from early on, that alternatives such as the ICE and BCOP test would at least be able to discriminate severe irritants. This very often was good enough, even if the substance in fact would be only moderately irritating[g]. In most cases when a substance would prove severely irritating according to the BCOP/ICE test, the only consequence for a producer of chemicals was that the substance had to be labeled as 'corrosive' (on packaging, etc.). Usually this was not really a problem[g], and in doing so the producer had warned the user of the corrosivity and fulfilled its duty with regard to substance safety.

An exception has to be made for substances that are manufactured by the chemical industry, but used as ingredients in cosmetics products. This is how the chemical industry is tied in with the cosmetics industry. For substances that fulfill this criterion, it has to be possible for the cosmetics company to demonstrate that it complies with the (generally stricter) regulations for cosmetics(Hartung 2010). Hence, methods that are not validated for this purpose cannot be used, so before the validation of the ICE and BCOP methods, Draize tests could be required. Often the producer (i.e. chemical industry) performs the tests necessary for its client. The introduction of REACH complicates this matter, as it will require mandatory safety testing for chemicals, including toxicology batteries with eye irritation endpoints. However, REACH does not impose the use of alternative methods[b], and with only partial replacement of the Draize test, this could mean a substantial increase in the amount of Draize tests performed by the chemical industry[a] (Hartung 2010). Also, it is unclear how exactly the Cosmetics Directive is compatible with the requirements set by REACH, especially for ingredients manufactured by chemical industry for use in cosmetics. The European Commission is currently investigating this issue (EC 2011).

5.2 Manufacturers' Associations

Manufacturers' associations represent companies from the cosmetics industry[b]. They provide a platform for collaboration, lobby for members' interests in political decisions and sponsor research programs through grants[b,c,i]. This also includes the funding of validation studies, as all members benefit from the availability of validated alternatives.

COLIPA in Europe has its own Steering Group for Alternatives to Animal Testing (AAT), which closely collaborates with ECVAM, the EC Scientific Committee on Consumer Safety (SCCS), and other groups having a legitimate interest in the outcome of the research (COLIPA, 2011). Other

international ties with ICCVAM, the OECD and trade organizations (CFTA) are also present. The AAT has its own research program on eye irritation, which builds on the experience of earlier validation studies and is focused on the identification of *in vitro* endpoints that are more predictive of the *in vivo* human response to chemical injury to the eye. This should enable the development of prediction models for pre-validation of new or improved/optimized *in vitro* methods that could proceed to formal validation.

The Personal Care Products Council (PCPC, formerly the CTFA) is active in the United States and represents the voice on scientific, legal, regulatory, legislative and international issues for the personal care product industry. Targeted heavily after the anti-Draize campaign in the US, the CFTA provided funding for the CAAT at the John Hopkins School of Medicine (Wilhelmus 2001), in part specifically focused on finding alternatives for the Draize test and organizing validation studies. Meanwhile, they were more successful than their European counterpart in preventing a ban on animal testing, which was being contemplated by the US authorities at a state level. Eventually only California implemented the legislation for a short while, until the Californian Governor vetoed it after persistent lobbying in combination with a statewide editorial campaign by the CTFA. While the CTFA early on was very important for research into alternative methods, recently COLIPA seems to be more proactive in such efforts[f].

5.3 Contract Research Organizations

CROs provide testing services for companies that are either unable to perform tests because they lack the necessary equipment, or choose to outsource this type of work for other reasons. Because of their particular job of performing contract research, CRO's tend to be specialized and deal with relatively large volumes of tests[b,f,k]. Consequently, improvements in alternative methods can provide substantial benefits and this does provide a stimulus to CRO's to develop alternative methods. It remains difficult for CRO's to profit from the development of an alternative by licensing it to other parties[d,h,l]. Instead, profits have to come from selling the testing service.

CRO's are still bound by regulatory requirements that apply to their clients, and can only use validated methods in the cases where the data will be submitted to authorities. Whenever this is not the case, CRO's can advise clients with the use of appropriate alternative methods that are not formally validated yet. TNO for instance, actively advised the use of the ICE test[a,c] even before it was endorsed by the OECD[18].

IIVS is a US based CRO that actively promotes the use of alternatives and as a result is active in research participation for the development of new methods. They maintain close ties with EPA and FDA in the US, which are responsible for most of the requirements to eye irritation testing[f]. Members of IIVS have also served as experts in several peer-review committees of ICCVAM, to prioritize the development of alternative methods and review validation studies. This demonstrates that people who work with alternative methods on a daily basis can provide the necessary, and highly valued, expertise in such committees[f].

[18] Of course, when appropriate for the desired purpose of the tests.

5.4 Public Research Support

Organizations in this actor-group form a special kind of actors that saw their creation go in line with the maturation of the safety-testing regime. Awareness of the 3R's and alternative methods had penetrated many public authorities. These saw the importance of their active support in the development and validation of alternatives to further the 3R principles. Also, they foresaw the magnitude of the validation efforts that would be required for substantial progress in this area. Europe was ahead in this regard, and 1991 marked the creation of ECVAM by the Commission of the European Communities, pointing to a requirement in Directive 86/609/EEC, which stated that '*the Commission and the Member States should actively support the development, validation and acceptance of methods which could reduce, refine or replace the use of laboratory animals '*(EC 1986).

The duties and tasks of ECVAM are defined as follows (Directive 2010/63/EU), and are generally exemplary for other public research support actors (EC 2010a):

a. Coordinating and promoting the development and use of alternatives to procedures including in the areas of basic and applied research and regulatory testing;
b. Coordinating the validation of alternative approaches at Union level;
c. Acting as a focal point for the exchange of information on the development of alternative approaches;
d. Setting up, maintaining and managing public databases and information systems on alternative approaches and their state of development;
e. Promoting dialogue between legislators, regulators, and all relevant stakeholders, in particular, industry, biomedical scientists, consumer organisations and animal-welfare groups, with a view to the development, validation, regulatory acceptance, international recognition, and application of alternative approaches.

The US followed with the creation of ICCVAM in 1994, under the NIH Revitalization act from 1993 that required '*the establishment of criteria for the validation and regulatory acceptance of alternative toxicological testing methods and the recommendation of a process to achieve the regulatory acceptance of scientifically valid alternative test methods*(NIH 1993)*'.* Despite their similarity in names, the set-up and roles of ECVAM and ICCVAM differ considerably. ICCVAM much more fulfills a role as regularly meeting advisory and peer-review body, resembling ESAC in Europe. However, it is made up mainly by representatives from regulatory agencies, which are exactly the ones excluded in ESAC (Bottini, Amcoff, and Hartung 2007). This has the advantage that, while decisions by ICCVAM are not necessarily binding and still have to be accepted by individual agencies, it does better pave the way for this step of acceptance. This often proves an obstacle in the European process, where implementation by member states is more separated from the decision-making process (i.e. after ESAC endorsement of a new method, individual countries still have to implement it in their regulations). Table 5.1 gives an overview of the organizational differences between ECVAM and ICCVAM.

Table 5.1: Organizational differences between ECVAM and ICCVAM (Source: Adapted from Hartung 2007)

	EU	**US**
Research and Development	ECVAM	-
Validation	ECVAM	NICEATM
Peer-review	ESAC	ICCVAM
Regulatory acceptance	Diverse	Agencies in ICCVAM

The importance of providing coordination activities in a network that is characterized by a need for cooperation between many different stakeholders cannot be underestimated. In this sense, public research support organizations fulfill a function as a central node in an otherwise rather decentralized network.

5.5 Research network

The research network is not clearly defined, but instead composed out of actors from different actor groups. This is especially the case for research in the form of validation studies, where representatives from different sorts of backgrounds participate in ad-hoc research projects[b,d,g], usually under the management of a dedicated 'management team' that is responsible for the continuation and completion of the study. The management team in the EC/HO study (1994) for instance, was composed of four representatives with background in industry, academia, the British Government and the European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC). In more recent studies, ECVAM is usually responsible for the management of the study (Eskes, 2011). ECVAM can (and often does) still appoint experts from different backgrounds to the management team[b]. The steering committee is commonly composed of –again- different experts that serve as an internal peer-review for feedback and input to ensure optimal results (Bottini, Amcoff, and Hartung 2007). The structure of the management and organization of a validation study is depicted in figure 5.2 in appendix D, along with a concise explanation.

The multi-actor network in figure 5.1 also represents the research network as consisting of different representatives from the actor groups that are crossed by the circle of the research network. These are industry, CROs, manufacturer's association and public research support actors. Academia is included in the research network through Universities that employ expert toxicologists.

Research into the development of *new* alternative methods is somewhat overshadowed by the vast research efforts that go into validating alternatives. This is especially the case with eye irritation alternatives, most of which have been developed during the 1980's. Research of this type does not take place in the same ad-hoc network fashion that characterizes validation studies. Instead, a lot of development takes place in-house, within R&D departments of companies and CRO's[c,d,h].

5.6 Public Authorities

Public authorities in the multi-actor network are divided by geographical boundaries, as regulations are region-specific. In the safety-testing regime, public authorities play a dual role. They provide the legislation that prescribes the necessary safety-standards, and at the same time have to agree upon the methods that are allowed for the measurement of these standards.

Global

On a global level, the OECD is the main authoritative actor. Its 'Test Guidelines' (TG) for chemicals form a toolbox of the most relevant internationally agreed upon testing methods, used by industry, academia and governments for the safety assessment of chemical products. When a new test is signed into a TG, the mutual acceptance of data agreement between member countries

ensures that test results have to be acknowledged as valid by the authorities in individual member countries[l].

The OECD commenced a meeting as early as 1996, to discuss aspects of validation studies[g]. It took until 2005 until an official publication was made on the requirements for validation procedures. While very thorough, the document counts 96 pages that represent the consensus of dozens of experts; it is so elaborative that some actors consider it a hampering factor[g]. Not all companies would see the need for such strict procedures, making it difficult to enforce them in practice.

Nonetheless, approval in the form of an OECD TG can be considered the highest available degree of acceptance.

Europe

European public authorities are considered to be in a leading position with regard to their effort to stimulate the development and validation of alternative methods. Implementations of Directive 86/609/EEC on the protection of animals, the Cosmetics Directive and subsequent amendments and the creation of a public research support infrastructure contribute to this position. The path towards regulatory acceptance in the EU starts with a completed validation studies and ESAC endorsement of the method in question[b]. Finally, National Coordinators of EU countries have to decide upon acceptance and state their position towards the method.

United states

Fourteen different regulatory agencies are represented in ICCVAM (ICCVAM, 2010), all of which require animal testing of some sort[f]. The major ones regarding their influence on eye irritation testing are FDA, EPA, to some extent CPSC and OSHA[f]. EPA required most of the Draize tests in their regulations on the safety of chemicals.

The National Institutes of Health (NIH) is the US's national medical research agency, as such it also provides in animal welfare legislation. The NIH Revitalization act of 1993 resulted in the establishment of ICCVAM.

5.7 Animal welfare organizations

Throughout the decades covered by the case study, the role of animal welfare organizations has shifted from simply opposing animal testing and contacting legislators and companies to inform them of this opinion (pre-1980) towards active involvement in the development of alternative methods by supplying research funds (i.e. the ICE funding from Dierenbescherming[g], AAVS funded research on the HET-CAM test[f]). Also, after the success of Spira's campaign, it became clear that mobilizing the consumers of cosmetics products was a lot more effective than trying to persuade regulator. As part of this strategy, animal welfare organizations also try to inform consumers on which companies do and do not use animal tests[f].

The involvement of animal welfare organizations is also clear in the shaping of legislation. Welfare groups lobbied heavily with the European Congress to install the Cosmetics Directive[g,l]. Lobbying for regulations has been more effective in the EU than in the US[f]. Advocacy with legislators also extends to collaborating with agencies that require animal testing, in efforts to minimize the amount of required tests[f].

5.8 SOCIETAL GROUPS

Societal groups form the link between animal welfare organizations and public authorities. Values of cosmetics users, such as whether or not cosmetics should be tested on animals, are voiced to public authorities through animal welfare organizations. At the same time, animal welfare organizations try to influence the public opinion through the creation of awareness on animal testing. Cosmetics users can influence the industry through buying power, and choosing for 'cruelty free' products.

6. Socio-technical analysis II: MLP dynamics

This chapter builds upon the case description and the multi-actor network to analyze the regime dynamics. The general trajectory as proposed by the MLP is followed, where niche developments and interactions in regime dimensions (technology, science, industry, policy, socio-cultural and users & markets) result in the uptake of innovation in the safety-testing regime. First, the niche developments are discussed. For the safety-testing regime, the important rules as stated in table 3.1 in the methodology section are analyzed on the basis of relevance as indicated by interviews or literature sources. By definition, the individual regime dimensions are not static, but interact with each other and these dynamics are indicated throughout the text. Finally, the chapter elaborates on the landscape pressures that have been identified and concludes with a timeline overview.

6.1 Niche Developments

The manifestation of alternative organotypic methods started in 1971 with Burton's publication on a subjective method for scoring eye irritation. In the following years, he developed the EET, which laid the basis for the current organotypic methods. However, the first moderate landscape pressure occurred in 1980, with the strong and successful anti-Draize campaign by Henri Spira[f]. At this time, the EET niche innovation was still underdeveloped: it had only been tested on a small amount of substances, the availability of rabbit slaughterhouses for the supply of the rabbit eyes was not widespread, and authorities did not accept the tests results for formal test procedures[c,g]. Therefore the technology could not take immediate advantage of the disruptions caused by the landscape pressure.

As a result, industry actors that already were part of the safety-testing regime responded by modifying the direction of development paths and innovation activities. They provided the protective spaces that characterize niche innovations. In case of the ICE test this was a dedicated group of researchers at TNO in the Netherlands[c,g]. They started using the EET, and an in-house validation project tried to establish its capability to replace the Draize test. Despite the project's good results, the safety-testing regime based on animal testing did not provide unanimous support for alternatives in these early days. There was no immediate 'rule change' in the regime to start using the EET and there still were conflicts and power struggles with the TNO management to put efforts into the further development of the EET test. This is demonstrated by the fact that external funding from an animal welfare organization (Dierenbescherming) was necessary[a,c,g] for a contract research organization such as TNO to facilitate the development of an alternative method. In addition, little support from the executive layer at TNO resulted in friction between the researchers that carried the niche technology and general management[a,c,g]. So the processes in the niche were still fueled mainly by welfare concerns and researchers' conscience[c,g,h]. Discussions on how to advance the development and acceptance of alternative methods were scheduled a couple of times a year, voluntarily, in the private time and on private costs of a small group of devotees[d,g] (such as Michael Balls, Herman Koeter, Horst Spielmann etc.).

The research subsidized by the Dierenbescherming eventually led to the use of chicken eyes instead of rabbit eyes and provided an even better demonstration of the tests capabilities. This

provided enough proof-of-principle to adjust the TNO management's opinion on the value of the ICE test as a viable alternative to the Draize test[a,g].

Over the same timeframe (1980s) another niche technology, cytotoxicity-based tests, saw a substantial increase in research efforts. The anti-Draize campaign put a possible Draize alternative on the agenda, and it became easier to obtain funding for supporting research projects. Simultaneously, research on cell culture technology was flourishing (Borenfreund and Borrero 1984), many researchers made use of this situation by developing eye irritation tests based on cytotoxicity, using target cells from non-ocular origin. However it was difficult for such tests to compare to the Draize test, because they cannot measure (decrease of) corneal opacity directly (Ubels and Clousing 2005). Organotypic methods are able to do this, and this was reason for Muir (1984) to start with the development of a method to assess the corneal opacity of bovine eyes, which provided the basis for the BCOP test. The niche developments of the BCOP test were guided by a deliberate research effort by Merck to find a test that could screen the toxicology of intermediate drug compounds for occupational hazard purposes[h]. This provided a protected space for the development of the BCOP, because occupational hazard regulations put the responsibility for safety upon the company and do not require data from acknowledged methods (such as the Draize test) in advance[h]. This allowed Merck to use internal validation criteria and the test was soon found good enough for the purpose of assessing occupational hazard. In fact, the results were promising enough for Merck to start validations with a wider range of substances, to allow wider acceptance of BCOP results (Gautheron et al. 1992).

From this point onwards, the advances in technological development of alternatives had not gone unnoticed by public authorities. However, before the ICE and BCOP could be used for formal safety-assessment (i.e. for chemicals and cosmetics), they had to be validated to demonstrate their equivalence to the Draize test. Legislators throughout Europe and the United States acknowledged this, and six large-scale validation studies were started at beginning of the 1990s (Table 4.1). As such, the organotypic methods became part of the safety-testing regime in symbiotic fashion: they were known by (and in some cases already used by) regime actors, but cumulative adjustments and reorientations of regime rules would have to occur before the methods could be fully taken up in the safety-testing regime. These rules and changes are discussed in the next section on regime dynamics.

6.2 Sociotechnical regime rules

Technology dimension

The original Draize test was not based on a technology so much as it was based on a living organism. Scoring was relatively subjective and hardly standardized (Frazier et al. 1987). This led to the articulation of a functional requirement, to standardize measurements for eye irritation. Especially for its major clinical manifestation: corneal opacity. This was known in the niche, and an opacitometer was developed in conjunction with the BCOP test for this purpose[d,h]. It remains an integral part of the BCOP test until today. It is also continuously being optimized, recently Johnson and Johnson published results from a laser-based meter that increased accuracy of the measurements (Van Goethem et al. 2010).

