Stem Cell Toxicology:
Ethical and Epistemic Constraints
on *In Vitro* Models

Grant Fisher

Abstract: The use of *in vitro* stem cell models in toxicology represents an important opportunity to engage with the interplay of ethical and epistemological issues in regulatory science and technology. Stem cell toxicology has been proposed to tackle epistemological, ethical as well as practical problems associated with the use of laboratory animals in toxicological studies to address a shortfall in chemical risk assessments. This paper argues that these developments are problematic if viewed as simply ameliorating these problems in the near term. Stem cell toxicology arises within a relatively novel intersection of the ethics and epistemology of pluripotent stem cell research and animal experimentation. It appears to require an expansion and a diversification of ethical and regulatory oversight due to epistemological and regulatory dependencies on therapeutic stem cell biology, the entrenchment of data from animal experimentation in toxicology, and the potentially novel implications of some aspects of the research. Understanding the role of stem cell toxicology models as model *for* will help to grapple their role in the transfer of knowledge between non-human animal models and humans as target systems. But advancing chemical risk assessment will not be a matter of simply addressing a normative problem by scientific and technological means.

Keywords: *toxicology, pluripotent stem cells, animal experimentation, replacement alternatives, ethics, in vitro models.*

1. Introduction

For many years there has been a strong motivation to find alternative methods to investigate the potential toxicological effects of chemical compounds on humans as well as their broader environmental impacts. Important pragmatic motivations include the cost of toxicology studies to government agencies and industry and the significant shortfall in the testing of a rapidly growing list of new chemical agents in circulation. In the US for example, the

Toxic Substances Control Act (TSCA) ensures that the US Environmental Protection Agency (EPA) compiles and publishes a list of chemicals that now stands at roughly 85,000 entries (United States Environmental Protection Agency Chemical Substance Inventory). The American Chemistry Council – a representative of the US chemical industry – argues that "the chemical industry is one of the most heavily regulated in the United States". While they add that the EPA has reviewed more than 36,000 chemicals and 2,700 have been subjected to regulatory action, whatever the presently accurate figure, regulatory screening is a central concern to government agencies and the chemical industry.

The heavy dependence on *in vivo* animal models in toxicological studies is evidenced by the huge volume of animals consumed. For example, the testing of new drugs for developmental toxicity in the US still follows the core directives of the US Food and Drug Administration (FDA) safety tests called the Goldenthal Guidelines introduced in 1966. An excess of 1200 animals (including fetuses) are consumed for safety tests (DeSesso 2017). It is estimated that 70% of the cost of European Union's 2006 regulations on animal experimentation is devoted to the testing of chemicals on animals, where 3,200 rats are consumed per chemical (Hou *et al*. 2013). Higher throughput, less laborious and cheaper methods include alternative animal models (for example, Zebrafish), the development of *in silico* computational toxicology (such as the study of quantitative structure activity relationships QSAR), and *in vitro* cellular assays including the use of stem cells.

Of course, the material costs of animal experimentation are but one significant factor. The ethical and regulatory implications of animal experimentation are another. In recent years there has been growing interest within the philosophy of chemistry on the ethics of chemistry and the governance of chemical risk. At the forefront of this work is engagement with the moral responsibilities of synthetic chemists and negative public responses to chemistry as a field. The two issues are intimately linked. Joachim Schummer argues that calculating chemical risks are difficult because "every new substance has an infinite potential of unpredictable properties […], such that risks are unpredictable" (Schummer 2001, p. 117). Given that risk assessments have a subjective component – "two people may differ in their moral judgment of a general risk inducing action, without having a superior moral level for 'objective' decisions" (*ibid*.) – this places increasing ethical responsibilities on chemists in their dealings with chemical risk and explains the often negative public reaction to new chemical substances. In a similar vein, Jean-Pierre Llored argues that the European Union's REACH (Registration, Evaluation, Authorisation and Restriction of Chemicals) regulations "enhances research in toxicology and ecotoxicology" while challenging existing approaches to safety (Llored 2017, p. 99). There are "alternative procedures using inherent

characteristics of substances, and amplify factors of damage or determinants of scale to identify filters, thresholds, and screening conditions" (*ibid.*). These regulatory changes contribute to the increasing adoption of alternative methods to replace *in vivo* studies. In the US, the Tox21 Consortium comprising federal agencies including the Environmental Protection Agency, the National Toxicology Program, and the Food and Drug Administration has become a major impetus to research on alternatives to whole animal models (National Research Council 2008). It is perhaps only becoming feasible to tackle the regulatory shortfall because of increased focus on alternative methods that aim at refining, reducing, or replacing the use of animals in toxicological testing. However, the ethical and epistemological implications of alternatives to the use of *in vivo* animal models in toxicology have thus far not received much attention in philosophy of chemistry.

In this paper, I focus on stem cell toxicology (SCT) – a major recent development aimed at tackling the normative problems of chemical risk assessment. My argument is that SCT does not overcome the ethical, regulatory, and practical problems needed to tackle the shortfall in chemical risk assessments at least in the near term. The problems are structural and these structural characteristics drive some of the ethical and epistemic constraints on the near term development of the field. SCT arises at the intersection of the ethics and epistemology of therapeutic stem cell biology and animal experimentation in toxicology. Rather than reduce the practical and regulatory pressures on the new research, it may well increase them. SCT is a nascent field intimately connected with therapeutic stem cell biology – the biomedical field associated with pluripotent stem cells (cells capable of self-renewal and differentiation into the three main tissue types). Recent work in the philosophy of biology has highlighted the interconnections of representations in therapeutic stem cell biology across a network of models including the cell cultures themselves (Fagan 2016). SCT, as a redeployment of this network to toxicology, is epistemologically dependent on the network of models in therapeutic stem cell biology. This dependence is especially obvious in the standardization of cell cultures, culture media, and laboratory protocols. Even established stem cell lines are subject to quality control before they can be investigated for their applicability to toxicology.

Due to the proposed use of human as well as non-human animal cells to develop *in vitro* models, SCT will inherit many of the well-known ethical problems and regulatory restrictions associated with therapeutic stem cell research.[1] In addition, while SCT is proposed as a strategy to develop replacement alternatives for non-human animals, agreement over evidential standards in toxicology is a vexed issue. In spite of the problems with animal experimentation, toxicology still remains dependent on animal experimentation data. The presumption that this data contributes to the validation of al-

ternative approaches to toxicological testing cannot be ruled out. This might be problematic if, on the one hand, the quality of this data is evinced given the well-known drawbacks to extrapolation from non-human animals to human targets in toxicology, while on the other hand it contributes to the validation process. If animal studies contribute to the validation of new models at least in the near term, then it will contribute to the ethical and regulatory constraints on SCT. There might also be a diversification of ethical and regulatory constraints. Although the ethical issues associated with SCT are very much an extension of existing problems associated with embryonic stem cell research and animal experimentation, potentially novel ethical problems arise, for example, with 3D brain organoid models that offer opportunities to explore developmental and other forms of neurotoxicity. They may introduce novel entities that require additional regulatory protections and ethical oversight.

The fundamental message I hope to convey is that for the above reasons, SCT will raise ethical and regulatory issues of potential interest to a wide range of stakeholders. If not communicated carefully – for example if the problems are treated as mainly *technical* problems associated with getting novel *in vitro* models to 'work' in toxicology – it could undermine stakeholder confidence in tackling the normative problems of chemical risk. Addressing the shortfall in toxicological testing requires less of an inclination towards seeking purely technoscientific fixes to normative problems and more recognition of the kinds of commercial and pragmatic pressures that underlie the current shortfall in toxicological testing.

The paper proceeds as follows. In Section 2, I discuss the 'high fidelity fallacy', which is used by some contemporary practitioners to motivate SCT. The fallacy is an important part of William Russell and Rex Burch's *Principles of Humane Experimental Technique* (1959). Drawing connections with a more recent view on the problems of inter-species extrapolation in philosophy of biology, I argue that some forms of SCT make the high fidelity fallacy redundant. I then outline how the embryonic stem cell test represents a central SCT model. In Section 3, I address problems for the standardization and validation of SCT models. Some of the normative implications of the dependency of SCT on therapeutic stem cell biology and a continued reliance of data from studies of non-human animals in toxicology are outlined in Section 4. By highlighting the juxtaposition of ethical and regulatory issues at the intersection of the ethics of animal experimentation and the ethics of pluripotent stem cell research, I argue that there may be an increase of the ethical and regulatory constraints on novel *in vitro* alternatives. Furthermore, the potential development of brain organoid models may diversify those constraints. In Section 5, I draw on some recent philosophical literature on *in vitro* models and the distinction between models *of* and models *for* in order to

outline an important generative function for SCT models. I then draw my conclusions.

2. The High Fidelity Fallacy and the Embryonic Stem Cell Test

Toxicity in humans is difficult to investigate for the obvious ethical reason that humans cannot be deliberately exposed to harm for scientific research. This has resulted in the use of laboratory animals such as the laboratory mouse and the rat as surrogates. The problem with using animals is that it sidesteps one ethical issue while introducing another one, and – as the thalidomide case tragically demonstrated – extrapolation of data from model organism to humans can be deeply flawed because of profound asymmetries in the effects of teratogens (compounds toxic to the developing embryo) on humans and laboratory animals. The thalidomide case stands out as a stark reminder of conflicting results from animal studies as well as with the unforeseen consequences (secondary effects) associated with chemical substances for humans (Ruthenberg 2018). As some practitioners have pointed out, the motivation for the study of developmental toxicity is in no small way associated with the thalidomide crisis (DeSesso 2017).

Recognition of the epistemological and ethical drawbacks to animal experimentation are perhaps most widely associated with Russell and Burch's influential *Principles of Humane Experimental Technique* (1959). Russell and Burch are best known for the 3 Rs (replacement, reduction, and refinement) in the ethics of animal experimentation, which has subsequently become part of a 'welfarist-reformist' attitude, meaning that researchers seek "successive short-term improvements to the status quo" in order to eventually realize the goals of animal rights by acknowledging the ethical limits to animal suffering while also acknowledging that human life deserves greater moral consideration (Franco 2013). Among other things, Russell and Burch also presented relatively early theoretical reasons to doubt the veracity of causal claims derived from mammalian models when extrapolated to humans as well as offering positive suggestions about what kinds of models might be used instead. And this has come to have some influence on the nascent field of SCT. Although hardly a ubiquitous motivation in the field, Russell and Burch's 'high fidelity fallacy' has come to be a spur to some SCT researchers (see Faiola *et al.* 2015). The appeal of SCT is that mouse and human pluripotent stem cells may offer potential *in vitro* replacements models for whole animal *in vivo* models. Human pluripotent stems cells offer the appeal of not crossing the species barrier. And by taking advantage of developments in the reprogram-

ming of human somatic cells to a pluripotent state – human induced pluripo-
tent stem cells (hiPSC) – it may be possible "to assess toxicity without the
ethical issues associated with the derivation and use of hESCs [human em-
bryonic stem cells]" from human embryos (*ibid.*, p. 5848).

The 'high fidelity fallacy' refers to the mistaken idea that mammals, due to
their relative phylogenetic similarity to humans, necessarily provide the best
model systems for human toxicological studies (Balls *et al.* 1995, p. 852).
Russell & Burch argued instead that in replacing one system with another,
discrimination trumps *fidelity*. Discrimination models able to reproduce a
particular property very well in a very *different* physical system are to be pre-
ferred for ethical and epistemic reasons to mammals with high *fidelity* in the
sense that model and target share *general* physiological and biochemical prop-
erties. Russell and Burch used the Herring Gull experiment to illustrate the
idea. The experiment demonstrates how the frequency of recently hatched
Herring Gull chick begging behavior increases when confronted by a stick
with a red tip and three well-marked white lines rather than a realistic model
of an adult Herring Gull head and bill (Balls 2013, pp. 12-13). An example
from toxicology is the *Limulus* amoebocyte lysate test, which remains a
standard drug test replacing rabbits by creating assays using blood cells from
the Atlantic Horseshoe Crab (*Limulus polyphemus*). The fidelity fallacy is
used to explain why model organisms, such as the laboratory rat, can fail to
demonstrate symmetric responses to humans when exposed to potential tox-
ins. Russell and Burch were proposing an ethically and epistemologically ad-
vantageous form of black boxing where the desired correlation between a
specific property instantiated in the models and target system is to be sought
irrespective of the biological or physical principles underlying the justifica-
tion of the model.