When the first large validation study (EC/HO) concluded that none of the included methods could replace the Draize test, two important recommendations were made regarding technology[d,g]: the inclusion of clear and specific testing protocols for each method (to prevent inter-lab discrepancies) and the specification of an accurate Prediction Model (PM) before entering a validation study. In subsequent studies, these requirements would be improved upon and optimization of both the protocol and PM became routine in the prevalidation of new methods[d,g] (Curren, 1995).

Another functional requirement comes from the pharmaceutical industry, where new compounds that are manufactured by R&D are usually only available in small quantities. A significant benefit from *in vitro* tests is that only a minimal amount of the compound is necessary when compared to an animal test[e,j], stimulating both the use and development of alternative methods for safety-assessment by the pharmaceutical industry.

Finally, the original Draize test evaluates eye irritation potential for a period up to three weeks[a]. This enables the study of reversibility of eye injuries. This criterion is considered important, because realistic exposure of irritating substances to human eyes involves short exposure periods after which the substance is washed out and injuries could be reversed[a]. Currently there is no method that can simulate recovery after eye injury, as in the Draize test, and modeling such a response without the use of a living organism is viewed as a barrier[f] (Piehl et al. 2011).

Further *rules* in the *technology* dimensions remained very stable and the ICE and BCOP methods as described in the OECD TG's basically have not changed over the years[c,d,g,i]. However, the BRD's from ICCVAM that formed the basis for OECD approval of the BCOP test did formulate additional functional requirements: the inclusion of histology as an endpoint and improvement of the sensitivity of the opacitometer (ICCVAM 2006a). These could potentially allow for the use of the BCOP test to assess mild irritants, and are currently being investigated by Johnson & Johnson[e,j].

Science dimension

Scientific inquiry has played a large role in the validation of the ICE and BCOP test, and other alternative methods. For illustration purposes, figure 6.1 below shows an ongoing increase in the amount of publications relating to alternative methods for the Draize test. The graph includes publications on new methods, improvements to existing methods, validation studies, reviews, etc. and is not limited to the BCOP and ICE test, because of limitations to the discriminating power of search strings. The increase shows the intensified research efforts up until (and most likely also after) OECD approval. Current research usually takes a tiered-testing strategy as starting point[b].

The main formal validation programs have been discussed in the case description (see table 4.1 for the overview of studies and results). For the science dimension, the most important changes in *rules* are improvements inherent to the design of validation programs, shifting organizational responsibilities for the management en execution of validation programs, and availability of funding for validation studies.

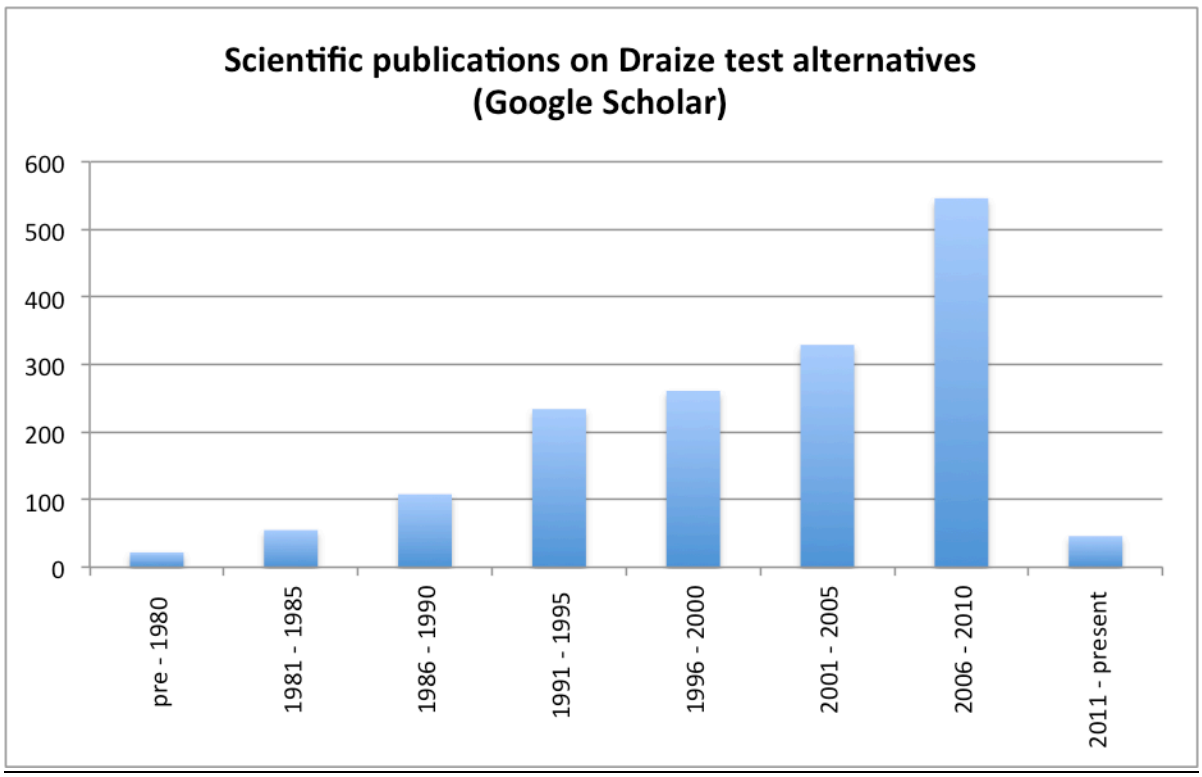

FIGURE 6.1: SCIENTIFIC PUBLICATIONS ON DRAIZE TEST ALTERNATIVES (SOURCE: DATA FROM GOOGLE SCOLAR, USING THE STRING [DRAIZE <AND> "ALTERNATIVE METHODS"])

IMPROVEMENTS IN VALIDATION STUDIES

After the incorporation of Prediction Models and specific protocol guidelines into validation studies as recommended by the EC/HO, the follow-up COLIPA study can be considered an improvement, but the results were similarly inconclusive. Three tests (FL test, RBC assay and TEA) showed 'promising results' (the ICE and BCOP were not included in the study), but did not match the Draize data sufficiently throughout the entire irritation scale (Gettings et al. 1996). The CFTA study took into account the variability of the results obtained by the Draize test, which was striking, but also did not conclusively validate a replacement test. The same holds for the other large validation studies that were performed during the 1990's. Two general observations were made throughout the studies that would later provide the basis for tiered-testing strategies (Balls et al. 1999):

- Different statistical methods, in particular multivariate analysis as opposed to linear regression, early on revealed that better prediction models could be developed based on combinations of *in vitro* endpoints[b,l]. The specialization of 'biostatistician' to properly analyze results was underdeveloped in the EC/HO study, but would be recognized later as a requirement in pre-validation studies[d].

- The predictivity of *in vitro* methods appeared to vary according to (1) the type of material being investigated and (2) the chemical class (low/severe irritant) of respective materials.

This led to a major shift in the way that actors, in particular public research support actors, thought about validating alternative methods[b,l]. It was becoming very clear that under current

circumstances, it would be impossible to find a single replacement for the Draize test. By recognizing the strengths and weaknesses of individual methods, the prediction models and protocols could be improved to obtain better predictability, albeit for a smaller group of substances[b,l,i]. For instance, the BCOP predicts severe irritants much more accurately than mild irritants[e,j,h], and its relative predictability increases when the false-positives/negatives in the lower irritancy range can be disregarded and accounted for with a different test; one that specifically predicts milder irritants. A tiered-testing strategy would use multiple tests to replace the Draize test, one for the mild irritancy range and one for the severe irritancy range. Validation of tiered-testing strategies could be performed retrospectively, i.e. with data already available from previous studies, eliminating the need to perform additional tests[b]. Since 2000, virtually all research efforts have taken tiered-testing into account (by specifying the irritancy range for which a method is suited); both for the development of new alternatives and the validation of existing ones.

Organizational responsibilities for validation studies

The multi-laboratory validation studies of the 1990s had in common that each was a major logistical undertaking, requiring a dedicated management team to ensure continuation[b,d]. While the later studies did take into account some of the findings and recommendations available from earlier studies, there was a considerable amount of difference in protocols, evaluated compounds, included test methods, and statistical approaches between the different validation studies (Table 4.1). Some of these can be considered slight improvements from experience. Other differences only result from the fact that inevitably, many choices have to be made when designing and managing a study. Differences in opinion and lack of coordination (with earlier/simultaneous studies) then lead to slight discrepancies between studies, for instance in the type of test compounds used. For each individual study this does not have to be a problem, but it does pose challenges when trying to compare data afterwards[a,c,f]. Moreover, it decreases efficiency because of partial duplicate work that results from entering multiple validation studies with the same test methods. All major validation studies in the 1990's overlapped in timeframe. There also was substantial overlap in the test methods included, and studies had a region-specific character (i.e. BGA/BMBF in Germany, COLIPA in Europe, CTFA and IRAG in the U.S., MHW/JCIA in Japan). With hindsight, improvements could be obtained with (even) better coordination between national validation organizations[l]. The role of ECVAM in this respect has evolved over the years since its establishment in 1991 and its American counterpart ICCVAM followed five years later in 1996.

While ICCVAM and ECVAM had always cooperated to some extent[l], they also had their differences with regard to requirements for alternative methods and for validation practices. These are often the result of different political decisions by authorities and the *policy* dimension is an integral part of this problem[b,c,l]. In the end, the regulatory bodies for instance decide upon the acceptable amount of false positives that still constitutes as 'safe' in validation studies[b,l]. Because clarity on requirements for validation studies is very important in the design of a study, the policy and science dimensions are closely linked and this point is discussed in the science dimension.

To advance the status of alternative methods for eye irritation testing, a thorough review was carried out in 2004 and subsequent retrospective validation studies were organized in a joint effort between ECVAM, ICCVAM, COLIPA and industry[b]. The resulting 'expert panel report' (ICCVAM 2006b) and 'background review documents' (ICCVAM 2006a) had been composed partly

for U.S. Federal agencies, but also anticipated on OECD requirements for successive endorsement by the OECD. In the end, this facilitated the OECD procedure substantially and the time between submission of the results and OECD endorsement was relatively short compared to the same procedure in other cases (ICCVAM 2010).

The most prominent parties involved meanwhile have acknowledged the importance of collaboration. On April 27, 2009, Canada, the European Union, Japan, and the United States signed a Memorandum of Cooperation (MOC) on International Cooperation on Alternative Test Methods (ICATM) (Wind et al. 2010). The agreement provides for enhanced cooperation and collaboration between four national validation organizations: ICCVAM, ECVAM, JCVAM and the Environmental Health Science and Research Bureau within Health Canada. Tests developed under the memorandum are expected to be more readily accepted by regulatory agencies by assuring international agreement on the scientific information demonstrating that the methods are reproducible and able to accurately identify product related health hazards (Wind et al. 2010). However, cooperation is not only the result of experience with earlier studies. The 7th amendment to the Cosmetics Directive can also be considered a driving force behind the coalition, as becomes clear from recital 10 that specifically encourages the recognition of alternative methods by non-member states (EC 2003). In practice, such recognition is achieved through collaboration with ICCVAM and to a lesser extent JCVAM (Bottini, Amcoff, and Hartung 2007).

EVOLUTION OF FUNDING FOR SCIENCE

Availability of adequate funds for research programs in a crucial *rule* in the science dimension. During the 1990s, in the first multi-laboratory validation studies it was the responsibility of the appointed management team to obtain funding for the study[c,d]. The manufacturers' associations and public authorities were able to free funds from allocated research budgets in most cases[d]. The emergence of ECVAM changed funding expectations and procedures, and for a while it was understood that ECVAM would arrange funding, at least for EU based validation studies[b,c,d,l]. This occurred in the period between the 1990's validation studies and the 2004 review that resulted in tiered testing strategies. No validations studies for eye irritation took place in this timeframe, but in other areas of toxicity testing, validation continued. A study on alternatives for skin irritation testing was funded with 800.000 Euro[b]. Such expenses were unsustainable for ECVAM (financed by the EC) and led to restructuration in the Joint Research Centre which ECVAM is part of[b]. In more recent studies, the prospective study on Human Tissue based models for eye irritation for example; finances are shared between ECVAM and COLIPA[b,l]. However, private/corporate labs are expected to cover their own expenses, while public labs to receive funding[b,c].

Formal regulations and elaborate demands for validation protocols have also resulted in increased expenses for validation studies[b]. In combination with the rapid introduction of new test methods (in area's other than eye irritation), validation procedures cannot keep up the pace, making it increasingly difficult to obtain funding for studies[b] (Bottini, Amcoff, and Hartung 2007).

INDUSTRY DIMENSION

There have been several major changes in *rules* in the *industry dimension* over the years, often intertwined with processes in the *policy* (guidelines and regulations) and *socio-cultural* (consumer values) dimensions that forced realignment of business practices. Before 1980 the industry performed safety tests such as the Draize test as required by law (Zurlo, Rudacille, and Goldberg

1994). The cosmetics industry operated under the *rules* that emerged under the Lash-Lure incident, where safety testing had become mandatory and animal tests were the norm. The chemical industry was in a similar situation where it was obligated to label products containing harmful substances, which also included eye irritation potential.

This regime was disrupted in 1980 through landscape pressure from the anti-Draize campaign and the resulting protests. Especially the cosmetics industry was impacted, as this also was the most prominent, with the pharmaceutical and chemical industry remaining under the radar of the general public. Changes in the public opinion led to changes in consumer demand[f,l] and the industry adapted by at least acknowledging the importance of animal-free testing procedures. Research programs on alternative methods were created and funded by large cosmetics companies in order to repair brand and corporate image. Because of the anti-Draize campaign fueling this process, there was an incentive to develop alternative methods to replace the Draize test, which contributed to the exceptionally large amount of alternatives[c,h] for eye irritation testing that would be developed in the upcoming years and that opened up the regime for existing niche technologies.

The public concern for cosmetics products that were not tested on animals also created marketing opportunities for companies like the Body Shop, which led to a slight change in product labeling practices, where variations of the sentence 'this product is not tested on animals' became custom. Nonetheless, *rules* regarding product-testing practices shifted more on the level of internal company procedures for testing, than for regulation on a formal level, because no alternative method(s) had been developed and validated to such an extent that they could fully replace the Draize test. So in many cases, formally, the Draize test was the only sufficient means to meet safety requirements. This resulted in the ICE and BCOP test coexisting with the original Draize test in symbiotic fashion. Cosmetics companies for instance would use the ICE or BCOP test for internal screening purposes to obtain data that would not have to face the scrutiny of regulatory authorities[e,j]. Cost advantages are a major stimulus for this adoption[e,f,j]. Also, CROs would perform the tests alongside Draize tests that had to be carried out anyways, to obtain comparison data that can be used for validation[a,c,h].

Finally, in practice, from the first years of the new millennium and onwards, data from the ICE and/or BCOP test were being accepted by several European Competent Authorities, particularly for the identification of corrosive/irritating substances (Eskes 2010). Even though not formally validated, the usefulness of the *in vitro* methods was recognized by several national regulatory agencies and within the industry itself, for specific and limited purposes. The confidence for such in-house use varies between companies, and is dependent of historical information on similar materials, an understanding of the method's limitations and characteristics, the availability of appropriate benchmarks and the technical expertise of the user (Eskes et al, 2005 (EC chapter 3)). Therefore, such use of *in vitro* methods was, and still is, company-specific, and often more related to the testing of finished products, especially in the case of cosmetics companies.

CONFIDENTIALITY OF INDUSTRY DATA

From the moment it became clear that Europe was going to push through the marketing ban for cosmetics containing ingredients that are tested on animals, in the 6th and 7th amendment to the Cosmetics Directive, foreign companies realized it would be unavoidable to invest in alternatives if they wanted to bring products to the European Market (Hartung 2008a). Most large

multinationals foresaw this problem and started investing in alternatives for in-house use[b,f,l,l]. However, these methods are proprietary and can be considered a strategic advantage because they allow for faster/better screening of new compounds[b,l,k]. Besides the methods themselves, data for specific methods on new compounds holds a competitive advantage as well[d,h,i]: in the first place because the company knows more about the compound itself, but in the second place because it provides a data point in the (internal) validation of the alternative method. Effectively, the company gains experience with the alternative method that could prove valuable in future testing (i.e. with similar compounds). In turn, industry is often reluctant to make such research data available[d,h,i] and valuable information can already exist without it being available to third parties. This rule impacted the BCOP test when it was being developed by Merck, and company lawyers would not allow the publication of the chemical structure of the compounds used for validation[h].

HARMONIZATION IN SAFETY STANDARDS BY GLOBAL INDUSTRY

While the requirements that result from the European legislation are a strong driving force for the acceptance of alternatives, there is an 'other-way-around' phenomenon especially visible in the chemical industry. The non-acceptance of novel methods in other parts of the world is a stumbling block for those who now have to perform traditional methods for some countries and alternative methods for others[b,l].

POLICY DIMENSION

Legislation has played a very large role in the development and validation of the ICE and BCOP test, and in the safety-testing regime in general. All *formal* changes in *rules* have been discussed thoroughly in the case description and figure 6.2 shows a timeline of the emergence and adaptation of relevant legislation[19]. This analysis focuses on the effects of these laws and regulations, demonstrating the interplay with other *dimensions* and *rules*.

SPECIFIC EYE IRRITATION POLICY / REGULATIONS

The foundation for the *policy rules* in the safety-testing regime was strongly established during the 1960's and early 1970's, with the incorporation of the Draize test into important FDA, EPA and CPSC guidelines for product safety (Ubels and Clousing 2005). This was effective in the sense that any large catastrophes such as the Lash Lure incident have not occurred with the newer guidelines in place. In turn, this is part of the reason why legislators have had faith in the results that were obtained with the Draize test and were dismissive of alternative methods with data that did not match those results 100%[f,,h,i].

Most influential has been the marketing ban, first introduced in 1993 with the 6th amendment to the Cosmetics Directive. Intended to turn into effect as of 1998, it was the main driver behind the ongoing efforts to validate alternatives to the Draize test (and also alternative methods in general). It also is the reason that Europe has advanced more in the field of *in vitro* toxicology than the US[d,f,h,l,k]. However, despite the legislator's intent to ban animal testing, the requirements for validations were underestimated[c,d] and in the case of eye irritation the validations failed. Other areas of toxicology testing showed similar progress: alternatives were available, but validation was

[19] Confirmed in during interviews[c,g]

not completed before the 1998 deadline of the marketing ban. In the end, the 6th amendment to the Cosmetics Directive only led to two postponement on the already foreseen phasing out of animal testing, on the grounds that 'insufficient progress had been made in developing and validating alternatives to the animal tests used for assessing cosmetic safety'(EC 2000), to discontent of legislators and animal rights groups. It was followed by the 7th amendment to the cosmetics directive in 2003, which is a unique legislation in many ways, for example, it phases out essential safety tests before alternatives are available. Recital 5 of the directive reads: "*Currently, only alternative methods which are scientifically validated by ECVAM or the OECD and applicable to the whole chemical sector are systematically adopted at Community level. However, the safety of cosmetic products and their ingredients may be ensured through the use of alternative methods which are not necessarily applicable to all uses of chemical ingredients. Therefore, the use of such methods by the whole cosmetic industry should be promoted and their adoption at Community level ensured, when such methods offer an equivalent level of protection to consumers*." This clearly indicates that the legislator wants to uncouple this industry from the possibly slower progress in other areas of application of alternative methods (Hartung 2008b).

Regional differences between legislators

It was not always clear beforehand, which results would be necessary for different regulatory bodies to approve of a new method. The BMBF study for instance, led German authorities to approve the HET-CAM test for certain (R41) classifications of chemicals as early as 1992[d], but other countries did not follow. A review by the EC in 2000 identified that many European countries already accepted data from the ICE or BCOP test for specific purposes, based on their interpretation of validation study results (Zuang et al. 2002). Because of that finding, harmonization has taken place between member countries and the retrospective validation studies have been performed[b]. Steps are being taken to minimize regional differences through better cooperation in future studies.

Understanding and consensus on scientific practices by regulators

Regulators in US throughout 1990s regarded result from alternative methods more as 'research' and not comparable to a Draize test[h]. Education of regulators is viewed as a hurdle that, if overcome, would benefit to the acceptance of alternative methods[f,h]. For ICE/BCOP this is still relatively simple (i.e. eye for an eye) but for more complex systems it would be good to remove the 'black box' idea of an alternative method that some substances are tested and data comes out, without really understanding what happens in the process[k]. The view that the Draize test is 'the thing that works' is still prevalent under some regulators in the US[f]. In Europe, regulators also are not always ready to accept scientific arguments over politically driven values[b] (i.e. the acceptable amount of false positives).

REACH legislation on chemicals

The new chemical legislation REACH represents another incentive for introducing alternative methods, at first glance addressing all the toxicological effects relevant for cosmetics. The situation for cosmetics, however, is very different to the one of REACH: While REACH would benefit from any reduction or refinement of animal tests and could stand areas without replacements, here, with an even shorter deadline, full replacement has to be achieved for some selected endpoints including eye irritation. In contrast, animal testing for REACH will start mostly after 2015 (Williams, Panko, and Paustenbach 2009). It is evident that the short timeline only

allowed for existing methods to be validated and did not permit any new developments. Indirectly, REACH undermines the ban on cosmetics testing, because many ingredients used in cosmetics are tested as chemicals if the annual tonnage of the substance is above 1 ton (Ogden 2010). REACH in certain circumstances requires animal tests for chemicals, and, cosmetics ingredients are chemicals, hence, REACH requires animal testing for cosmetics ingredients. However, the Cosmetics Directive prohibits animal testing of cosmetics ingredients 'for the purpose of this Directive'. While REACH promotes the use of alternatives wherever possible, it does mandate animal tests[a], also for endpoints that are likewise required for cosmetics testing including eye irritation. The necessity for a complete replacement for the Draize test is therefore even greater in the period of REACH implementation and provides a significant impetus for continuing development of alternatives.

OECD TESTING GUIDELINES AND ITS IMPLICATIONS

A misunderstanding by national legislators about the OECD TG's creates problems in the regulatory implementation of validated methods: The TG's represent a 'toolbox' of standardized methods, they do not represent the mandatory inclusion of a specific test method in national legislation of OECD member countries. But if Europe would require an *in vitro* test for eye irritation in its regulation, the inclusion of the test in a TG allows companies outside the EU to carry out the test in any OECD member country, and it would still have to be accepted in Europe[g]. However, member countries are not obliged to start requesting data from this eye irritation test.

In agreement with the OECD Council decision on Mutual Acceptance of Data (MAD) from 1981, data from OECD TG's should be accepted for purposes of assessments in all OECD member countries (OECD 1981). Still, when member countries have not had the opportunity to include tests in their own country, they sometimes use that as a reason to block OECD acceptance[d]. Instead member countries should be much more relaxed about the adaptation of the methods described in the TG, because TG's do not prescribe individual regulations, but only serve to agree on the right way to use a test. The fact that such discussions takes place about tests that are not always applied in the respective national regulation, demonstrates that legislators recognize the importance of global standards and possible impact on national legislation, as has been the case with the ICE and BCOP test after their implementation in the OECD TG[g].

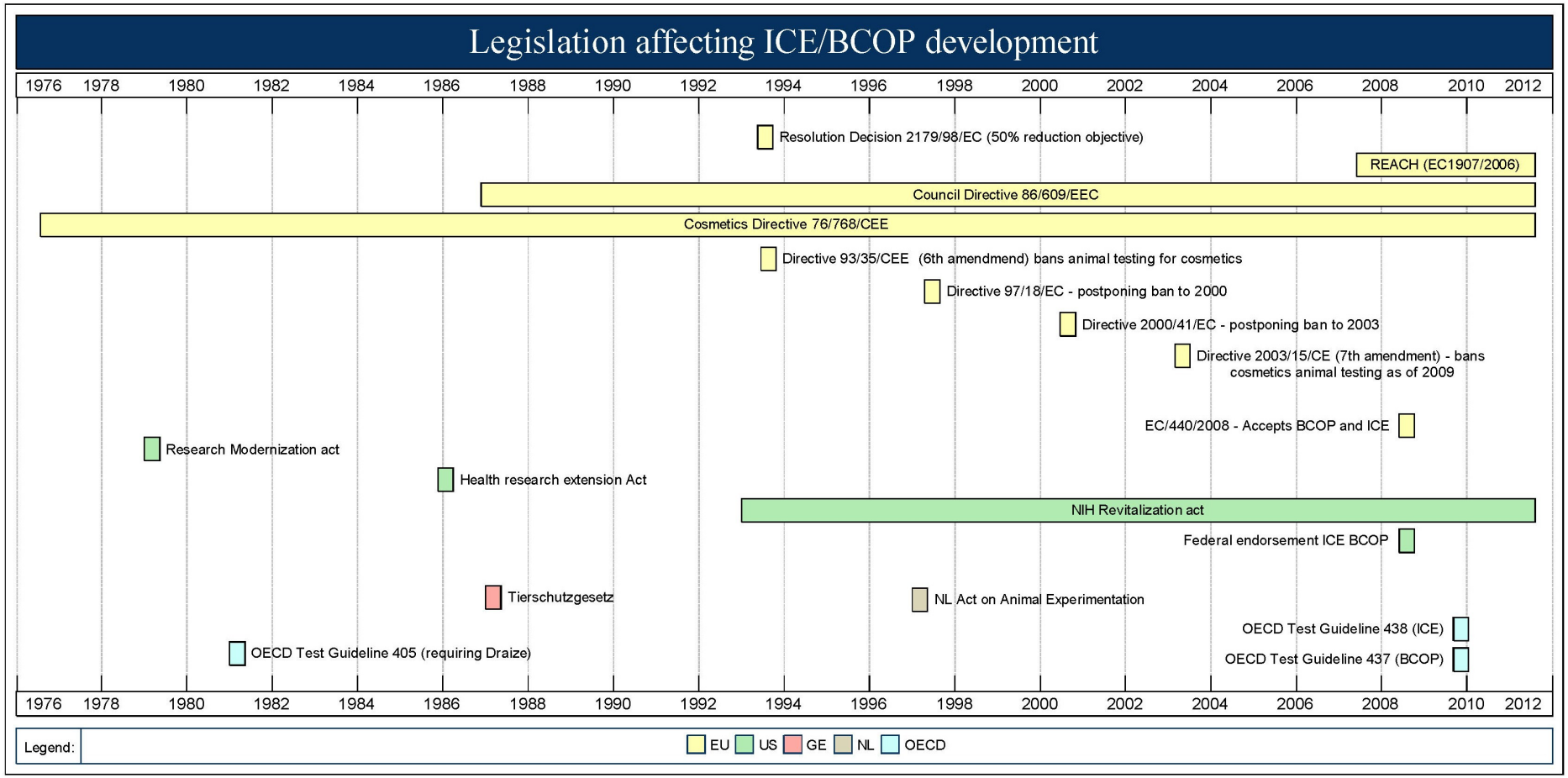

FIGURE 6.2: LEGISLATION AFFECTING ICE/BCOP TEST DEVELOPMENT

Socio-cultural dimension

The *socio-cultural* aspects surrounding the Draize test have been very influential in the development and validation of the ICE and BCOP test, and also played an important role in the shaping of the public opinion on animal welfare in general. The anti-Draize campaign is considered an external landscape pressure (and is discussed in the section on landscape developments) which has embedded certain values in to the regime. These gradual changes in *rules* have occurred over time, and are discussed in the remainder of this section.

Evolution of the public attitude towards animal experimentation

Several studies have been concerned with the public attitude towards animal experimentation. Many people generally approve of animal research for human benefit. In 1989 a study found that the majority of U.S. citizens agreed that the use of research animals progresses the medical development (Harvey and Shubat 1987). In Europe, higher levels of opposition were found with over 50% of the population being opposed to animal research in the majority of EU countries (Pifer and Shimizu 1994).

Table 6.1: Trends in public opinion on animal experimentation (Source: vonRoten 2009)

Evolution of the Percentage of Acceptance of Animal Experimentation in European Countries*

Country	*2001*	*2005*	*Trend*
Spain	56	59	3
Portugal	66	54	–12
Belgium	45	52	7
Denmark	67	51	–16
Greece	67	51	–16
Finland	53	49	–4
Netherlands	46	45	–1
Italy	43	40	–3
Germany	41	40	–1
United Kingdom	43	40	–3
Sweden	50	40	–10
France	40	37	–3
Ireland	41	36	–5
Switzerland	48	35	–13
Austria	37	33	–4
Luxembourg	38	30	–8
EU15	45	43	–2

Table 6.1 above demonstrates a slightly declining trend in the acceptance of animal experimentation in European Countries over the period 2001-2005, based on a Eurobarometer survey in the EU (von Roten 2009). Data from the same survey in 2010 did not differ significantly from the 2005 data in table 5.1. (EC 2010b) Similar to the other previously mentioned studies, the survey relates to research on animals for the purpose of 'resolving human health problems' and on the basis of the Eurobarometer it can be concluded that the public opinion on animal

experimentation (for resolving health problems) has remained relatively stable over the 2001-2010 period. However, when testing for cosmetics and household products these percentages usually drop considerably. A study in 1990 ((Freeman and Ward 1990) found that one-third of the respondents tolerated the use of animals for such purposes. In a 1999 poll for the New Scientist, a British organization (MORI 2010) found that 38% of respondents condoned the testing of cosmetics on mice for allergic reactions.

These data are in line with observations regarding the alternatives for eye irritation, where cosmetics (and not pharmaceutical) companies were influenced by public pressure. Even after the anti-Draize campaign, the public opinion is still regarded an important driving force behind validation efforts by many different actors[b,d,f,h,l]. For a great part, this has to do with the regulations that have been put in place as a result of welfare and lobby groups asserting this opinion towards legislators [b,d,g,h,l,k,l]. The Cosmetics Directive, and its amendments in particular, are believed to reflect the cultural value that safety testing for 'trivial' cosmetic products does not justify the accompanying animal suffering[i,k].

VALUE OF EYE SIGHT AND EMOTIONAL RESPONSE TO EYE DAMAGE

The feelings that humans associate with impaired vision and eye damage have played a remarkable role in the history of the Draize test. Of course, the initial reports on the Lash Lure incident containing explicit graphic material of blistered eyelids paved the way for the pre-1980 safety-testing regime. Eyes are considered delicate and sensitive organs, damage in the form of impaired vision or blindness can have a profound impact on one's future and is often irreversible. Hence, people respond emotionally aversive to prospective eye injuries, much stronger when compared to skin damage for instance[a,b,c].

A similar response was used decades later in the anti-Draize campaign, showing severely injured rabbit eyes to generate public attention. Meanwhile, exactly because of toxicity testing, cosmetics safety had increased substantially and could be relied upon by consumers, which allowed for the opposition to the Draize test. The assumption that cosmetics would remain safe if animal tests would be abolished was implicit in this case.

Our caution with regard to eye damage would later (and still) prove to be an issue in the validation of the ICE and BCOP test. When validation studies led to the comparison of data between different tests, regulators had to decide upon the percentage of false-positives and false-negatives that would be acceptable. Despite the (known) shortcomings of the Draize data, regulatory representatives up until today would require a 0% false negative rate[b] (i.e. harmful substance inadvertently classified as safe) for alternative methods in validation against the Draize test, with no restrictions on the false positive rate[20]. This complicates the validation of alternatives considerably, and while many experts agree that a higher (but still conservative) false negative rate would be scientifically favorable, it appears difficult to prevent regulators from yielding to an emotional response of 'safety-first"[b,h,l].

[20] Compare for instance the false positive/negate rates of 20%/20% for skin irritation alternatives, which have followed the validation process much more quickly.

Users, markets and distribution networks dimension

Rules in this dimension have arguably changed least visibly, but do have some profound implications for both the validation of the ICE and BCOP test and implications for the safety-testing regime in general. With industry actors both developing alternative methods and being their only user, this dimensions shows overlap with the industry dimension.

Eye irritation test users

Users of eye irritation test for a long time have remained stable throughout the safety-testing regime. They are described more elaborately in the multi-actor network analysis (i.e. pharmaceutical, cosmetics and chemical industry). Although the pharmaceutical industry represents only a small part of eye irritation test usage, its presence has remained constant over the years. Eye irritation tests are used for occupational hazard screening and pharmaceutical companies such as Johnson and Johnson actively participate in improving alternative methods, as demonstrated by the recent advances in opacitometer technology for the BCOP test (Van Goethem et al. 2010).

The main user of eye irritation tests is the cosmetics industry[a,c,c,g], which is bound by legislation to foresee in the safety of its products. Besides, new substances are often screened in-house before they appear in consumer products. *In vitro* methods are widely used in this case because of the safety requirements are somewhat less stringent, and companies can use internal standards and experience to evaluate innovative methods, without the lengthy process of convincing legislators of the reliability of such methods[b,l,k]. This is not to say that cosmetics companies have lower standards or are careless compared to regulators, on the contrary, the prescreen standards are very high because it is in the best interest of the company to select the right (i.e. safe, non-irritating) substances as early as possible[h,l,j]. Use for screening purposes also takes a form of bureaucracy out of the equation, because methods do not have to face formal validation first. This beneficial for the development of alternative methods, which, when developed, can be submitted to ECVAM or ICCVAM for review when it is deemed suitable or important. For finished products and ingredients thereof, safety has to be demonstrated through officially acknowledged methods[a,c]. The use of eye irritation tests has also remained constant in this case with no major changes in *rules*. Moreover, the acceptance of the ICE and BCOP methods through OECD TG 405 in practice changed little for cosmetics companies[c,g]. This has to do with the fact that both tests are still only validated for the classification of severe eye irritants. While this significantly reduces cruelty to animals, it also means that animal tests are still required for moderate and mild irritants. Obviously, cosmetics should fall under the latter category to begin with, which is why the cosmetics industry has invested considerably in *in vitro* alternatives for the identification of mild irritants, for which two tests (EpiOcular and SkinEthic) are currently being validated for use in tiered testing strategies in a prospective study by ECVAM and COLIPA[b]. Figure 4.1 in the case description demonstrates this problem. Despite the relatively limited coverage of the Draize test, it currently still remains the only (formally) accepted method for demonstrating non-irritancy, indicting the need for validation of EpiOcular or a similar method.