Russell and Burch's fallacy represents an early recognition of a problem
now well-established among critical appraisals of the status quo in animal ex-
perimentation. In some ways at least, the fidelity fallacy shares concerns
raised in Hugh LaFollette and Niall Shanks's (1995) account of the failure of
animal tests to establish the causal aspects of human disease. LaFollette and
Shanks ascribe this failure to the "modeler's functional fallacy": "Functional
similarity does not guarantee underlying causal similarity, nor does it make
such similarity probable" (*ibid.*, p. 150). Hence even if, for example, cats,
rats, pigs, and humans all metabolize phenol in order to aid excretion, the
mechanisms vary considerably (*ibid.*). Furthermore, while a common evolu-
tionary heritage suggests that animal models may provide "hypothetical ana-
logue models" that can spur basic research in biomedical science based on
functional similarities, they do not provide "causal analogue models" that
represent the mechanism of most interest that depend on evolved properties
and subsystems (*ibid.*, p. 151). So, while non-human primates demonstrate

phylogenetic continuity with humans, this does not guarantee similarity of mechanism either, for to draw an inference such as this would be to commit the "modeler's phylogenetic fallacy" (*ibid.*).

Russell & Burch's aim was to suggest how one might look for more reliable models of biomedical phenomena in phylogenetically dissimilar models (discrimination models). In this sense then, the high fidlelity fallacy is similar to LaFollette and Shanks's modeller's phylogenetic fallacy. The use of discrimination models, however, might nonetheless commit the modeller's *functional* fallacy. While discrimination models do not rely on phylogenetic similarity, presumably the reproduction of a particular property in a different system would appear to depend on an attribution of functional similarity in order for the discrimination model to be applicable to its target in toxicological studies. In any case, as I will elaborate below, the problem with the high fidelity fallacy is that perhaps it is redundant in some cases, such as in *in vitro* models utilizing *human* pluripotent stem cells. But first, what kinds of systems are stem cell toxicology models?

Arguably the most successful SCT model at present is an *in vitro* assay known as the murine embryonic stem cell test (EST), which has been validated by the European Centre for Validation of Alternative Methods (ECVAM) (Spielmann *et al.* 2001, pp. 31-32). It is the only validated *in vitro* pluripotent stem cell assay that does not destroy animals directly because it uses established murine stem cell lines. The murine EST uses suspension cultures called embryoid bodies to derive pluripotent embryonic stem cells. These cells are used to determine the toxicity of potential teratogens by investigating the ability of the stem cells to differentiate into healthy beating cardiocyocyte (cardiac muscle) cells. It is also used to determine cytotoxicity (cell viability) and is widely used in the drug industry to test lead compounds in preclinical trials (Seiler *et al.* 2011, p. 963) and by the US Environmental Protection Agency (Liu *et al.* 2017, p. 1529).

Although widely used and validated, the murine EST suffers a number of drawbacks. First and most obviously, it utilizes murine pluripotent stem cells. *Prima facie*, extrapolation to human targets suffers the same problems as the animal models it was intended to replace (Liu *et al.* 2017, p. 1529). Second, it focused on a specific morphological endpoint – the formation of beating cardiac muscle cells. However, this reveals little information about developmental toxicity in other tissues, such as nerve cells, or more complex systems like organs as well as the associated metabolic activity that this kind of *in vitro* models leaves out. An added difficulty with the model is that it is difficult to acquire accurate data on the beating heart muscle cells. It requires considerable experience and skill to make the careful observations necessary to collect data and avoid error (Seiler *et al.* 2011). Other forms of the murine EST focus instead on gene and protein expression while others shift from

murine to human embryonic stem cells (hESC) in order to overcome the problems of interspecies extrapolation. Important examples of the human EST can even take metabolomics into account where the molecular by-products of cell metabolism are used as biomarkers of developmental toxicity. These can be used to construct quantitative predictive models as well as to gain insights into underlying biochemical mechanisms in the early stages of human development. There is an understandable desire to improve the validity and generality of EST across a network of models: from murine to human stem cells, cardiac muscle cells to nerve and other cells, from morphological to molecular endpoints, by integration with metabolomics, generating predictive statistical models and mechanistic modeling of metabolomic pathways, and so on.

Although the high fidelity fallacy may motivate the idea that mammalian models are an unreliable guide to developmental toxicity in humans, it does not necessarily motivate the idea that SCT are comparable replacements to discrimination models (if that is indeed what some toxicologists intended by referring to the fidelity fallacy in motivating SCT). To describe them as 'discrimination' models seems to be in tension with the phylogenetic advantages to procuring *human* cells to model human developmental toxicity. Rather, one might argue that SCT models and their derivatives possess an epistemic advantage relative to alternative bioassays, particularly non-human animals. An argument from epistemic advantage might attempt to demonstrate that human pluripotent stem cells are the best source of stem cells for *in vitro* models for regulatory toxicology, thereby ameliorating problems of interspecies extrapolation. Furthermore, the use of 'reprogrammed' human somatic cells (hiPSC) might allow for the representation of the effects of human genetic diversity on toxicological responses at different doses. SCT would be best envisaged as moving beyond the distinction between 'fidelity' and 'discrimination' models. Presumably the potential uses human pluripotent stem cells might be put – for example as a basis for more complex human 'organoid' assays as I will discuss below – would also tend to suggest that developments in SCT undercut the problems associated with the fidelity fallacy.

3. Standardization and Validation

Nevertheless, SCT is facing considerable technical challenges. An important issue in the drive to reduce or replace animal experimentation in toxicology generally concerns how to standardize laboratory practices, techniques, and protocols to ensure quality control and how to validate new models. Standardization and validation would seem to presuppose agreed evidential stand-

ards and relevant background assumptions in order to compare the performance of alternative models in preparation for regulatory approval. Problems of standardizing laboratory protocols, cultures, and other factors can inhibit the reliability of *in vitro* models. For example, primary laboratory rat cells tend to provide phenotypically stable *in vitro* models due to their standardization *in vivo* – as a consequence of the primary cells being derived from model organisms. Models using isolated cells, such as hiPSC model cardiomyocytes, are derived from multiple sources, using different (often proprietary) protocols, cell cultures, conditions, and maturation times that can lead to heterogeneous cell populations (Gintant *et al.* 2017, p. 4). Furthermore, there are important differences in phenotype, metabolism, and gene expression rendering hiPSC cardiomyocyte models more like fetal cells than adult cells (*ibid.*). *In vitro* hiPSC models in toxicology may generate problems for *intra*-species extrapolation.