This brings us to the chemical industry as *user* of eye irritation tests, which used to be bound by a range of European Directives covering the manufacturing of chemicals and appropriate safety classification (Williams, Panko, and Paustenbach 2009). Depending of the type of substance, its predicted use, and company and regulative standards, chemical manufacturers would decide if eye

irritation testing would be relevant. The ICE and BCOP test were informally allowed for the classification of severe /extreme irritants early on[g], and their usage continued over the years in steady fashion.

REACH legislation might change this in the near future[a]. The EU decided that at present, we know very little about chemicals from a safety point of view, because in the past, the burden of proof for the (un)safety of chemicals lied upon the authorities (Williams, Panko, and Paustenbach 2009). They needed to demonstrate that a chemical was unsafe, before imposing any restrictions, opposite to the situation in the cosmetics industry (Ogden 2010). As a result, many chemicals have been on the market for many years without being 'properly tested', according to the EU. The introduction of REACH regulates the registration, evaluation and authorization of substances that are manufactured or imported into the EU and puts a duty on industry to actively collect information on the safety of chemicals, also for substances that are already on the market. This means that thousands (Hartung 2010a) of substances will have to be tested for eye irritation in the upcoming years, substantially increasing the share of the chemical industry as users of eye irritation tests, both *in vitro* and *in vivo*. REACH does allow for the use of alternative methods, but does not *require* them in the way the Cosmetics Directive does, so animal tests will have to be used when no alternatives are available[a]. This includes the lower ranges of eye irritation, for which methods are currently under validation, but not formally adopted yet.

Markets and distribution networks

Markets and *distribution networks* have adapted under the macro level *landscape* pressure of globalization, which is both a driver and a barrier for alternative methods (Bottini, Amcoff, and Hartung 2007). The harmonization of approaches to the development and validation is a prerequisite for international trade markets and change gives opportunities to update such approaches (driver). Simultaneously, multinational companies will use animals for safety-testing until the last national market has updated its regulations[b,l] (barrier). Hence, there is a clear role for bodies like the OECD, because they shape the worlds' most important economic areas by setting standards, which are typically followed by the rest of the world.

Processes in such bodies however, are based on slow consensus processes, which can hardly cope with the pace of new technologies. This is part of the reason why the older ICE and BCOP test have gained validation and more modern tests and strategies (i.e. EpiOcular and tiered testing) are still under review[b,f]. In a more general sense, this also holds for other technologies and products (i.e. computational methods, -omics, recombinant products, nanotechnology): as long as our knowledge in the life sciences doubles every couple of years (Popper 2002), it will be difficult to afford standardization, validation and peer-review processes that can span over 10+ years (Hartung 2010b). Advances in global distribution networks also are far ahead of a global consensus on regulatory requirements. Perhaps because of this, it is unsurprising that globally operating companies have contributed a lot to the progress that has been made with alternatives for eye irritation testing. While also being a matter of capacity and capability, the perceived vested interest for global distribution cannot be ignored.

6.3 Landscape pressure

Two distinct landscape pressures have been identified, these are societal pressure and globalization.

Societal pressure

As has shortly been mentioned in the niche analysis, the first moderate landscape pressure emerged in 1980 with the anti-Draize campaign by Henry Spira. While animal rights activists already had targeted the Draize test on ethical grounds in earlier years (Wilhelmus 2001), Spira succeeded in influencing the public opinion through the effective use of mass media. He was the first to effectively mobilize the public opinion[f], voicing protests and demanding a solution to abolish the Draize test. The New York Times ran two consecutive full-page advertisements, and advertisements in other newspapers followed, along with international publications in Britain, Canada and Australia. Revlon, the company that was targeted in the ads gave in to the increased public pressure weeks later, by donating $750,000 over three years to Rockerfeller University (CAAT 2011). This, in effect, transformed the search for alternative methods from a minor antivivisectionists' issue into a large-scale operation that was supported by a multi-million dollar corporation and linked to a respectable research institute (Sina and Gautheron 1998). Public awareness on the experimental use of animals in the manufacturing of cosmetics was raised internationally, and resulted in a reconfiguration of the safety-testing regime, allowing the alternative (organotypic) methods from the niche to be gradually incorporated into the regime. Through the targeting of consumer demand[f] (instead of companies or regulators directly) the societal pressure also went from being an exogenous pressure to becoming an integral part of the safety-testing regime over the following decade[21]. Cosmetics companies now had to take into account animal welfare concerns of their consumers in their operations and the changes in values were translated into legislation in 1993 with the 6th amendment to the Cosmetics Directive.

Globalization

The fact that national boundaries are losing importance, i.e. globalization, is currently felt in many fields (Castro Díaz-Balart and Pérez Rojas 2002; Hopp 2002), including the field of alternatives[k]. The pressure is coming from disparities between the rate at which globalization is taking place science and markets (fast adaption to globalization) and regulations (nationally-bound, little globalization). Science is global and knows no boundaries. Progress in the safety-testing regime is very much based on scientific progress and development[j], which can provide new alternative methods and quickly spread knowledge of their value and limitation in the use for safety-assessment (by scientific standards). Similarly, market forces increasingly have a global reach with the establishment of growing multi-national companies and communication and transportation systems (i.e. internet and global postage) that allow the retail of from any place in the world (Jennings 2006). However, in legislation there is a 'compartmentalized' approach in national regulations, where methods can remain 'frozen' in guidelines that cannot follow the pace of science because political decisions on consumer safety that have to be made on a national level (Bottini, Amcoff, and Hartung 2007).

[21] This is also why the public opinion from 1990 and onwards is discussed in the socio-cultural dimension of the safety-testing regime.

6.4 The MLP transition pathway

Summing up, the uptake of the ICE and BCOP test into the safety-testing regime has been analyzed through the MLP. Figure 6.3 displays a timeline with the most important rules that have been discussed in the analysis. With regard to the transition pathway that has been followed, it can be concluded that the safety-testing regime has undergone a transformation path: moderate landscape pressure in the form an effective anti-Draize campaign happened at a moment when organotypic niche-innovations had not been sufficiently developed to fully substitute the Draize test. The safety-testing regime was disrupted enough for regime actors, industry and public authorities in particular, to pick up on the niche innovation. They responded by modifying the innovation activities that were needed to get organotypic methods validated for use in formal regulatory safety-assessment. From the moment that the first validation studies took place, the ICE and BCOP test have coexisted in symbiotic fashion next to the Draize test. Cumulative adjustments and reorientations in the safety-testing regime had to take place to further validate the ICE and BCOP test: legislation was created that forced the development of alternatives methods. Industry in collaboration with public authorities and public research support actors started extensive validation programs, which have improved considerably through learning, by including prediction models, protocols and a tiered testing strategy. Under the landscape pressure of globalization, the EU and US saw increased collaboration that resulted in a retrospective validation study, which paved the way for US federal endorsement and OECD acceptance of the ICE and BCOP test.

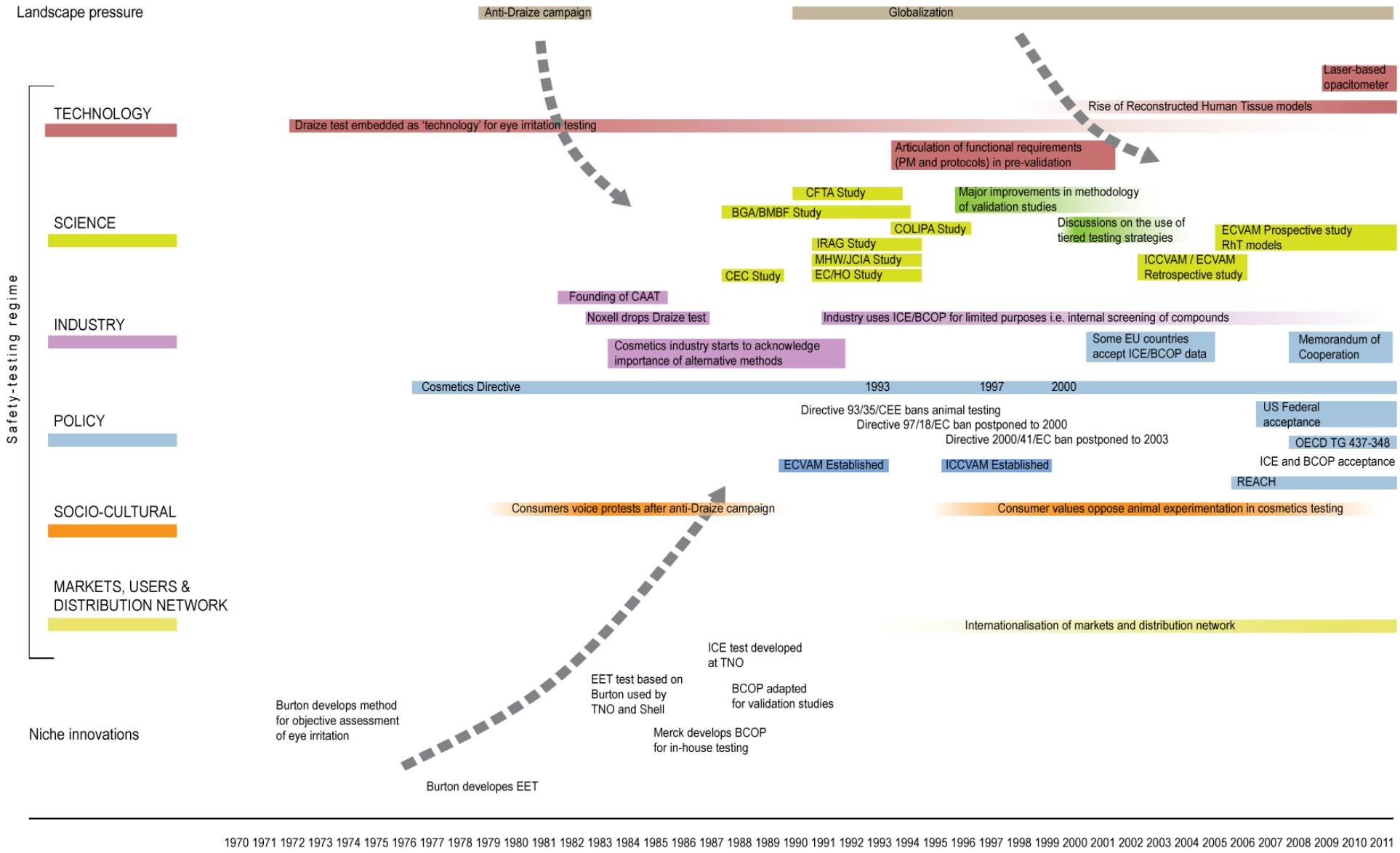

FIGURE 6.3: MULTI-LEVEL OVERVIEW OF THE SAFETY-TESTING REGIME

7. Conclusion

In order to better understand innovation and transitions in toxicological safety testing, this research has studied the development and validation of the Bovine Corneal Opacity and Permeability (BCOP) test and the Isolated Chicken Eye (ICE) test in the context of the ongoing transition in the field of safety testing towards the use of alternative (non-animal) models. The multi-level perspective (MLP) was applied as an analytical framework for its ability to examine complex, long-term transitions involving many different actors. It was used to answer the main research question:

W*hich sociotechnical factors influenced the successful technological development and regulatory validation of the BCOP and ICE test methods for replacing the Draize eye irritation test in the field of toxicological safety testing?*

The results are based on comprehensive review of existing literature and interviews with international experts in the field of eye irritation testing. The sub-questions '*what technologies are used in the BCOP and ICE test methods?*' and *'how were the BCOP and ICE tests developed and validated'* have been answered in detail throughout chapter 3. The BCOP and ICE test can be considered as organotypic methods, using isolated organs of animals to assess a toxicological endpoint representative of eye irritation potential. They are the least technologically complicated when compared to the other categories: cytotoxicity (i.e. using a target cell of non-ocular origin) and reconstructed human tissue (i.e. using human-derived cells) methods. The ICE test was developed at the Dutch TNO in a research project aiming to improve the Enucleated Eye Test[22] (EET). The BCOP resulted from a research project at Merck to develop an *in vitro* test for occupational hazard screening of intermediate pharmaceutical compounds. Public authorities and manufacturers' organizations in cooperation with industry, legislators, animal welfare organizations and academia undertook enduring validation projects to establish the ICE and BCOP's equivalence to the Draize test. Over a period of 25+ years, the validation efforts resulted in OECD acceptance of the ICE and BCOP test for the evaluation of severe eye irritants, in the mean while shaping the present-day validation process for (alternative) toxicological safety tests.

Analyzing the results from a multi-level perspective, demonstrated the transition towards the use of alternative methods that has been unfolding over the last decades, in a regime characterized as the 'safety-testing' regime. This regime includes actors and rules[23] pertaining to the development, production, validation, use and regulation of tests for the assessment of possible adverse effects of chemicals on living organisms. The analysis answers the sub-question '*which barriers had to be overcome for the partial replacement of the Draize test?*' and the main research question.

Organotypic methods were created in a niche that emerged from the 1971 publication of Burton on the objective measurement of corneal thickness of exposed rabbits eyes, which led to the development of the EET ten years later. This laid the basis for other organotypic methods. Moderate landscape pressure on the safety-testing regime occurred in 1980 with the anti-Draize campaign, but the EET niche-innovation was still underdeveloped: it had only been tested on a

[22] Another organotypic method developed and introduced by Burton (1981) during 1978-1981
[23] See figure 5.1 for an overview of actors and section 6.2 and figure 6.3 for regime rules.

small amount of substances and supply of rabbit eyes was limited. Therefore the EET could not take immediate advantage of the disruption caused by the landscape pressure. As a result, industry actors that were already part of the safety-testing regime responded by modifying the direction of development paths and innovation activities. For the ICE test, dedicated researchers at TNO provided a protective space, improving upon the EET by using chicken eyes and gathering data on more test substances. Merck developed the BCOP for occupational hazard assessment of intermediate drug compounds, which provided a protective space because occupational hazard regulations do not require the use of validated methods, but leave safety responsibility upon the company. Expectations of both the ICE and BCOP's tests' potential to quickly replace the Draize were high, after small in-house validation tests showed good correlation with Draize data.

Advances in technological development of alternatives had not gone unnoticed by public authorities. However, before the ICE and BCOP could be used for formal safety-assessment (i.e. for chemicals and cosmetics), they had to be validated. Legislators throughout Europe and the United States acknowledged this, and the European Commission and the British Home Office started the first large-scale validation study in 1991. As such, the ICE and BCOP test became part of the safety-testing regime in a symbiotic fashion: they were known by (and in some cases already used by) regime actors, but cumulative adjustments and reorientations of regime rules would have to occur before the methods could be fully taken up in the safety-testing regime.

The safety-testing regime at that time was poised by several problematic *rules:* first, the general success of the Draize test in safeguarding the public from major eye irritation catastrophes, had led legislators to view the test as a 'golden standard'. Second, there was little experience in conducting large validation studies to establish the reliability and relevance of an alternative test, leading to (with hindsight) sub-optimal studies. Regulators decided that any replacement should be as 'good' as the Draize test. This made it difficult for BCOP and ICE to enter the existing safety-testing regime, since fundamental problems with the Draize test itself (subjective scoring in test results, lack of control over rabbit behavior during the exposure period, and unpredictable mechanistic effects when testing for solids) are widely considered by experts to have caused variability in the Draize data itself. In turn, it becomes difficult to validate any alternative using data from a flawed test to begin with. Rules in the science and policy dimension of the safety-testing regime clashed in this regard: legislators were very slow to recognize and acknowledge the problem, demanding a zero percent false negative rate from alternative methods while using the Draize data as standard for validation. Even the 2003 ECVAM/ICCVAM retrospective study included this requirement, despite repeated efforts from ECVAM and industry experts to gain recognition for the Draize test flaws. From a socio-cultural perspective, the problematic Draize test data and the inclination of regulators to adhere to the Draize test as golden standard, were the most important barriers withholding the uptake of the ICE and BCOP test into the safety-testing regime.

Two related sociotechnical factors eventually were key to the continuation of validation efforts: changes in the values of consumers that animal tests should not be necessary for the production of cosmetics were articulated to the European Congress by animal welfare organizations, and effectively turned into legislation with the 6th amendment to the Cosmetics Directive in 1993. The amendment posed strict deadlines to ban the marketing of cosmetics in the EU for products tested on animals. This forced both the cosmetics industry and ECVAM to find solutions that would get alternative test methods for eye irritation validated. The dynamics between the socio-cultural,

policy and science dimension of the safety-testing regime, i.e. consumer values affecting legislation, which in turn influence industry requirements, was repeated when the marketing ban was postponed, but reinforced in 1997, 2000 and 2003 respectively. Eventually the Cosmetics Directive proved to be the main driving force behind the ongoing efforts to obtain validation for the BCOP and ICE test. As such, the Cosmetics Directive realigned the safety-testing regime to allow the BCOP and ICE to be adopted formally.