The use of hiPSC in EST for studies of developmental toxicity – where the aim is to determine the effects of potential teratogenic agents on the capability of pluripotent stem cells to differentiate and form healthy cardiomyocytes and other cells, such as hepatocytes and neural cells – obviously requires cell lines of sufficient quality to ensure the veracity of the toxicological test. A recent report by the US National Academies of Sciences, Engineering, and Medicine highlights how cell cultures of all kinds "should be characterized sufficiently before, during, and after experimentation. Genetic variability, phenotypic characteristics, and purity should be reported in published literature or on publicly accessible web sites or interfaces" (National Academies of Sciences, Engineering, and Medicine 2017, p. 68). Quality control in all *in vitro* modeling remains an important issue and indicates that the quality of pluripotent cell lines to be used in toxicology depends on reliable standards.

Compare the above to quality control of hiPSC *within* therapeutic stem cell biology. The reprogramming of murine somatic cells and its subsequent extension to human somatic cells (Takahashi & Yamanaka 2006, Takahashi *et al.* 2007) offered a much sought after alternative to the derivation of human pluripotent stem cell lines from human embryos. In order to validate hiPSC models, practitioners had to determine their pluripotent potential. This initiated studies drawing on the more established knowledge and practices of human embryonic stem cell biology. hESC are the epistemic standard used to determine the pluripotent differentiation potential of alternative stem cell derivation methods and cultures including hiPSC (Fagan 2013). hiPSC cannot be expected to replace hESC in the near term in therapeutic stem cell biology and developmental biology (Maienschein 2014). So long as hESC lines fulfill an indispensable epistemological function in the validation of the pluripotent potential of hiPSC lines, they cannot simply by-pass the ethical is-

sues associated with hESC, such as the destruction of human embryos including those associated with the procurement of human ova. hiPSC fails as a 'technical solution' for ethical problems in human embryonic stem cell research (Devolder 2015).

A similar predicament arises when pluripotent stem cells are redeployed in toxicology. For example, toxicologists working for the Stem Cell Toxicology Group at the US National Institute of Environmental Health Sciences claim that the uptake of hESC in toxicology has been relatively slow in spite of the envisaged advantages of hESC in not crossing the 'species barrier' and that this might be as a consequence of ethical and legal restrictions (Luz *et al*. 2018, p. 32). While they argue that this means there are significant drawbacks with hESC compared to hiPSC (as well as little evidence that hESCs are 'better' developmental toxicity assays), use of the latter – even well-established hiPSC lines – depend on standards established for hESC lines in therapeutic stem cell biology in to order determine and maintain the quality of all stem cell lines. While the use of hiPSC lines in toxicology possess potential epistemological advantages, such as helping to understand differences in the response to toxins within the population (National Academies of Sciences, Engineering, and Medicine 2017, p. 57) and "personalized drug design" and "personalized toxicology" (Sahu 2017), the supposed ethical and regulatory advantages of hiPSC compared to hESC to toxicology are not clear cut for reasons analogous to those in therapeutic stem cell biology.

While standardization of cell lines in toxicology is dependent on the practices of therapeutic stem cell biology, the *validation* of these alternative models is another matter. Validation of alternative models in toxicology aims to establish applicability, relevance, or 'fitness-of-purpose' to a specific problem in toxicology as well as their reliability as testing strategies that are reproducible across laboratories. One important asymmetry between therapeutic stem cell biology and SCT is that unlike the validation of novel stem cell assays in therapeutic stem cell biology, toxicology appears to lack consensus on whether there is a 'gold standard' – an agreed evidential standard of validation – or where there is agreement on evidential standards, toxicologists are divided as to its quality (National Academies of Sciences, Engineering, and Medicine 2017, p. 111). Model systems including the rodent cancer bioassays, and reproductive and developmental toxicity studies in rodents, among other model systems that have "inherent shortcomings and imperfections", have become "nearly indispensable for risk assessment" (*ibid.*, p. 108). One problem here is the potential role of this data in validation studies. If data from non-human animal experimentation contributes to establishing the applicability of alternative models as well as their predictive value, then this might undermine the veracity of that project given the problems of extrapolation discussed above. Other problems concern how limitations associated with

existing assays motivate novel forms of whole animal assays. For example, assays such as the US EPA's 'ToxCast' have relatively narrow applicability due to their design for the needs of the pharmaceutical industry. Both Tox-Cast and Tox21 lack assays for carcinogenicity, while 'cell-based' assays often miss biological responses at higher levels of biological complexity (*ibid*., p. 67). Among the US National Academies of Sciences, Engineering, and Medicine (NASEM) recommendations to address these problems are "targeted rodent tests", genetically diverse rodent species (to tackle a lack of knowledge of population wide sensitivities to toxicity) and transgenic model organisms alongside expanding the application of -omic technologies (*ibid.*, p. 4; p. 68). Toxicology appears to be 'locked in' to the technoscience of non-human animal *in vivo* models at least for the immediate future.

Presumably there is a rationale to underwrite the epistemological legitimacy of *in vivo* data despite the known problems. According to NASEM, species differences are predominantly down to differences in pharmacokinetics and metabolism, which might be clarified by mechanistic molecular level studies (*ibid.*). That inter-species differences might be predominantly a matter of differences in pharmacokinetics and metabolism suggests a potential advantage to *in vitro* models that abstract away from some of these features by omitting them, as in non-human and human forms of EST. One implication might be that non-human *in vitro* pluripotent stem cell assays do not necessarily recapitulate precisely the same problems of extrapolation as *in vivo* animal assays. Nevertheless, mechanistic studies required to underwrite causal claims from *in vivo* test system to human targets would be needed to legitimate the use of legacy *in vivo* toxicity studies to validate *in vitro* models.[2]

Yet it remains clear that the epistemological differences between *in vivo* and *in vitro* experimental systems can be profound. For example, developmental toxicologist may use animal models to investigate how maternal metabolism can render normally benign compounds toxic in ways that cannot be captured by *in vitro* pluripotent stem cell assays (Luz *et al*. 2018, p. 36). And for other reasons, even the validated murine EST leads to false negatives for known, strongly embryotoxic substances such as methyl mercury (Seiler *et al*. 2011, p. 962). Establishing exactly how, why, and in which contexts false negatives arise are not simply problems of abstraction and idealization in modeling. There are other contributing factors, such as errors resulting from attempts to coordinate results across laboratories with different protocols and even different classifications of toxicity. As well as the inability of the early murine EST to predict the strongly embryotoxic effects of methyl mercury, they also fall short in predicting the embryotoxicties of heavy metals such as cadmium and arsenic, but this can be down to problems with statistical modeling or the choice of toxicity classification scheme. In other cases of

toxicity, errors can result when models are constructed to investigate different stages of the developmental process (Seiler et al 2011, p. 963). Then again, stem cell toxicology studies using *human* cells are much less advanced than the use of murine stem cell systems. Species specific information regarding molecular mechanisms cannot be directly transferred to human pluripotent cells such as hESC, making it difficult to validate SCT models because there are "fewer standardized differentiation procedures for hESC" (*ibid.*; Yao *et al*. 2016, p. 443).