Partly because of the Cosmetics Directive's driving force and partly because the industry and public authorities had learned from the first validation studies, the methodology of validation studies during the 1990s and early 2000s saw considerable improvements with the introduction of Prediction Models and strict protocols in a pre-validation process. Also, enhancements in statistical methods improved the accuracy of results and helped to shift thinking in the safety-testing regime towards 'tiered testing strategies', where multiple tests are used to cover the entire spectrum of eye irritancy.

In practice, the safety-testing regime already partially allowed data from several alternative methods, including the BCOP and ICE tests. Competent authorities in different countries would accept results for the classification of severe irritants, on the basis of their opinion and interpretation of the validation studies that were performed during the 1990s. A strong country-bias could be observed, for instance with legislators in Germany accepting the German-invented HET-CAM essay over other tests. This partial uptake of the ICE and BCOP test in the safety-testing regime was beneficial for reducing the amount of animals used, but also decreased the motivation for industry to continue formal validation programs.

Coordination of research activities was greatly facilitated by ECVAM (and ICCVAM), in bringing together industry, scientists, contract research organizations and public authorities in workshops and validation studies. Under the globalization of markets and knowledge, multinational companies are operating worldwide and scientific knowledge (for instance on validation) can spread almost instantaneously. Policy and legislation however, are still very bound to regional differences. There was (and is) an increasing need for collaboration between policy-makers on a global scale in order to effectively reach consensus on the requirements of alternatives and validation studies. The past (and current) situation where validation procedures take much longer than the technological development of new methods creates a barrier for getting enough new methods validated, that can only be overcome by optimizing validation procedures and ensuring the mutual acceptance of results by public authorities in different countries. In the case of eye irritation, the joint ECVAM-ICCVAM retrospective study that started in 2005 was exemplary in the sense that it was set up with OECD criteria in mind. Eventually it resulted in US federal endorsement of the ICE and BCOP tests for the identification of severe irritants in 2007. The OECD adapted its Testing Guideline 437 and 438 in 2009 to the same extent, resulting in formal uptake of the BCOP and ICE into the safety-testing regime, being recognized by all 34 member states.

Summing up, the uptake of the ICE and BCOP test into the safety-testing regime has been analyzed through the multi-level perspective. With regard to the transition pathway that has been followed, it can be concluded that the safety-testing regime has undergone a transformation path: moderate landscape pressure in the form an effective anti-Draize campaign happened at a moment when organotypic niche-innovations had not been sufficiently developed to fully

substitute the Draize test. The safety-testing regime was disrupted enough for regime actors, industry and public authorities in particular, to pick up on the EET niche innovation. Regime actors responded by modifying the innovation activities that were needed to get the ICE and BCOP test validated for use in formal regulatory safety-assessment. From the moment that the first validation studies took place, the ICE and BCOP test have coexisted in symbiotic fashion next to the Draize test. Cumulative adjustments and reorientations in the safety-testing regime had to take place to validate the ICE and BCOP test: legislation was created that forced the development of alternative methods; industry in collaboration with public authorities and public research support actors started extensive validation programs, which have improved considerably through learning, by including prediction models, protocols and a tiered testing strategy; under the landscape pressure of globalization, EU and US authorizes increased their collaboration that resulted in a retrospective validation study, which paved the way for US federal endorsement and OECD acceptance of the ICE and BCOP test.

From an innovation perspective this study has demonstrated the large impact that policy, in particular regulative force and decision-making in politics, has on innovation activities in a sector that is heavily regulated to warrant public safety. Policy however can be a dual edged sword: on one side, the Cosmetics Directive certainly stimulated the validation efforts for alternative methods. On the other side, without the very strict regulatory requirements for validation, more innovative methods would be used by industry today, and it is not necessarily clear that public safety would have suffered from somewhat more lenient validation study requirements.

Insights for the advancement of future alternative methods

This study has found many factors that are important for the successful regulatory validation of alternative methods. Some findings have already been addressed (i.e. inclusion of prediction models, definition of protocols, pre-validation, etc.), but other insights could prove valuable to stimulate innovation in the safety-testing regime.

- Education of legislators

Many of the struggles in validation procedures appear to come from the fact that legislators, who are in the position of making important decisions on validation criteria, sometimes have little affinity with the workings of alternative methods and implications that result from not using live animals (i.e. always testing a 'proxy' of what can be measured in animals; eye damage versus corneal permeability). Education of legislators would be beneficial to reduce the knowledge gap that now, sometimes, exists between experts on the science of validation and statistics, and experts on the politics of guarding public safety.

- Introduction of positive stimuli to develop & use alternatives

Virtually all of the regulations and laws that have been discussed in this report are in place to *prohibit* certain things. For example, the Cosmetics Directive bans animal testing, Test Guidelines prescribe which methods to use and inherently prohibit the use of other methods. The success stories in other areas that have been obtained by introducing positive stimuli seem not to have reached the safety-testing regime. For instance, patent expiry dates on innovative substances could be extended, if it can be demonstrated that the substance has been developed solely through the use of alternative methods.

- CREATE AN INTERNATIONAL HARMONIZATION BODY FOR VALIDATION REQUIREMENTS AND CLASSIFICATION SCHEMES

In light of the ongoing globalization, the safety-testing regime would significantly benefit from one central actor that can oversee and manage global validation efforts, while at the same time harmonizing the differences in labeling systems that currently still exist. The sharing of work of validation studies requires a continuous platform that can act with a global voice. The OECD itself can hardly meet the demand for coordinated validation studies, while experiences gained in costly validation studies are too painful to be repeated.

Future alternative methods will further decrease animal use for toxicological safety testing, providing more effective, efficient and economically viable testing procedures in the process. Innovations that are expected to make this possible (reconstructed human tissue models in the short term, genomics and computational modeling in the long term[g]) could benefit substantially from implementation of the points made in this section.

8. Discussion

This section discusses some of the findings and limitations to the research, and provides suggestions for further research.

8.1 Research quality

With regard to construct validity, this research encountered several difficulties, especially with the implicit overlap between regime dimensions in the MLP. For instance, with the cosmetics and pharmaceutical industry responsible for both developing and using alternative methods for safety assessment, in addition to participating in their validation studies, it was difficult to consistently analyze the processes in the single dimensions of *science* (i.e. validation studies)*, industry* (i.e. production)*, users & markets* (i.e. use of alternatives). In addition, the MLP does not prescribe how broad or narrow empirical topics should be delineated. The 'regime' notion is an analytical concept that can be applied to empirical topics of different scopes. The type of scope (i.e. narrow, primary fuels or broad, electricity systems) has implications for regime characteristics (i.e. amount/type of actors, prevalent rules) (Geels 2011). This leads to a degree of flexibility on the part of the researcher, and demarcation of concepts is dependent of the subject being studied. To this end, using multiple sources of evidence and multiple interviews optimized construct validity of this research. The inclusion of interviewees from every actor group ensured a balanced view on the subject and consistency between interview results indicates consensus and validity.

As far as internal validity, the validity of the supposed causal relationships is concerned, the MLP itself does not propose any causal relations to be tested, but employs 'process theory' as explanatory style (Geels 2002). Process theories do not explain dependent variables as being caused by independent variables, but instead explain outcomes in terms of sequences of events and their timing (Langley 2007). Some 'historical causality' is presented in the case study, in the sense that events are often considered as if in some way being agents that can bring about other historical events (i.e. the cosmetics directive causes the ongoing validation efforts) but these are not explained through statistical causality. Such event sequences are supported throughout the case study with multiple expert statements from interviews and/or publications. For major and contemporary events, which are often also documented, this results in adequate support for their internal validity. However, for a case study with a timeframe spanning multiple decades, smaller events that occurred further on in the past (i.e. niche developments) could have been more difficult for experts to accurately recall, particularly with timing and effects between matters that occurred thirty years ago. Again, this is triangulated by interviews as much as possible, but it also poses the most vulnerable part of the research concerning internal validity.

The external validity of the results is relatively low because of the peculiar characteristics of the safety-testing regime (i.e. producers also serve as the only users, absence of 'free market forces' for toxicology tests, the demand for tests is directly shaped by regulations, etc.). With regard to the generalization of empirical findings (i.e. consumer values opposing animal testing are reflected in legislation and in turn cause market reactions forcing the continuity of validation efforts) one should be very cautious, and a compelling case can only be made for generalization within (very) similar areas such as other toxicology tests.

The final quality criterion is reliability, which can be established by ensuring that a replication of the study by another investigator would also replicate the obtained results. Several precautions have been taken to optimize the reliability of this research; the first is the documentation of literature sources throughout the study. Because of the 'pioneer' status of eye irritation alternatives, comparably many articles have been published on the subject. This includes review documents and annotated historical accounts. As a result, the factual events gain in reliability. Secondly, to determine (historical) causality and dynamics, this research leans on expert interviews. The views represented in this research are combined from those of respectable experts, many of whom dedicated an entire career to *in vitro* toxicology testing. There was a good triangulation of interview results for many of the more 'obvious' findings (i.e. influence of the flaws in the Draize test data). For several of the more 'subtle', actor-specific findings (i.e. the availability of R&D funding within a pharmaceutical company) triangulation was not always possible because of the decision (for feasibility reasons) to interview only a single representative from every actor-group. However, the interview protocol is documented (see appendix A) and the interviews were recorded to increase the accuracy in processing of the data. Replication of the interview findings should be possible. The procedures for analyzing the results are described to the best extent in the methodology section. The next step in analyzing the interviews was the grouping of findings from the interviews along the dimensions provided by the MLP. The overlap of dimensions and difficulties in defining clear operational boundaries (as discussed in construct validity) resulted in a certain amount of subjectivity in assigning statements to a specific dimension. However, the exercise can be repeated and the empirical findings would remain similar, regardless of the dimensions they are assigned to.

8.2 Research limitations and further research

The use of the MLP as an analytical tool to study transitions has both its strengths and weaknesses. While it allows for the analysis of complex transitions of sociotechnical systems, it necessarily has to cover a large amount of broad, all-encompassing concepts, for which the demarcation is not always clear beforehand. For instance, there is no clear conceptual delineation of what does and does not belong to a 'socio-technical regime', these boundaries are up to the researcher to decide and dependent of the subject in question. While the notion 'safety-testing regime' is deemed adequate for this research, other investigators could argue to narrow or broaden this notion, for instance by also including animal testing for different purposes as (toxicological) safety testing.

The distinction in rules between formal, normative and cognitive rules as proposed by Geels (2007), however valuable of a theoretical insight, in practice turned out problematic to use as an analytical tool. For instance, the rules that are proposed as examples in table 3.1 can make data gathering a difficult exercise: how to know when all rules have been identified to completeness? This research relied on expert interviews to ensure the inclusion of *important* rules (on basis of expert opinion), because identifying, for a technology dimension, all: search heuristics, routines, guiding principles, technical problem agenda's, problem solving strategies, technical recipes would be a daunting task. Furthermore, when rules have been identified, the MLP does not seem to adhere any importance or conceptual consequence as to whether a rule is formal or cognitive. In line with criticisms by Genus and Coles (2008) that better analytical operationalization and specification of the regime concept is necessary, additional research could aim to improve upon

the quality of operational measures for rules, by reviewing the types of rules that have been most prevalent in studies that used the MLP.

An additional limitation that was observed in this research, is the relative lack of agency in the MLP. Earlier, Smith et al (2005) found the MLP to be 'too descriptive and structural, leaving room for greater analysis of agency' (p.1492). Their initial emphasis laid on the governance of transitions. While this research dealt with a narrower unit of analysis (i.e. uptake of technology into a regime), more attention to the role of power and politics could create more specific pointers for influencing contemporary transitions. For instance, lobbying practices (of animal welfare groups, but also politically engaged researchers, industry representatives) appear to be of great influence in the shaping of (important) regulations and political decisions on alternative methods. And while the phenomenon of lobbying practices is acknowledged in the MLP, it fails to address the influences that individual actors can have in the process. Particularly in heavily regulated markets that are prevalent in the life sciences, the influence that relatively small, singe actors can exert on important political decision-making processes can provide valuable steering points in contemporary transitions. Geels (2011) dismisses the lack of agency by stating that 'trajectories and multi-level alignments are always enacted by social groups' (p.29), but the fact that agency often comes up as limiting characteristic of the MLP (Berkhout, Smith, and Stirling 2004; Genus and Coles 2008; Pierick and van Mil 2009; Smith, Stirling, and Berkhout 2005) suggests relevancy for additional research that tries to incorporate agency in the MLP. An approach taken by Grin et al. (2010) to draw on political science theory and incorporate the role of power in the MLP could be valuable, in particular, in heavily regulated markets.

The choice for the uptake of the ICE and BCOP test as unit of analysis, as opposed to the transition of the entire safety-testing regime, also made it difficult to pinpoint the relevance for including or excluding some of the sociotechnical factors. Several changes in the sociotechnical regime, such as the notion of localization of regulations versus the globalization of science have certainly influenced the safety-testing regime, but not necessarily through events instigated by the BCOP or ICE test. Despite this, they were important in shaping the regime that exists today and could therefore be noteworthy. This was solved by drawing the line with factors that were mentioned in the interviews in relation to the BCOP or ICE test. A recommendation for further research then is to consider the entire transition towards alternative methods to better establish all the processes that have influenced the transition (and consequently, give up some detail in the analysis of niche-innovations).

Another option for further research would be to compare the findings from this study with similar cases where the replacement of animal-based tests with *in vitro* methods was attempted and either did or did not succeed. This would create a better insight in the factors that actually stimulate replacement and allows for a higher degree of external validity by testing variations in context.

REFERENCES

Balls, M. 1995. "Validation of Alternative Tests in the European Union." *Current Problems in Dermatology* 23:265-274. Retrieved 18 October 2011.

Balls, M., N. Berg, L. H. Bruner, R. D. Curren, O. De Silva, L. K. Earl, D. J. Esdaile, J. H. Fentem, M. Liebsch, Y. Ohno, M. K. Prinsen, H. Spielmann and A. P. Worth. 1999. "Eye Irritation Testing: The Way Forward." *ATLA Alternatives to Laboratory Animals* 27(1):53-77. Retrieved 11 October 2011.

Balls, M., P. A. Botham, L. H. Bruner and H. Spielmann. 1995. "The EC/HO International Validation Study on Alternatives to the Draize Eye Irritation Test." *Toxicology in Vitro* 9(6):871-929. Retrieved 11 October 2011.

Balls, M. and J. H. Fentem. 1997. "Progress Toward the Validation of Alternative Tests." *ATLA Alternatives to Laboratory Animals* 25(1):33-43. Retrieved 18 October 2011.

Baxendale, I. R., J. J. Hayward, S. V. Ley and G. K. Tranmer. 2007. "Pharmaceutical Strategy and Innovation: An Academics Perspective." *ChemMedChem* 2(6):768-788.

Bergek, A., S. Jacobsson, B. Carlsson, S. Lindmark and A. Rickne. 2008. "Analyzing the Functional Dynamics of Technological Innovation Systems: A Scheme of Analysis." *Research Policy* 37(3):407-429.

Berkhout, F., A. Smith and A. Stirling. 2004. "3. Socio-Technological Regimes and Transition Contexts." *System Innovation and the Transition to Sustainability: Theory, Evidence and Policy*:48.

Borenfreund, E. and O. Borrero. 1984. "In Vitro Cytotoxicity Assays. Potential Alternatives to the Draize Ocular Allergy Test." *Cell Biology and Toxicology* 1(1):55-65. Retrieved 19 October 2011.

Bottini, A. A. and T. Hartung. 2009. "Food for Thought... on the Economics of Animal Testing." *Altex* 26(1):3.

Bottini, A. A., P. Amcoff and T. Hartung. 2007. "Food for Thought ... on Globalisation of Alternative Methods." *Altex* 24(4):255-269.

Brantom, P. G., L. H. Bruner, M. Chamberlain, O. De Silva, J. Dupuis, L. K. Earl, D. P. Lovell, W. J. W. Pape, M. Uttley, D. M. Bagley, F. W. Baker, M. Bracher, P. Courtellemont, L. Declercq, S. Freeman, W. Steiling, A. P. Walker, G. J. Carr, N. Dami, G. Thomas, J. Harbell, P. A. Jones, U. Pfannenbecker, J. A. Southee, M. Tcheng, H. Argembeaux, D. Castelli, R. Clothier, D. J. Esdaile, H. Itigaki, K. Jung, Y. Kasai, H. Kojima, U. Kristen, M. Larnicol, R. W. Lewis, K. Marenus, O. Moreno, A. Peterson, E. S. Rasmussen, C. Robles and M. Stern. 1997. "A Summary Report of the COLIPA International Validation Study on Alternatives to the Draize Rabbit Eye Irritation Test." *Toxicology in Vitro* 11(1-2):141-179. Retrieved 11 October 2011.

Bruner, L. H., G. J. Carr, M. Chamberlain and R. D. Curren. 1996. "Validation of Alternative Methods for Toxicity Testing." *Toxicology in Vitro* 10(4):479-501.

Burton, A. B. G. 1971. "A Method for the Objective Assessment of Eye Irritation." *Food and Cosmetics Toxicology* 10(2):209-210,IN7,211-217. Retrieved 19 October 2011.

Burton, A. B. G., M. York and R. S. Lawrence. 1981. "The in Vitro Assessment of Severe Eye Irritants." *Food and Cosmetics Toxicology* 19(4):471-480. Retrieved 11 October 2011.

Bussiere, J. L., P. Martin, M. Horner, J. Couch, M. Flaherty, L. Andrews, J. Beyer and C. Horvath. 2009. "Alternative Strategies for Toxicity Testing of Species-Specific Biopharmaceuticals." *International Journal of Toxicology* 28(3):230.

CAAT. 2011. " Johns Hopkins Center for Alternatives to Animal Testing.", Retrieved Sep 15, 2011. (http://caat.jhsph.edu/).

Carlsson, B. and R. Stankiewicz. 1991. "On the Nature, Function and Composition of Technological Systems." *Journal of Evolutionary Economics* 1(2):93-118.

Castro Díaz-Balart, F. and H. Pérez Rojas. 2002. "Globalization, Science and Development." *Perspectives on Global Development and Technology* 1(3-4):322-339. Retrieved 19 October 2011.

CEC. 1991. "Collaborative Study on the Evaluation of Alternative Methods to the Eye Irritation Test." *Commission of the European Communities* EC Document XI/632/91, V/E/I/131/91.

Chamberlain, M., SC Gad, P. Gautheron and MK Prinsen. 1997a. "Irag Working Group 1: Organotypic Models for the assessment/prediction of Ocular Irritation." *Food and Chemical Toxicology* 35(1):23-37.

Chamberlain, M., S. C. Gad, P. Gautheron and M. K. Prinsen. 1997b. "Organotypic Models for the assessment/prediction of Ocular Irritation." *Food and Chemical Toxicology* 35(1):23-37. Retrieved 11 October 2011.

Clark, G. R. 1958. "John H. Draize, the Man and the Scientist." *J. Soc. Cosmet. Chem.* 9(120-1).

Coenen, L. and F. D. Lopez. 2008. "Comparing Systemic Approaches to Innovation for Sustainability and Competitiveness.".

Coleman, R. A. 2011. "Human Tissue in the Evaluation of Safety and Efficacy of New Medicines: A Viable Alternative to Animal Models?" *ISRN Pharmaceutics* ISNR Pharmaceutics.

Collins, F. 2008. "Robots could Reduce Animal Tests." *BBC News,* .

Dahl, H., B. Gluud, P. Vangsted and M. Norn. 1985. "Eye Lesions Induced by Mustard Gas." *Acta Ophthalmologica* 63(SUPPL. 173):30-31. Retrieved 11 October 2011.

Draize, J. H. 1955. "VIII. Dermal Toxicity." *Food Drug Cosmetic Law Journal* 10:722-32.

Draize, J. H. and E. A. Kelley. 1959. "The Urinary Excretion of Boric Acid Preparations Following Oral Administration and Topical Applications to Intact and Damaged Skin of Rabbits." *Toxicology and Applied Pharmacology* 1(3):267-276. Retrieved 11 October 2011.

Draize, J. H., G. Woodard and H. O. Calvery. 1944. "Methods for the Study of Irritation and Toxicity of Substances Applied Topically to the Skin and Mucous Membranes." *J.Pharmacol.Exp.Ther.* 82:377-390. Retrieved 11 October 2011.

Durkheim, E. 1997. *The Division of Labor in Society.*Free Pr.

EC. 2011. "Extended Impact Assessment on the REACH Proposal.", Retrieved 09/10, 2011. (http://ec.europa.eu/enterprise/sectors/chemicals/documents/reach/archives/impact-assessment/index_en.htm#h2-extended-impact-assessment-(eia)).

------. 2010a. "Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the Protection of Animals used for Scientific Purposes (Text with EEA Relevance)." *Official Journal of the European Union* L267/33.

------. 2010b. "Special Eurobarometer on Science and Technology.", Retrieved 10/10, 2011. (http://ec.europa.eu/public_opinion/archives/ebs/ebs_340_en.pdf).

------. 2007. "Fifth Report from the Commission to the Council and the European Parliament on the Statistics on the Number of Animals used for Experimental and Other Scientific Purposes in the Member States of the European Union." Union COM/2007/675 final.

------. 2003. "Directive 2003/15/EC of the European Parliament and of the Council of 27 February 2003 Amending Council Directive 76/768/EEC on the Approximation of the Laws of the Member States Relating to Cosmetic Products (Text with EEA Relevance)." *Official Journal of the European Union* OJ L 66, 11.3.2003, p. 26–35.

------. 2000. "Commission Directive 2000/41/EC of 19 June 2000 Postponing for a Second Time the Date After which Animal Tests are Prohibited for Ingredients Or Combinations of Ingredients of Cosmetic Products (Text with EEA Relevance)." *Official Journal of the European Union* OJ L 145, 20.6.2000, p. 25–26.

------. 1997a. "1996 Commission Report on the Development, Validation and Legal Acceptance of Alternative Methods to Animal Experiments in the Field of Cosmetics
." *Official Journal of the European Union* COM/97/0182.

------. 1997b. "Commission Directive 97/18/EC of 17 April 1997 Postponing the Date After which Animal Tests are Prohibited for Ingredients Or Combinations of Ingredients of Cosmetic Products (Text with EEA Relevance)." *Official Journal of the European Union* OJ L 114, 1.5.1997, p. 43–44.

------. 1993a. "Council Directive 93/35/EEC of 14 June 1993 Amending for the Sixth Time Directive 76/768/EEC on the Approximation of the Laws of the Member States Relating to Cosmetic Products." *Official Journal of the European Union* OJ L 151, 23.6.1993, p. 32–37.

------. 1993b. "Fifth European Community Environment Programme: Towards Sustainability.", Retrieved 09/09, 2011. (http://europa.eu/legislation_summaries/other/l28062_en.htm).

------. 1986. "Council Directive 86/609/EEC of 24 November 1986 on the Approximation of Laws, Regulations and Administrative Provisions of the Member States regarding the Protection of Animals used for Experimental and Other Scientific purposes
." *Official Journal of the European Union* OJ L 358, 18.12.1986, p. 1–28.

Edquist, C. and B. A. Lundvall. 1993. "Comparing the Danish and Swedish Systems of Innovation." *National Innovation Systems: A Comparative Analysis*:265-298.

Elzen, B. and A. Wieczorek. 2005. "Transitions Towards Sustainability through System Innovation." *Technological Forecasting and Social Change* 72(6):651-662.

Eskes, D. 2010. "Application of Alternative Methods in the Regulatory Assesment of Chemical Safety Related to Human Eye Irritation & Severe Irritation." *Orange House Partnership*.

Eskes, C., S. Bessou, L. Bruner, R. Curren, J. Harbell, P. Jones, R. Kreiling, M. Liebsch, P. McNamee, W. Pape, M. K. Prinsen, T. Seidle, P. Vanparys, A. Worth and V. Zuang. 2005. "Eye Irritation." *ATLA Alternatives to Laboratory Animals* 33(SUPPL. 1):47-81. Retrieved 19 October 2011.

FDA. 2004. "Innovation Or Stagnation: Challenge and Opportunity on the Critical Path to New Medical Products." *US Food and Drug Administration*.

FRAME. 2010. "Http://www.Frame.Org.Uk.", Retrieved 01/08, 2011.

Fraser, D. 1999. "Animal Ethics and Animal Welfare Science: Bridging the Two cultures1." *Applied Animal Behaviour Science* 65(3):171-189.

Frazier, J. M., S. C. Gad, A. M. Goldberg and J. P. McCulley. 1987. *A Critical Evaluation of Alternatives to Acute Ocular Irritation Testing.* Vol. 4. Liebert Press. ed. New York, N.Y.: MA Liebert.

Freeman, C. and C. Perez. 2000. "Structural Crises of Adjustment, Business Cycles and Investment Behaviour." *Technology, Organizations and Innovation: Theories, Concepts and Paradigms*:871.

Freeman, L. and A. Ward. 1990. "
Animal Uproar; when Testing for Human Health and Safety is the Issue, Beauty-Aids Marketers Struggle to Define Humane." *Advertising Age* 61(S1-S2).

Gautheron, P., M. Dukic, D. Alix and J. F. Sina. 1992. "Bovine Corneal Opacity and Permeability Test: An in Vitro Assay of Ocular Irritancy." *Fundamental and Applied Toxicology* 18(3):442-449. Retrieved 11 October 2011.

Gautheron, P., J. Giroux, M. Cottin, L. Audegond, A. Morilla, L. Mayordomo-Blanco, A. Tortajada, G. Haynes, J. A. Vericat, R. Pirovano, E. Gillio Tos, C. Hagemann, P. Vanparys, G. Deknudt, G. Jacobs, M. Prinsen, S. Kalweit and H. Spielmann. 1994. "Interlaboratory Assessment of the Bovine Corneal Opacity and Permeability (BCOP) Assay." *Toxicology in Vitro* 8(3):381-392. Retrieved 11 October 2011.

Geels, F. 2005. "Co-Evolution of Technology and Society: The Transition in Water Supply and Personal Hygiene in the Netherlands (1850-1930) - A Case Study in Multi-Level Perspective." *Technology in Society* 27(3):363-397. Retrieved 18 October 2011.

Geels, F. W. 2011. "The Multi-Level Perspective on Sustainability Transitions: Responses to Seven Criticisms." *Environmental Innovation and Societal Transitions* 1(1):24-40. Retrieved 18 October 2011.

------. 2010. "Ontologies, Socio-Technical Transitions (to Sustainability), and the Multi-Level Perspective." *Research Policy* 39(4):495-510. Retrieved 18 October 2011.

------. 2007. "Transformations of Large Technical Systems: A Multilevel Analysis of the Dutch Highway System (1950-2000)." *Science Technology and Human Values* 32(2):123-149. Retrieved 18 October 2011.

------. 2006a. "Co-Evolutionary and Multi-Level Dynamics in Transitions: The Transformation of Aviation Systems and the Shift from Propeller to Turbojet (1930-1970)." *Technovation* 26(9):999-1016. Retrieved 18 October 2011.

------. 2006b. "The Hygienic Transition from Cesspools to Sewer Systems (1840-1930): The Dynamics of Regime Transformation." *Research Policy* 35(7):1069-1082. Retrieved 18 October 2011.

------. 2006c. "Major System Change through Stepwise Reconfiguration: A Multi-Level Analysis of the Transformation of American Factory Production (1850-1930)." *Technology in Society* 28(4):445-476. Retrieved 18 October 2011.

------. 2005. "Processes and Patterns in Transitions and System Innovations: Refining the Co-Evolutionary Multi-Level Perspective." *Technological Forecasting and Social Change* 72(6 SPEC. ISS.):681-696. Retrieved 18 October 2011.

------. 2004. "From Sectoral Systems of Innovation to Socio-Technical Systems: Insights about Dynamics and Change from Sociology and Institutional Theory." *Research Policy* 33(6-7):897-920. Retrieved 18 October 2011.

------. 2002. "Technological Transitions as Evolutionary Reconfiguration Processes: A Multi-Level Perspective and a Case-Study." *Research Policy* 31(8-9):1257-1274. Retrieved 18 October 2011.

Geels, F. W. and R. Kemp. 2007. "Dynamics in Socio-Technical Systems: Typology of Change Processes and Contrasting Case Studies." *Technology in Society* 29(4):441-455. Retrieved 18 October 2011.

Geels, F. W. and J. Schot. 2007. "Typology of Sociotechnical Transition Pathways." *Research Policy* 36(3):399-417. Retrieved 18 October 2011.

Genus, A. and A. M. Coles. 2008. "Rethinking the Multi-Level Perspective of Technological Transitions." *Research Policy* 37(9):1436-1445.

Gerlach, A. 2001. "Sustainable Entrepreneurship and Innovation." *Centre for Sustainability Management, University of Lueneburg, Luenebourg*.

Gettings, S. D., R. A. Lordo, K. L. Hintze, D. M. Bagley, P. L. Casterton, M. Chudkowski, R. D. Curren, J. L. Demetrulias, L. C. Dipasquale, L. K. Earl, P. I. Feder, C. L. Galli, S. M. Glaza, V. C. Gordon, J. Janus, P. J. Kurtz, K. D. Marenus, J. Moral, W. J. W. Pape, K. J. Renskers, L. A. Rheins, M. T. Roddy, M. G. Rozen, J. P. Tedeschi and J. Zyracki. 1996. "The CFTA Evaluation of Alternatives Program: An Evaluation of in Vitro Alternatives to the Draize Primary Eye Irritation Test. (Phase III) Surfactant-Based Formulations." *Food and Chemical Toxicology* 34(1):79-117. Retrieved 19 October 2011.

Grabowski, H., J. Vernon and J. A. DiMasi. 2002. "Returns on Research and Development for 1990s New Drug Introductions." *PharmacoEconomics* 20(S3):11-29.

Grin, J., J. Rotmans, J. Schot, F. W. Geels and D. Loorbach. 2010. "Transitions to Sustainable Development: New Directions in the Study of Long Term Transformative Change." *Routledge, New York*.

Hansson, S. O. and C. Rudén. 2003. "Improving the Incentives for Toxicity Testing." *Journal of Risk Research* 6(1):3-21. Retrieved 11 October 2011.

Hartung, T. 2010. "Food for Thought… on Alternative Methods for Chemical Safety Testing." *Altex* 27(1):3.

Hartung, T. and G. Daston. 2009. "Are in Vitro Tests Suitable for Regulatory use?" *Toxicological Sciences* 111(2):233.

Hartung, T. 2010a. "Evidence-Based Toxicology - the Toolbox of Validation for the 21 St Century?" *Altex* 27(4):253-263. Retrieved 18 October 2011.

------. 2010b. "Lessons Learned from Alternative Methods and their Validation for a New Toxicology in the 21st Century." *Journal of Toxicology and Environmental Health - Part B: Critical Reviews* 13(2-4):277-290. Retrieved 11 October 2011.

------. 2008a. "Food for Thought... on Alternative Methods for Cosmetics Safety Testing." *Altex* 25(3):147-162. Retrieved 11 October 2011.

------. 2008b. "Food for Thought... on Animal Tests." *Altex* 25(1):3-16. Retrieved 11 October 2011.

Harvey, L. K. and S. Shubat. 1987. "AMA Surveys of Physician and Public Opinion on Health Care Issues." *American Medical Association*.

Hekkert, MP, R. A. A. Suurs, SO Negro, S. Kuhlmann and R. Smits. 2007. "Functions of Innovation Systems: A New Approach for Analysing Technological Change." *Technological Forecasting and Social Change* 74(4):413-432.

Hill, J. W. and A. W. Langvardt. 2010. "
Healthcare and the Life Sciences in Transition." *Business Horizons, Elsevier* 52(2):87-91.

Hoogma, R., R. Kemp and J. Schot. 2002. "Experimenting for Sustainable Transport. the Approach of Strategic Niche Management." *Spon Press, London/New York*.

Hopp, V. 2002. "Life Sciences and Globalization - Two Modern Marketing Slogans Or a Challenge for the Future?" *Chemical Engineering and Technology* 25(2):45-52. Retrieved 19 October 2011.

Hottendorf, G. H. 1987. "Species Differences in Toxic Lesions." *Human Risk Assessment-the Role of Animal Selection and Extrapolation (M.V.Roloff, Ed.), Taylor and Francis, Philadelphia*:87-95.

ICCVAM. 2010. "In Vitro Test Methods for Detecting Ocular Corrosives and Severe Irritants - OECD Test Guideline Proposals.", Retrieved 10/05, 2011. (http://iccvam.niehs.nih.gov.proxy.library.uu.nl/methods/ocutox/ivocutox/OECD.htm).

------. 2006a. " Background Review Document: Current Status of in Vitro Test Methods for Identifying Ocular Corrosives and Severe Irritants: Bovine Corneal Opacity and Permeability Test Method." *NIH Publication no: 06-4512.*

------. 2006b. "Test Method Evaluation Report: In Vitro Ocular Toxicity Methods for Identifying Severe Irritants and Corrosives." *NIH Publication no: 07-4517*.

------. 1997. "Validation and Regulatory Acceptance of Toxicological Test Methods." *National Institute of Environmental Health Sciences, NIH, DHHS*.

Jacobsson, S. 2002. "Universities and Industrial Transformation: An Interpretative and Selective Literature Study with Special Emphasis on Sweden." *Science and Public Policy* 29(5):345-365.

Jennings, L. 2006. "The Effects of Globalization on Freight Transportation." *2006 IIE Annual Conference and Exhibition*. Retrieved 19 October 2011.

Jucker, M. 2010. "The Benefits and Limitations of Animal Models for Translational Research in Neurodegenerative Diseases." *Nature Medicine* 16(11):1210-1214.

Kemp, R. 1994. "Technology and the Transition to Environmental Sustainability* 1:: The Problem of Technological Regime Shifts." *Futures* 26(10):1023-1046.

Knapp, A. 1949. "Ophthalmology in the War Years. Volume 2 (1944--June 1946)." *Archives of Ophthalmology* 41(1):110.

Koeter, H. and M. K. Prinsen. 1985. "Comparison of in Vivo and in Vitro Eye Irritancy Test Systems: A Study with 34 Substances." *Alternative Methods in Toxicology (USA)*.

Langley, A. 2007. "Process Thinking in Strategic Organization." *Strategic Organization* 5(3):271.