Alternatively, more evidence of the effects of chemicals on humans might be gained through human epidemiological studies. While even moderate defenders of the use of animals in toxicology recognize the epistemological and ethical limitations of animal studies, calls for a shift in the standards of evidence to human epidemiological studies may not only demand unethical methods, they may be used by industry to delay regulatory action when "high-powered animal tests [on carcinogenicity] are more reliable than typical low-powered, human epidemiological tests" (Shrader-Frechette 2008, p. 3). Such a view may go some way to suggesting epistemic and political strengths to animal studies in toxicology. But redeploying pluripotent stem cells from therapeutic stem cell biology to regulatory toxicology to provide alternatives to animal experimentation without an independent epistemological standard of sufficient quality to validate novel *in vitro* and *in vivo* assays for toxicity testing may also contribute to the slow regulatory acceptance of alternatives to animals. Even in the presumably less problematic case of alternatives to whole animal studies of embryotoxicity, such as the murine EST, validation does not coincide with regulatory acceptance (Luz *et al*. 2018, p. 36).

To summarize, the redeployment of pluripotent stem cells for toxicological testing transfers some epistemological characteristics of therapeutic stem cell research and is dependent on this research for the standardization of its experimental practices and models. Therapeutic stem cell biology is a complex interconnected network of stem cell systems or representations (Fagan 2016). SCT can be thought of as an attempt to extend part of this network to toxicology. Furthermore, in spite of the vexed issue of evidential standards within toxicology, SCT is dependent on the existing entrenched practices of toxicological testing and chemical risk assessment using *in vivo* data and models. But SCT problematizes the use of non-human and human tissues in additional ways. In the next section, I argue that the intersection of the ethics of pluripotent stem cell research and animal experimentation is significant and problematic in ways that might have important implications for stakeholder responses to the nascent field. While this need not entail new ethical issues, there is at least one example of the potential direction of research that could have more novel ethical and regulatory implications.

4. Increase and Diversification of Ethical and Regulatory Constraints

There are potentially a wide range of stakeholder positions on the ethics of both animal research and the ethics of human pluripotent stem cell research. The deployment of human pluripotent stem cells for *non*-therapeutic ends – such as in toxicology – is a comparatively under-explored issue. Like therapeutic stem cell biology, the aims of SCT are in the public interest but the redeployment of models for the development of alternative toxicological assays will not only rehearse existing ethical problems and regulatory concerns surrounding therapeutic stem cell biology. An ethics of SCT focused on replacement alternatives in toxicology is structured by the intersection of the ethics of embryonic stem cell research and the ethics of animal experimentation. While *in vivo* experiments have also played an important role in the development of therapeutic stem cell biology, developments in SCT engage with perhaps a broader range of views when the desire to reduce potential harms to the public as a result of the shortfall in toxicological testing is cast in the context of a 'comparative' ethics associated with the intersection of the ethics of human embryonic stem cell research and the ethics of non-human animal experimentation. There may well be an increase of ethical and regulatory concerns about the research arising from conflicting views on the veracity of data from animal studies alongside the continued dependency of toxicology on animal data – even when attempts are made to provide alternative models using human cells – alongside ethical and regulatory burdens associated with the procurement of human tissues. The intersection is likely to increase the number of stakeholders and increase stakeholder concerns about the direction of this research. This will require care in how research and regulatory developments are communicated so as to not inflate expectations about the ways in which alternative technologies can address ethical and regulatory problems in toxicology.

The intersection of the ethics of human embryonic stem cell research and the ethics of animal experimentation does not simply represent a contrast between anthropocentric and non-anthropocentric ethics. For some, the failure to recognize the significance of the fidelity fallacy has resulted not only in the continued suffering of animals, but also in the continued suffering of humans due to the continued dependence on animals in flawed pharmacological testing (Balls 2014). One might think of the interlinking of harms to humans and animals resulting from the *status quo* in pharmacology and toxicology, informally, as an 'argument from shared (non-human animal and human) ethical burdens' in the service of replacement alternatives. Harms to sentient non-human animals and the indirect and unintentional harm to humans due to the flawed predictive nature of non-human animal models gen-

erate obligations to seek safer replacement alternatives. The idea, then, is that experimentation on non-human animals harms those animals and as a consequence of the predictive failures of these experiments it indirectly harms the purported human beneficiaries of that research. The aim of the tests is to use animals as surrogates for humans who cannot be experimented on directly, but due to the problems of inter-species extrapolation, it turns out that harms to humans can occur by proxy through a poor choice of model.

Perhaps a problem with an argument from shared ethical burdens is that in the case of SCT, there are good reasons to believe that seeking replacements for animal experiments will increase the demand for human tissues and it might also increase the demand for non-human animal studies if this data contributes to the validation of alternative models using human cells in spite of the problems with this data. The potential of SCT to contribute to the ethical and epistemological reform of toxicology should not be given a gloss. By redeploying human pluripotent stem cell research to toxicology, ethical and regulatory restrictions may increase considerably beyond the current constraints in toxicological modeling. It is not surprising then that some toxicologists note that an impediment to the development and validation of a human EST is the ethical problem of embryo destruction in the derivation of cells (Jannuzzi *et al.* 2016). Problems also arise with the use of established hESC cell lines and not simply for the derivation of primary cells extracted from human embryos. The use of established hESC lines may address regulatory issues in jurisdictions where research on human pluripotent stem cells is legally permitted but the derivation of stem cell lines is not permitted. Nevertheless, the ethical and legal restrictions surrounding hESC have slowed the uptake of these assays in SCT (Luz *et al.* 2018, p. 32). And while hiPSC research is particularly important in SCT and is being used to extend the EST to culture hepatocytes, which are potentially crucial assays for toxicology, the uptake of iPSC, like hESC, is also relatively slow due to technical difficulties with the cell lines such as the 'bias' towards cell lineage of origin that can undermine their use in developmental toxicity testing (*ibid.*).