Lilienblum, W., W. Dekant, H. Foth, T. Gebel, JG Hengstler, R. Kahl, P. J. Kramer, H. Schweinfurth and K. M. Wollin. 2008. "Alternative Methods to Safety Studies in Experimental Animals: Role in the Risk Assessment of Chemicals Under the New European Chemicals Legislation (REACH)." *Archives of Toxicology* 82(4):211-236.

Lundvall, B. Å. 1992. "National Innovation Systems: Towards a Theory of Innovation and Interactive Learning." *London: Pinter*.

Malerba, F. 2002. "Sectoral Systems of Innovation and Production* 1." *Research Policy* 31(2):247-264.

Markard, J. and B. Truffer. 2008. "Technological Innovation Systems and the Multi-Level Perspective: Towards an Integrated Framework." *Research Policy* 37(4):596-615.

Matthews, R. A. 2008. "Medical Progress Depends on Animal models–doesn't it." *JR Soc Med* 101(2):95-98.

Miller, J. L. 1935. "The Influence of Claude Bernard's Experimental Methods on Medicine." *The Journal of Laboratory and Clinical Medicine* 20(10):1001-1008. Retrieved 11 October 2011.

MORI. 2010. "Views on Animal Experimentation.", Retrieved 10/10, 2011. (http://www.ipsos-mori.com/researchpublications/researcharchive/poll.aspx?oItemId=2563).

Muir, C. K. 1985. "Opacity of Bovine Cornea in Vitro Induced by Surfactants and Industral Chemicals Compared with Ocular Irritancy in Vivo." *Toxicology Letters* 24(2-3):157-162. Retrieved 19 October 2011.

------. 1984. "A Simple Method to Assess Surfactant-Induced Bovine Corneal Opacity in Vitro: Preliminary Findings." *Toxicology Letters* 22(2):199-203. Retrieved 11 October 2011.

Nelson, R. R. and S. G. Winter. 1982. *An Evolutionary Theory of Economic Change.*Belknap Press.

Nichols, L. M., P. B. Ginsburg, R. A. Berenson, J. Christianson and R. E. Hurley. 2004. "Are Market Forces Strong enough to Deliver Efficient Health Care Systems? Confidence is Waning." *Health Affairs* 23(2):8.

NIH. 1993. "The National Institutes of Health (NIH) Revitalization Act of 1993." *Public Law no. 103-43* 42 U.S.C.A. §289g-1, -2.

OECD. 2009a. "Test no. 437: Bovine Corneal Opacity and Permeability Test Method for Identifying Ocular Corrosives and Severe Irritants." *OECD Guidelines for the Testing of Chemicals, Section 4: Health Effects*. Retrieved 19 October 2011.

------. 2009b. "Test no. 438: Isolated Chicken Eye Test Method for Identifying Ocular Corrosives and Severe Irritants." *OECD Guidelines for the Testing of Chemicals, Section 4: Health Effects*. Retrieved 19 October 2011.

------. 2002. "Test no. 405: Acute Eye Irritation/Corrosion." *OECD Guidelines for the Testing of Chemicals / Section 4: Health Effects*.

------. 1981. "Council Decision Concerning the Mutual Acceptance of Data (MAD) in the Assessment of Chemicals.
." [C(81)30].

Ogden, T. 2010. "REACH-how is it Going?" *Annals of Occupational Hygiene* 54(1):1-4. Retrieved 18 October 2011.

Ohno, Y., T. Kaneko, T. Kobayashi, T. Inoue, Y. Kuroiwa, T. Yoshida, J. Momma, M. Hayashi, J. Akiyama and T. Atsumi. 1994. "First-Phase Validation of the in Vitro Eye Irritation Tests for Cosmetic Ingredients." *In Vitro Toxicology* 7(2):89-94.

Olson, H., G. Betton, D. Robinson, K. Thomas, A. Monro, G. Kolaja, P. Lilly, J. Sanders, G. Sipes and W. Bracken. 2000. "Concordance of the Toxicity of Pharmaceuticals in Humans and in Animals." *Regulatory Toxicology and Pharmacology* 32(1):56-67.

Pape, W. J. W., U. Pfannenbecker, H. Argembeaux, M. Bracher, D. J. Esdaile, S. Hagino, Y. Kasai and R. W. Lewis. 1999. "COLIPA Validation Project on in Vitro Eye Irritation Tests for Cosmetic Ingredients and Finished Products (Phase I): The Red Blood Cell Test for the Estimation of Acute Eye Irritation Potentials. Present Status." *Toxicology in Vitro* 13(2):343-354. Retrieved 11 October 2011.

Parsons, T. 1949. *The Structure of Social Action.*Free Press New York.

Piehl, M., M. Carathers, R. Soda, D. Cerven and G. DeGeorge. 2011. "Porcine Corneal Ocular Reversibility Assay (PorCORA) Predicts Ocular Damage and Recovery for Global Regulatory Agency Hazard Categories." *Toxicology in Vitro*. Retrieved 19 October 2011.

Pierick, E. and E. van Mil. 2009. *Multi-Level Perspective Nader Beschouwd: Aangrijpingspunten Voor Transitie Richting Biobased Economy?*LEI Wageningen UR.

Pifer, L., K. and E. Shimizu. 1994. "Public Attitudes Toward Animal Research: Some International Comparisons." *Soc. Anim.* 2: 95-113.

Popper, K. R. 2002. *Conjectures and Refutations: The Growth of Scientific Knowledge.*Psychology Press.

Potter, D. 1989. "Ida Mann--Her Wartime Career, 1939-1949." *Australian and New Zealand Journal of Ophthalmology* 17(1):95-101. Retrieved 11 October 2011.

Price, J. B. and I. J. Andrews. 1985. "The in Vitro Assessment of Eye Irritancy using Isolated Eyes." *Food and Chemical Toxicology* 23(2):313-315. Retrieved 19 October 2011.

Prinsen, M. K. and H. B. W. M. Koeter. 1993. "Justification of the Enucleated Eye Test with Eyes of Slaughterhouse Animals as an Alternative to the Draize Eye Irritation Test with Rabbits." *Food and Chemical Toxicology* 31(1):69-76. Retrieved 11 October 2011.

Rip, A. and R. Kemp. 1998. "Technological Change." *Human Choice and Climate Change* 2:327-399.

Rotmans, J. 2007. "Duurzaamheid: Van Onderstroom Naar Draaggolf." *Jansen En Van Driel*.

Rougier, A., M. Cottin, O. De Silva, R. Roguet, P. Catroux, A. Toufic and K. -. Dossou. 1992. "In Vitro Methods: Their Relevance and Complementarity in Ocular Safety Assessment." *Lens and Eye Toxicity Research* 9(3-4):229-245. Retrieved 19 October 2011.

Rowan A.N., Loew F.M. and J.C. Weer. 1995.
*the Animal Research Controversy : Protest, Process & Public Policy : An Analysis of Strategic Issues.*Center for Animals & Public Policy, Tufts University School of Veterinary Medicine.

Russell, W. M. S., R. L. Burch and C. W. Hume. 1959. *The Principles of Humane Experimental Technique.*Methuen London, UK.

Schaeffer, C. 2005. "Searching for Innovation: Drug Research." *The American Anti-Vivisection Society* CXIII(3).

Schein, P. S., R. D. Davis, S. Carter, J. Newman, D. R. Schein and D. P. Rall. 1970. "The Evaluation of Anticancer Drugs in Dogs and Monkeys for the Prediction of Qualitative Toxicities in Man." *Clinical Pharmacology and Therapeutics* 11(1):3-40.

Schot, J., R. Hoogma and B. Elzen. 1994. "Strategies for Shifting Technological Systems* 1:: The Case of the Automobile System." *Futures* 26(10):1060-1076.

Schot, J. and F. W. Geels. 2008. "Strategic Niche Management and Sustainable Innovation Journeys: Theory, Findings, Research Agenda, and Policy." *Technology Analysis and Strategic Management* 20(5):537-554. Retrieved 18 October 2011.

Scott, L., C. Eskes, S. Hoffmann, E. Adriaens, N. Alepée, M. Bufo, R. Clothier, D. Facchini, C. Faller, R. Guest, J. Harbell, T. Hartung, H. Kamp, B. L. Varlet, M. Meloni, P. McNamee, R. Osborne, W. Pape, U. Pfannenbecker, M. Prinsen, C. Seaman, H. Spielmann, W. Stokes, K. Trouba, C. V. d. Berghe, F. V. Goethem, M. Vassallo, P. Vinardell and V. Zuang. 2010. "A Proposed Eye Irritation Testing Strategy to

Reduce and Replace in Vivo Studies using Bottom-Up and Top-Down Approaches." *Toxicology in Vitro* 24(1):1-9. Retrieved 19 October 2011.

Sina, J. F. and P. Gautheron. 1998. "An Historical Perspective." *In Vitro and Molecular Toxicology: Journal of Basic and Applied Research* 11(4):316-326. Retrieved 19 October 2011.

Smith, A., A. Stirling and F. Berkhout. 2005. "The Governance of Sustainable Socio-Technical Transitions." *Research Policy* 34(10):1491-1510.

Spielmann, H., M. Liebsch and S. Kalweit. 1996. "Results of a Validation Study in Germany on Two in Vitro Alternatives to the Draize Eye Irritation Test, the HET-CAM Test and the 3T3 NRU Cytotoxicity Test." *ATLA* 24:741-858. Retrieved 11 October 2011.

Ubels, J. L. and D. P. Clousing. 2005. "In Vitro Alternatives to the use of Animals in Ocular Toxicology Testing." *Ocular Surface* 3(3):126-142. Retrieved 11 October 2011.

Van Goethem, F., E. Hansen, M. Sysmans, A. De Smedt, P. Vanparys and J. Van Gompel. 2010. "Development of a New Opacitometer for the Bovine Corneal Opacity and Permeability (BCOP) Assay." *Toxicology in Vitro* 24(6):1854-1861. Retrieved 19 October 2011.

Vanparys, P., G. Deknudt, M. Sysmans, G. Teuns, W. Coussement and H. Van Cauteren. 1993. "Evaluation of the Bovine Corneal Opacity-Permeability Assay as an in Vitro Alternative to the Draize Eye Irritation Test." *Toxicology in Vitro* 7(4):471-476. Retrieved 11 October 2011.

von Roten, F. C. 2009. "European Attitudes Towards Animal Research." *Science Technology & Society* 14(2):349.

VWS. 2011. "Actieplan Dierproeven En Alternatieven. Ministerie Van Volksgezondheid, Welzijn En Sport.". (http://www.rijksoverheid.nl/bestanden/documenten-en-publicaties/richtlijnen/2011/09/19/actieplan-dierproeven-en-alternatieven-2011-2021/actieplan-alternatieven-voor-dierproeven.pdf).

Wanningen, V. 2009. "Policy in Socio-Technical Transition: The Case of Consumer Privacy Protection for Radio Frequency Identification (RFID) in Retail." *Master Thesis Philosophy of Science Technology Society*.

Wilhelmus, K. R. 2001. "The Draize Eye Test." *Survey of Ophthalmology* 45(6):493-515. Retrieved 11 October 2011.

Williams, E. S., J. Panko and D. J. Paustenbach. 2009. "The European Union's REACH Regulation: A Review of its History and Requirements the EU REACH Regulation: A Review E. S. Williams Et Al." *Critical Reviews in Toxicology* 39(7):553-575. Retrieved 18 October 2011.

Wind, M., D. Blakey, H. Kojima, J. Kreysa and W. Stokes. 2010. "International Cooperation on Alternative Test Methods (Status Report on ICATM)." *ALTEX* 27(Special Issue).

Worth, A. P. and M. Balls. 2002. *Alternative (Non-Animal) Methods for Chemicals Testing: Current Status and Future Prospects.*Fund for the Replacement of Animals in Medical Experiments.

Yin, R. K. 1993. *Case Study Research: Design and Methods.*Sage Publications, Inc.

ZonMw. 2011. "Dierproeven Begrensd - Http://www.Zonmw.nl/nl/onderwerpen/alle-Programma-s/dierproeven-Begrensd/.", Retrieved 01/08, 2011.

Zuang, V. 2002. "ECVAM's Research and Validation Activities in the Fields of Topical Toxicity and Human Studies." *ATLA* 30:119-124.

Zuang, V., C. Eskes, C. Griesinger and T. Hartung. 2007. "ECVAM Key Area Topical Toxicity: Update on Activities." *AATEX* 14(Special Issue):523-528.

Zuang, V., M. Balls, P. A. Botham, A. Coquette, E. Corsini, R. D. Curren, G. R. Elliott, J. H. Fentem, J. R. Heylings, M. Liebsch, J. Medina, R. Roguet, J. J. van de Sandt, C. Wiemann and A. P. Worth. 2002. "Follow-Up to the ECVAM Prevalidation Study on in Vitro Tests for Acute Skin Irritation. the European Centre for the Validation of Alternative Methods Skin Irritation Task Force Report 2." *Alternatives to Laboratory Animals : ATLA* 30(1):109-129.

Zurlo, J., D. Rudacille and A. M. Goldberg. 1994. *Animals and Alternatives in Testing: History, Science, and Ethics.*Mary Ann Liebert.

APPENDICES

A: INTERVIEW PROTOCOL

Category	Concept	#	Questions
General			Can you first elaborate a bit on your background regarding alternative methods for animal testing, in particular as far as the replacement of the Draize test is concerned?
Niche innovations	Origin of technology		What was the first domain of application where the alternative technology arose?
	(External) influences on niche creation		What originated the development of so many alternative testing strategies?
	Novelties on the basis of expectations and visions		At that time, how did the involved researchers/actors interpret the possibilities of the technology?
			What visions and expectations of the alternative method(s) existed in the early days of development, and how did these change over time?
			Who or what influenced the beliefs of potential for the alternative method(s)?
	Learning processes on multiple dimensions		Were there any interactions between different methods (types of technologies), i.e. were all novel testing strategies developed stand-alone, or have improvements or changes (in methods or protocols) occurred due to superior practices in other methods ?
Sociotechnical regime - Technology	Formal rules		Were there any clear functional requirements for alternative methods to be considered for validation studies or safety testing applications? If so, who articulated these requirements?
	Resources and R&D subsidies		What were the main sources for R&D funds for <specific test> ?
			Have any subsidies been used, or were they available at the time?
			Who is responsible for the financing of validation studies and where do the necessary funds come from?
			Are you satisfied with the availability of financial resources for both the development and validation of alternatives for the Draize test?
			Are there financial stimuli to start developing alternative methods? Is there a way to make money off an alternative method once it is validated?
Sociotechnical regime – Science	Formal research programs		Besides the five major multilaboratory studies that have been undertaken, are you aware of any other formal research programs that were beneficial to the replacement of the draize test? If so, what was their role/impact?
	Criteria and methods of knowledge production		Have there been (m)any changes in the way research and validation programs have been designed and undertaken over the years, and if so could you identify changes that made it possible/easier for the altenative methods to be validated?
			Could you identify significant changes in other scientific methods and findings – for instance statistical approaches – that made it possible/easier for alternative methods to be validated?

Sociotechnical regime – Policy	Verification		Please consult the accompanying diagram on policy measures that have impacted the development and validation of the ICE and BCOP test (but also other alternative methods). To your best knowledge, can you judge the correctness and completeness of this diagram?
			Could you judge the legislative measures on their relative importance?
	Regulation impact		Do you have an opinion about the effectiveness of legislation in speeding up the development and validation process of the alternative methods?
	Regulation interaction		Can you identify any external circumstances that clearly influenced certain legislative measures and developments?
			Could you describe the interaction between regulatory authorities and private companies or research organizations in the validation stage?
Sociotechnical regime – Socio-cultural	Sociocultural developments		Would you consider the public opinion to have influenced the development of alternative methods and/or legislation? Can you try to describe how?
			Has this influence been on the Draize test specifically or do think other animal testing practices have been influenced equally?
Sociotechnical regime - Users, markets and distribution networks	User practices and preferences		Who are the main users of the draize test and the altenative methods?
	Construction of markets through users		Have users been explicitly involved the development and validation of the alternative methods?
			How would you describe 'the market' for alternative methods for eye irritation testing?
	Responsibilities		Who started validation procedures for the alternative methods to the draize test? Are there any specific responsibilities for starting such procedures?
Sociotechnical landscape			Are there any general developments or trends in society that you feel are worth mentioning, that have influenced the development and validation of the ICE and BCOP/ alternative tests?
Open questions relating to overview			Can you identify barriers in the validation process?
			What do you think, in the end, made the difference that resulted in the acceptation of the BCOP and ICE test method?
			Is there any other information that you feel like mentioning in relation to the previous questions or topics?
			Relevant contacts?