Another concern with SCT in the reform of animal experimentation in toxicology is that it might be subsumed or sidelined by narrowly anthropocentric ethical concerns about the impacts of these new technologies. SCT entails broadening ethical concerns to include those associated with human stem cell research. Perhaps for some stakeholders it could entail too high an ethical burden. For example, human tissues are subject to more stringent protections than those afforded to animals and raise additional concerns over privacy and informed consent in human tissue procurement and storage, let alone concerns over the ethics of human embryo research. Whether this is the case is an empirical issue. Nonetheless, it seems reasonable to expect that as increasing work on EST using human pluripotent stem cells advance, there

will be increasing stakeholder interest in the ethics of SCT given that the ethics of human embryonic stem cell research and animal experimentation intersect.

Furthermore, while SCT may increase the need for ethical engagement and regulatory oversight, it could lead to a *diversification* of ethical problems and institutional oversight. Among the flaws attributed to the use of non-human animals to infer causal claims about toxicity in humans are the differences in metabolism and toxicokinetics between model and target. Model organisms can therefore misrepresent metabolism and the toxicokinetics in humans. But *in vitro* stem cell models, such as a human EST, are of limited scope since they focus on a single endpoint (such as the differentiation of healthy cardiomyocytes) and are abstractions in the sense metabolism is left out. As indicated above, some toxicologists refer to these abstractions as a means to explain why animal studies can fail to provide reliable data on human toxicology. In SCT, the problem of abstraction has perhaps more generative implications in the sense that it is an incentive to design model systems capable of better representing the toxicity of compounds.

Take the development of EST since the early 1990s. The scope of EST has been extended to include multiple endpoints not only associated with developmental toxicity but also functional toxicity as part of a growing network of SCT models. There is a tendency in the development of *in vitro* models such as human EST towards models purported to make increasingly general claims about their targets using the same kinds of physical systems governed by at least some of the same underlying scientific principles. As one source puts it, "[…] it seems clear that cell biology is entering a new era in which mechanistic studies can be conducted in systems that are close in physiology to human biology and that cell biologists should enthusiastically embrace this shift in emphasis" (Drubin & Hyman 2017).[3]

Drubin & Hyman refer here – among other things – to the developing technoscience of organoids. Organoids are self-organizing 3D structures derived from stem cells capable of recapitulating some aspects of organ function (Huch *et al.* 2017, p. 938). Among other things, organoids offer the potential to reduce animal experiments and may help in "closing the gap between preclinical drug development and human trials" (Bredenoord *et al.* 2017, p. 7). While organoid systems provide a new model system, much like other novel applications of human pluripotent stem cell research, organoid research will not abolish embryo research but might increase it instead because validation of these novel systems may require comparison with 'normal' human tissues at least in the near term (*ibid.*, p. 3). Others see these developments as a threat to model organisms and basic research from application-driven, human tissue derived systems, such as organoids derived from hESCs (Duronio *et al.* 2017, Huch *et al.* 2017). One of these sources argues

"organoids will not and should not replace non-vertebrate model organisms as discovery tools […] the best use of organoids may be to transfer knowledge acquired in model organisms to humans" (Duronio *et al*. 2017, p. 1386).

Human organoid systems may possess significant advantages to both animal models and EST. For example, recent work by James Thomson – whose research was responsible for the derivation of the first hESC lines – uses hESC to develop a 3D model of neural cells using bioinformatics to correlate toxicity affecting cell physiology with changes in gene expression profiles (Hou *et al*. 2013). This work is part of the National Institute of Health's Microphysical Systems Initiative,[4] which is motivated in part by well-known failures of animal experiments to warn of toxicity in humans (especially the thalidomide case) and because cognitive changes like autism are difficult or impossible to model in animals. *In vitro* models of human neurophysiology – or 'mini-brains' – may be capable of modeling functional aspects of the human brain to improve toxicological testing, pharmacology, and disease modeling (Pamies *et al*. 2017).

Speculation on the future development of brain organoids includes the possibility of conscious systems. Although these are currently only hypothetical, they may challenge an advantage to the use of human EST with respect to the ethics of animal experimentation. Given that *in vitro* assays based on human pluripotent stem cell lines (such as a human form of EST) are not conscious entities, an ethical asymmetry might be drawn between them and conscious *in vivo* non-human animals such as mice and rats. However, developments in brain organoid research to study neurotoxicity motivated in part by epistemic and applicability problems with animal models, suggests that if mini-brains became phenomenally conscious then their use should be subject to ethical and regulatory restrictions. Julian Koplin and Julian Savulescu (2020) argue that any advanced brain organoid system should be regulated under existing frameworks for stem cell research until the point at which consciousness is reached. While establishing when this happens in an (at present hypothetical) *in vitro* system would be a difficult issue, there are some forms of experimentation that are nonetheless permissible even on conscious entities. If this were not true, Koplin and Savulescu argue, then it would be impermissible to experiment on animals for worthwhile human ends. Consciousness does not mean we should not use mini-brains (assuming there is no other viable alternative) since, like animals, we do not afford them the same protections as person and, unlike human fetuses, lack the potential to become persons. We would, however, afford them similar protections to those currently afforded to laboratory animals.

In the context of reform of toxicology, if a motivation for developing organoid models is to replace non-human animals in neurotoxicity testing for

ethical reasons, then it would be dubious to attempt to do so by simply introducing *another* model system that raises the same or similar ethical objections regarding harms to conscious entities that do not count morally as persons. The argument that it would be permissible to use brain organoids in experiments for ends that promote human welfare is predicated on the assumption that we already tolerate it when we use conscious animals for legitimate human ends under ethical and regulatory oversight in biomedical and toxicological research. So, to be fair, the context is somewhat different. For example, Koplin & Savulescu suggest a reformulation of Russell & Burch's 3Rs alongside adopting a more comprehensive account of moral principles needed for brain organoid research (*ibid.*, p. 763). Since an aim of brain organoid technology is to study phenomena for important human ends that are not well instantiated in animal models, there are of course good ethical and epistemological reasons for doing so.

But at the same time, brain organoid models in the hypothetical advanced form we are considering here would nevertheless represent a potential *diversification* of ethical constraints on toxicology. Conscious brain organoids would be novel entities requiring new regulatory oversight in ways perhaps analogous to model organisms within an additional institutional arrangements given that protections would go beyond existing protections for stem cell research. And the issue, again, is that if the aims of toxicological reform are indeed to address the ethical, regulatory, and economic burdens of animal experimentation, there is something troubling about the strategy of doing so by creating potentially novel morally considerable entities as well as additional forms of oversight. In any case, we should not expect a novel technological intervention into normative problems of chemical risks to undercut the existing ethical and regulatory obstacles to tackling the shortfall in toxicity testing.