B: INTERVIEW RESULTS

Table 1: Interview results as *rules*

	FORMAL/REGULATIVE, NORMATIVE & COGNITIVE
TECHNOLOGY	Draize test 'technology' fundamentally flawed, resulting in variable data[1, 2, 3, 8, 11, 12]. Important functional requirements for alternatives were standardized measurement (1980s), prediction models (1990s) and optimized standard protocols for performing the test (1990s) [1,9,12]. The Background Review Documents, published in 2006 from the ICCVAM/ECVAM retrospective study articulated additional functional requirements for the BCOP test to allow the identification of mild irritants: including histology as an endpoint and improving opacitometer sensitivity[10]. R&D funds for ICE test through 'dierenbescherming' [4, 7] in the Netherlands, because funding from within TNO was not available because of a lack of 'belief' in the technology from TNO management. Funding for BCOP test in-house from Merck/MSD, from allocated budget to advance the drug discovery process[8] (not specific to develop alternatives). General good availability of research funds in 1980s because of public concern for animal welfare[7]. Cell culture technology was flourishing through scientific advances in 1980s[3]. Generally, commercial potential is only a small stimulator for the development of alternatives[8]. There are ways to make money off tests (sell tests, sell apparatus used in the tests, perform tests for others) and this does happen, but is not a primary goal in most cases[5, 12]. It can be considered a small market, and not all companies succeed[8]. New compounds from R&D in the pharmaceutical industry are produced in small quantities. A significant benefit from *in vitro* tests is that only a minimal amount of the compound is necessary when compared to an animal test[5]. Some companies publicly distance themselves from testing on animals and embrace technology from alternative methods[12]. Pharmaceutical companies have the responsibility to identify occupational hazard for employees and take adequate safety measures. Eye irritation testing in this case is not obligated, but companies do perform such tests by normative, internal safety standards[5,10]. Sometimes this is accompanied by additional technological development, for instance Johnson & Johnson incorporating histology testing as an additional endpoint, and improving the sensitivity of the opacitometer[5,10]. For J&J, the driver for improved testing is purely scientific. If new methods obtain more accurate results, it is worthwhile to invest in their development[10]. Organotypic tests, BCOP and ICE test in particular (and potentially other tests) are prone to regulations on the transport of animal organs[2]. For instance, the UK does not know free transport of chicken eyes[3]. The BCOP was 'basically useless' for a couple of years during the crisis of the Mad Cow Disease, which prohibited any transportation of bovine organs[4]. It is a lot easier to standardize measurements on cell based / human tissue based tests, therefore they have potential for replacement of the Draize (and ICE/BCOP) in the future[4,7,11]. The large amount of alternatives, besides the BCOP and ICE test, is partially a result from every country having it's own method and being familiar with that[3,4, 8]. Insistence on a capability for *in vitro* tests to predict reversibility is seen as a barrier, and difficult to assess without the use of a living organism[6].
SCIENCE	Six multi-lab validation studies on Draize alternatives including BCOP and ICE in 1990s (table 4.1).

	The specialization of 'biostatistician' was underdeveloped during the first major validation studies and there was little experience with this matter. There wasn't always agreement between the statistician and other researchers from the study, on the 'best' way to analyze results[4] (in particular for EC/HO study, which partly failed through underdeveloped statistics[8]). Eventually, a (very informal) meeting to overcome this problem was held with a small group of influential people and led to the development of the concept of pre-validation[4]. Change in approach to validation: from one total replacement to the Draize test to several tests for different irritancy ranges in tiered-testing strategy (2000-2003)[2, 12]. Retrospective validation study for tiered-testing strategy in 2004[2,11,2]. Prospective validation study on Human Tissue Based methods currently ongoing[2, 12]. Organizational responsibilities for validation studies have evolved from having a regional-specific character to increased cooperation between regions through coordinating bodies[12]. The multi-lab studies were funded mainly through 'general' (i.e. not specifically allocated for validation) research budgets of involved public authorities, and grants from manufacturers organizations. For the EC/HO study, the European Commission decided that a replacement should be available for the Draize test and provided funding[4]. Funding for contemporary studies is usually the responsibility of coordinating bodies (i.e. ECVAM, ICCVAM) but increased expenses and increased amount of studies require sharing of the funding between different parties[2, 3, 11] (i.e. ECVAM and COLIPA for the Human Tissue Based models). In Germany it is possible to obtain government funding for validation studies, because of the realization that ECVAM cannot provide funding for all the validation studies[4]. ICCVAM usually does not fund validation studies, contrary to ECVAM[4, 11]. Two different approaches to validation in the early years[8,9]: (1) validation as a process for the development of new tests whereby test are improved and 'validated' in an iterative cycle and (2) validation as a method to ensure that a test actually measures the endpoints that are defined in advance for a specific purpose. Different statistical methods would lead to different interpretations of the results in validation studies. Becomes more important to specify the satisfactory results from statistical methods up front[4,8]. Functional requirements for validation studies set by OECD in 2003 in an elaborate document. While technically very comprehensive, it also represents an economical factor that companies on themselves can/will not spend the necessary amounts of money to comply with all the requirements[7]. AAVS funded research into HET-CAM[6]
INDUSTRY	Formal testing practices can still require the Draize test for mild to moderate irritants[3, 8, 9, 12]. Confidential nature of information on the types of substances[8] used in products and their toxicology[1, 9] hampers the sharing of such data. Cosmetics industry has 'learned to live' with the ban on animal tests and have gained experience with using *in vitro* methods to assess product safety[3]. As long as animal tests are still required in some operational area's (geographically) of multinational companies, efforts to reduce animal tests elsewhere are hampered[3, 12]. Cosmetics industry (much more than the pharmaceutical and chemical industry) under normative pressure from society embraced the image of producing products that are not tested on animals[4, 9]. Routines for internal safety-test screening within companies are quickly adopted to use alternative methods, often because of cost advantages[5,10]. For instance, P&G in the US has been using BCOP for a long time and also submitting data to regulatory agencies[6]. A hurdle for industry is to sufficiently develop prevalidation models for *in vitro* methods that are used/developed in house, before they can undergo formal validation procedures[4].

	Strategies of companies to patent methods are often not to protect their intellectual property for licensing purposes, but to ensure that the knowledge can be used by other companies as well, however it may/can not be patented again by others; a so called 'use-patent'[8, 11].
POLICY	Regulation for worker safety in the US influenced BCOP development. In particular, OSHA requirements[6]. Regulators in US throughout 1990s regarded result from alternative methods more as 'research' and not comparable to a Draize test[8]. Education of regulators (comment related to the US) is viewed as a hurdle that, if overcome, would benefit to the acceptance of alternative methods. For ICE/BCOP this is still relatively simple (i.e. eye for an eye) but for more complex systems it would be good to remove the 'black box' idea of an alternative method that some substances are tested and data comes out, without really understanding what happens in the process[11]. The view that the Draize test is 'the thing that works' is still prevalent under some regulators in the US[6]. More political pressure in Europe, mainly through Cosmetics Directive[1, 4, 5, 6, 8, 9, 12]. Also more willingness to accept data for severe irritants from ICE/BCOP test[2,8]. In the US, even if alternatives suggested high irritancy, data would not always be accepted because no 100% certainty compared to Draize test[8]. This started to change gradually from de mid 1990s and onwards for the classification of positives[9]. Regulators in Europe accept 'positives' (i.e. irritating substances) much quicker than 'negatives' (i.e. non irritating substances) because the risks for the public are higher in the latter case[12]. Regulators then tend to adhere to politics instead of neutral scientific discussion, and for instance demand 0% false positive rates. The EC around 2002 informally stated that, under public pressure, that data from the ICE and BCOP tests for severe irritants would be accepted[4]. However, this led to the situation where companies would submit *in vitro* data to the EC and Draize data to the US and/or Japan[4]. As long as there is no global agreement, animal tests will still be used. Overview of relevant laws and directives in figure 6.2, confirmed by interviews[3, 7, 12]. OECD Test Guidelines only represent a standard 'toolbox' of tests that are considered best practice. They do not directly change the individual regulations of member countries. Many government agencies require 'the safety of a (pharmaceutical) product' in very broad terms, without specifying in the public domain which tests exactly are required to demonstrate this safety. Formally this is left up to the company. However, the norm is that in practice, agencies expect to see a certain combinations of tests. Companies do not quickly deviate from these, even if they feel data from better (in-house) tests would be superior, because it causes delays with authorities to explain/defend the results[8]. Regulators in the US tend to favor tests that are widely used/available in the US, both for the safety assessment of products, but also for inclusion in validation studies[4]. Still, US regulator support is needed to obtain OECD acceptance[4], making it more difficult for methods developed in Europe to obtain OECD validation without proper collaboration with other regions such as the US[4]. Methods that are patented to restrict their use and availability generally are not allowed in validation studies[2,3,4,12] and frowned upon by regulators[11]. As a solution, the BCOP was patented with a so-called 'use patent', which keeps the intellectual property rights with Merck while allowing other companies to make use of the technology[8]. This has the advantage over not filing for a patent at all, that it prevents other companies from patenting similar technology later on and using that patent to enforce licensing costs. The BCOP now is a classic example of a test that every laboratory with the appropriate hardware can perform[11]. No specific legislation in the US has influenced the advancement of validation studies in the case of eye irritation alternatives[8, 11]. Regulatory labeling systems in the United States have evolved to different standards for different applications. For the validation of *in vitro* methods, harmonization of labeling systems is important[3, 11]. Current efforts try to obtain harmonization (both within the US and globally)[6, 12]. The disadvantage of the use of large panels (at ICCVAM) for review procedures, is that group dynamics can and do sometimes hamper the decision making process, in particular because of very 'conservative' or 'overly scientific' input[6]. This, in part, is a result of open meetings at

	ICCVAM, which are in principle available for anyone to attend[6].
SOCIO-CULTURAL	Animal welfare groups strongly lobby in the EU parliament to recognize animal welfare and regulate animal testing[3, 4, 6, 7]. In the US there also has been lobbying and active involvement from animal rights groups in the form of advocacy towards legislators to stimulate the use and development of alternatives, although with less influence on the eventual regulatory process[6]. Europeans generally place greater value on animal welfare in the case of cosmetics testing; Americans are more inclined to favor product safety of cosmetics even if it means more animal tests[8]. Symbolic meanings of eyes[1]. Emotional response to impaired vision[1] is significant[3]. The amount of animals used in eye irritation studies is relatively low, but unknown to the general public[3]. Before the successful anti-Draize campaign by Henry Spira, the perspective of animal rights organizations including AAVS was that they had to try and change the regulatory requirements for the Draize test. The more pragmatic and 'innovative' approach by Spira (and afterwards) was to incentivize companies to actively pursue alternative methods, and in particular make use of the fact that companies tend to adhere to what their consumers want[6].
USERS, MARKETS AND DISTRIBUTIONS NETWORKS	Main user of eye irritation tests is the cosmetics industry[1, 7, 12, 10]. Chemical industry uses eye irritation tests[3, 2, 7, 10, 12]; is to a lesser extend bounded to Cosmetics Directive[1,3, 12] but will have to test chemicals that are on the market under REACH[1,3, 7, 12]. The cosmetics industry tests mainly for mild irritants, as there should not be severe irritants in cosmetics to begin with; this is problematic as long as there are no validated alternatives (for eye irritation) for lower irritancy ranges, the ICE and BCOP are only very partially effective in reducing animal tests[4]. The market for eye irritation has not changed considerably over the years; however in the near future REACH could increase the share of the chemical industry through mandatory testing[7]. The users for eye irritation tests are very much the same as the producers (i.e. industry)[4].
NICHE DEVELOPMENTS	Reason for using chicken eyes with the ICE test, as alternative to Draize: 'an eye for an eye', provides an obvious correlation with the human eye[3]. High expectations of the quick replacement of 'acute tests', because of their apparent simplicity and, early on, generally good indications from in-house studies that similar results to the Draize tests were obtained[3, 4, 7, 8,]. Similar high expectation from the EC/HO study to validate the ICE/BCOP test upon completion[3,7,8]. During development of the ICE tests, there was resistance with the general management of TNO to the efforts to replace animal tests, because it was not clear if it would be worthwhile[3, 7]. BCOP test developed by Merck to test for occupational hazard screening, and only afterwards omitted into validation studies with the aim to be used as toxicology tests for product safety[8].
LANDSCAPE DEVELOPMENTS	When we move to a global economy, which has been happening over last decades, measures have to be put into place for the harmonization of classification systems[11].

C: List of interviewees

#	Interviewee	Function	Date
1-a	Symposium Nederlandse Vereniging Toxicologie	'Alternatieve benaderingen in de risicobeoordeling	12-04-2011
2-b	<name>	ESTIV, ECVAM	26-05-2011
3-c	<name>	TNO	25-05-2011
4-d	<name>	Freie Universität Berlin Former: ZEBET (DE Centre for the Documentation and Evaluation of Alternatives to Testing in Animals)	31-05-2011
5-e	<name>	Johnson & Johnson	30-05-2011
6-f	<name>	AAVS	01-06-2011
7-g	<name>	Orange House Partnership, EFSA, OECD, TNO	19-05-2011
8-h	<name>	Merck/MSD	31-05-2011
9-i	<name>	Grocery Manufacturers Association (GMA), Procter & Gamble, Gilette	27-05-2011
10-j	<name>	Johnson & Johnson	30-05-2011
11-k	<name>	IIVS, OECD TG panels and ICCVAM review panels	02-06-2011
12-l	<name>	ECVAM (EC JRC)	25-05-2011

* Names have been removed for privacy purposes. Contact the author for more information on the interviewees.

D: Structure of a validation study

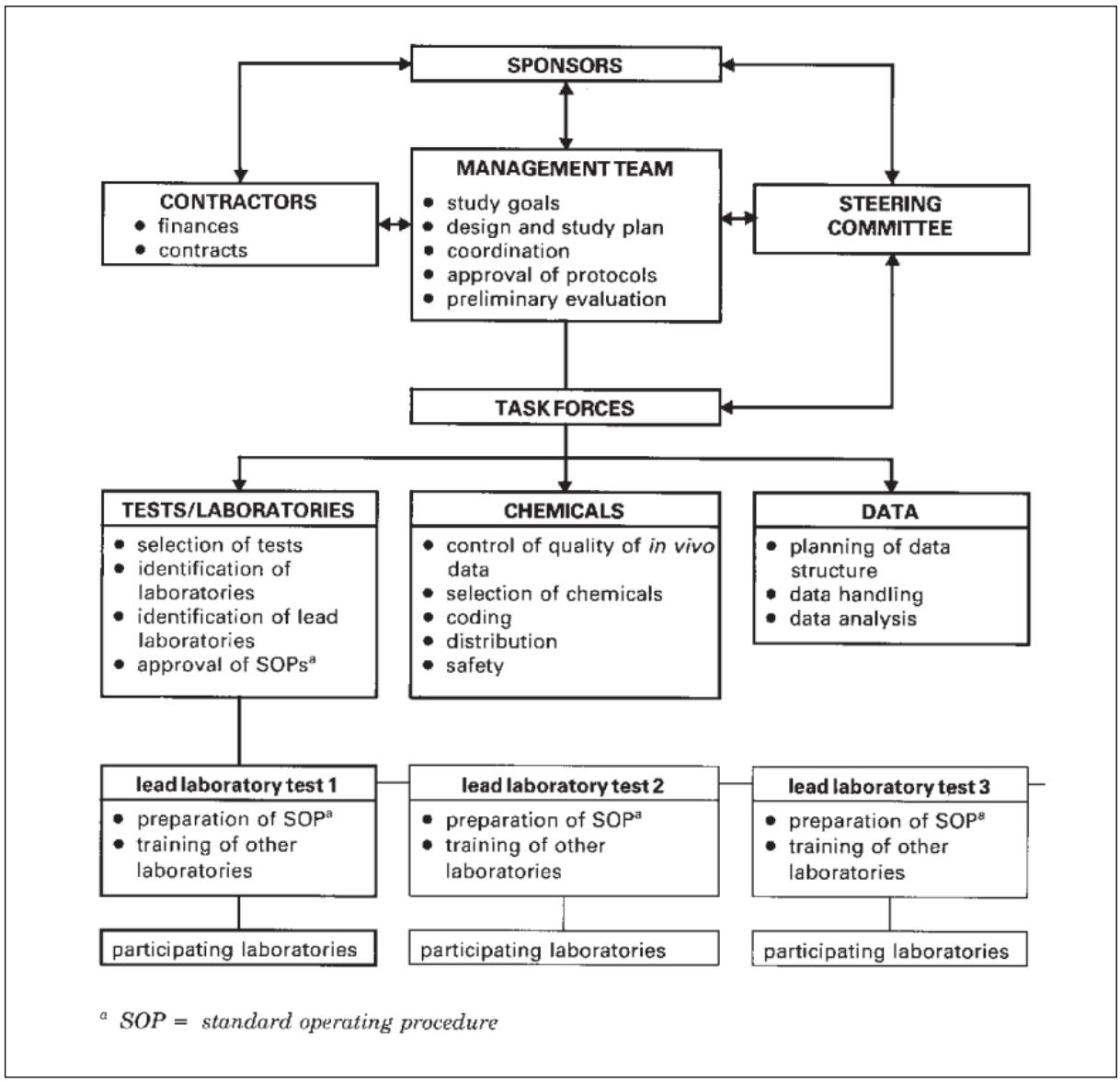

Figure 5.2: Structure of the management and organization for validation studes (Source: Balls et al, 1995)

Sponsors are generally 'found' by the management team, which is also responsible for (obtaining) financing for the study. Participating laboratories have to be selected, preferably on the basis of demonstrable competence in the test that is being validated. The management team and steering committee in turn can create 'task forces' and charge them with specific tasks such as the selection of laboratories and tests to be included, overseeing data collection, arranging transportation between labs, etc. The structure of such validation has generally remained the same over the years, and is depicted in figure 5.2.