In summary, an argument from shared ethical burdens may not help to encourage novel *in vitro* stem cell assays to tackle the ethical problems associated with animal research if there is still a significant contribution of animal data to the validation of alternatives. Below I will briefly sketch a suggestion that attempts to distinguish a more modest and transitory role for animal data. SCT also appears to increase the ethical and regulatory constraints on toxicological research due to its dependencies on therapeutic stem cell biology. The ethical problems are not new but the juxtaposition of the ethics of stem cell research and animal experimentation is likely to result in increased stakeholder and regulatory scrutiny of the field. Furthermore, there are potentially novel ethical issues arising from advanced developments in *in vitro* modeling that might have applications in toxicology. Brain organoid models may require a diversification of ethical and institutional oversight in addition

to the increase of constraints arising from the intersection of therapeutic stem cell biology and animal experimentation.

5. *In Vitro* Models *Of* and *For*

One positive suggestion I would like to make before I close is that SCT practitioners are constructing and developing models with an important generative function in toxicology. SCT models can be thought of as an *in vitro* version of a kind of experimental model Marcel Weber calls "*in vivo* representations" – "a representation that shows the typical features of a model, such as being idealized, simplified, multiply-realizable, and so on, but where the thing that does the representing is alive" (Weber 2014, p. 765). SCT models might then be described as a form of '*in vitro* representation'. However, according to Weber, experimental models "provide knowledge of causal processes that generalize to systems where biologically and chemically different kinds of causes are at work" (*ibid.*, p. 764). In other words, they are like computational models used to construct simulations in the sense that they function as stand-ins, the systems modeled are multiply realizable by a range of potentially very different biological processes. SCT models as a form of '*in vitro* representation' would then differ from Weber's idea of *in vivo* representations not merely because of the material difference between *in vivo* and *in vitro* systems but also because *in vitro* representations lack this sense of multiple realizability and yet nonetheless 'stand in' for their target systems. But crucially, *in vitro* representations in toxicology studies stand in for *both* their targets *and* (at least purportedly) the animal models they are intended to replace. However, this does not mean that an *accurate* representational function is what most importantly establishes the current function of the models used in SCT.

Drawing on Evelyn Fox Keller's (2000) distinction between models *of* and models *for*, Emanuele Ratti (2018, p. 787) argues that in molecular biology, scientists can adopt alternative cognitive dispositions towards the same model.[5] On the one hand, they may focus on the explanatory power or representational accuracy of a model. On the other, scientists may be concerned with the *redeployment* of the model in ways that can result in new forms of experimental intervention, rather than the explanatory or representational aspects of the model. I would suggest that Ratti's elaboration of the models-of/models-for distinction can be extended to SCT. As *in vitro* models *of*, pluripotent stem cell models in therapeutic and developmental biology provide crucial mechanistic insights into cell differentiation and development enabling novel experimental interventions as *in vitro* models *for* translational

medicine. Moreover, the redeployment of pluripotent stem cell lines as models *for* developmental toxicology suggests new forms of experimental intervention. For example, the construction of the murine EST and later human forms of EST based on hESC and hiPSC are used to probe the effects of teratogens on the capacity of *in vitro* models to issue in healthy cardiomyocytes, hepatocytes, as well as other cells. This was done not simply to provide more accurate representations of the target systems, but rather to seek novel experimental interventions capable of offering plausible higher throughput models to study toxicity without relying on *in vivo* models for ethical and pragmatic reasons.

This redeployment has proved problematic. But this is perhaps in part due to a conflation of the different dispositions towards SCT models. For novel *in vitro* models to constitute models *of* human target system, they would already have needed to demonstrate what is in need of support, namely, that they are veridical surrogates for the study of human developmental toxicity. Admittedly, however, it is not just the epistemological uncertainties regarding the validation of these models that is in question but also their ethical advantages. In toxicology, as I hope will now be clear, epistemological and ethical factors interact in model design, standardization, and validation. As *in vitro* models *for*, SCT models play more of a legitimate generative function because they offer novel means of experimental intervention not possible prior to the redeployment. Due to this generative aspect to *in vitro* experimental models for, the ethical as well as epistemological strengths and weaknesses emerge along with these interventions as the models and their derivatives, such as organoid models, are used to extend the network of emerging models in toxicology. Furthermore, it can make somewhat more sense of the task of validating SCT models by referring to (among other sources) data from *in vivo* animal studies. Aside from the successful validation of the murine EST, the idea that toxicology remains significantly dependent on data from model organisms is problematic if these models as regarded as model *of* their targets. But as models *for* it may not be quite so difficult to reconcile the relationship of evidence to models if, as Duronio *et al.* (2017) appear to suggest, models in biomedicine and toxicology can be used to 'mediate' between model organisms and humans. Although the details of this transfer of knowledge between model organisms and humans via organoid models in this case would need to be clarified, the idea that novel experimental systems permitting new means of intervention in this context would lie somewhere between LaFollette and Shanks's distinction between hypothetical analogue models and causal analogue models. In any case, as *in vitro* experimental models *for*, SCT and their derivatives remain tied to the technoscience of model organisms and hence are unlikely to replace them in the near term. And in addition, as a redeployment of models in therapeutic stem cell biolo-

gy, they will not offer quick technical solutions to the normative problem of chemical risk.

6. Conclusion

The work on replacement alternatives alongside the reduction and refinement of animal models in toxicology is very important for ethical and epistemic reasons. However, SCT models and their derivatives do not overcome the ethical problems they were intended to solve. At present, they would appear to suggest an increase and diversification of ethical and regulatory constraints on toxicology. However, they serve a potentially crucial generative function as *in vitro* models *for* by offering novel forms of experimental intervention capable of probing the flawed transfer of knowledge between non-human animals and their human target. The broader message I hope this paper conveys is that there should be more focus on the ethics of replacement alternatives because of the possible negative public and policy responses if the kinds of drawbacks – not simply the technical drawback – are not acknowledged. Perhaps the predicament surrounding the shortfall in toxicological testing, which is an ongoing long term normative problem of chemical risk, will inevitably results in attention on 'downstream' technological responses to the problem. It remains to be seen whether the kinds of developments discussed here are not only technically viable, but also ethically viable and in terms of regulatory oversight. Moreover, we need to cultivate more 'upstream' analyses in a way that I take recent developments in the ethics of chemistry to be suggesting. For example, by scrutinizing the commercial incentives and pressures that contributed to the shortfall in toxicology testing in the first place might focus more attention on the decisions and motivations that have resulted in such a predicament, since it seems reasonable to suggest that the problems are best addressed not *post hoc* but also by employing more regulatory foresight.

Acknowledgements

Notes

[1] I will not discuss these issues at any length here. For a detailed introduction, see Devolder 2015.

[2] However, this would be an example of the 'extrapolator's circle' – the idea that once we have established that non-human animal models instantiate the causal similarities needed to underwrite inferences from model to target, it seems dubious that we would need the animal model to fulfill such a function (LaFollette & Shanks 1995, p, 157). For an account of how extrapolation from non-human animals to humans are legitimate in spite of causal disanalogies, see Steel 2008.

[3] One advantage is the capacity to develop organoids and while there are drawbacks (*e.g.*, the lack of vascularization) these kinds of problems are merely ones of 'engineering' (*ibid*.).

[4] See for example https://ncats.nih.gov/tissuechip/about.

[5] I thank an anonymous referee for drawing my attention to developments in the literature on models-of and models-for.

References

Ankeny, R.A. & Leonelli, S.: 2011, 'What's So Special about Model Organisms?', *Studies in History and Philosophy of Science*, **42**, 313-323.

Balls, M.; Goldberg, A.M; Fentem, J.H.; Broadhead, C.L. & Burch, R.L.: 1995, 'The Three Rs: The Way Forward: The Report and Recommendations of ECVAM Workshop 11', *Alternatives to Laboratory Animals*, **23**(6), 838 – 866.

Balls, M.: 2013, 'The Wisdom of Russell and Burch: Fidelity and Discrimination', *Alternatives to Laboratory Animals*, **41**, 12-13.

Balls, M.: 2014, 'Animal Experimentation and Alternatives: Time to Say Goodbye to the Three Rs and Hello to Humanity?', *Alternatives to Laboratory Animals*, **42**, 327-333.

Bredenoord, A.L.; Clevers, H. & Knoblich, J.A.: 2017, 'Human Tissues in a Dish: The Research and Ethical Implications of Organoid Technology', *Science*, **355** (6322), eaaf9414.

Chemical SafetyFaacts.org., 'Debunking the Myths: Are there Really 84,000 Chemicals?' [available online: https://www.chemicalsafetyfacts.org/chemistry-context/debunking-myth-chemicals-testing-safety/ accessed 14th June 2019].

DeSesso, J.: 2017, 'Future of Developmental Toxicity Testing', *Current Opinion in Toxicology*, **3**,1-5.

Devolder, K. 2015, *The Ethics of Embryonic Stem Cell Research*, Oxford: Oxford University Press.

Drubin, D.G. & Hyman, A.A.: 2017, 'Stem Cells: The New "Model Organism"', *Molecular Biology of the Cell*, **28**, 1409-1411.

Duronio, R.J.; O'Farrell, P.H.; Sluder, G. & Su, T.T.: 2017, 'Sophisticated Lessons from Simple Organisms: Appreciating the Value of Curiosity-driven Research', *Disease Models and Mechanisms*, **10**, 1381-1389.

Fagan, M.B.: 2013, *The Philosophy of Stem Cell Biology: Knowledge in Flesh and Blood*, Basingstoke: Palgrave MacMillan.

Fagan, M.B.: 2016, 'Generative Models: Human Embryonic Stem Cells and Multiple Modeling Relations', *Studies in History and Philosophy of Science*, **56**, 122-134.

Failoa, F.; Yin, N.; Yao, X. & Jiang, G.: 2015, 'The Rise of Stem Cell Toxicology', *Environmental Science & Technology*, **49**, 5847-5848.

Franco, N.H.: 2013, 'Animal Experiments in Biomedical Research: A Historical Perspective', *Animals*, **3**, 238-273.

Gintant, G & Braam, S.: 2017, 'Stem Cell-Derived Models for Safety and Toxicity Assessments: Present and Future Studies in the 'Proclinical Space', in: Clements, M. & Roquemore, L. (eds,): *Stem Cell-derived Models in Toxicology*, New York: Springer, pp. 1-15.

Hou, Z.; Zhang, J.; Schwartz, M.P.; Stewart, R.; Page, C.D.; Murphy, W.L & Thomson, J.A.: 2013, 'A Human Pluripotent Stem Cell Platform for Assessing Developmental Neural Toxicity Screening', *Stem Cell Research & Therapy*, **4** (Suppl 1), S12.

Huch, M.; Knoblich, J.A.; Lutolf, M.P. & Martinez-Arias, A.: 2017, 'The Hope and the Hype of Organoid Research', *Development*, **144**(6), 938-941.

Jannuzzi, A.T.; Ozcagli, E.; Kovatsi L.; Goumenou, M. & Tsatsakis, A.M.: 2016, 'Using Stem Cells in Toxicological Assessments', *Journal of Medical Toxicology and Clinical Forensic Medicine*, **2**, 1-2.

Keller, E.F.: 2000, 'Models of and Models for: Theory and Practice in Contemporary Biology', *Philosophy of Science*, **67**, Supplement (Proceedings of the 1998 Biennial Meetings of the Philosophy of Science Association. Part II: Symposia Papers), S72-S86.

Kirk, R.G.W.,: 2018, 'Recovering the *Principles of Humane Experimental Technique*: The 3Rs and the Human Essence of Animal Research', *Science, Technology, and Human Values*, **43**(4), 622-648.

Koplin, J.J. & Savulescu, J.: 2019, 'The Moral Limits of Brain Organoid Research', *Journal of Law, Medicine and Ethics*, **47**, 760-767.

LaFollette H. & Shanks, N.: 1995, 'Two Models of Models in Biomedical Research', *The Philosophical Quarterly*, **45**, 141-160.

Liu, S.; Yin, N. & Faiola, F.: 2017, 'Prospects and Frontiers of Stem Cell Toxicology', *Stem Cells and Development*, **26**(21), 1528-1539.

Luz, A.L. & Tokar, E.J.: 2018, 'Pluripotent Stem Cells in Developmental Toxicity Testing: A Review of Methodological Advances', *Toxicological Sciences*, **165**, 31-39.

Llored, J-P.: 2017, 'Ethics and Chemical Regulation: The Case of Reach', *Hyle: International Journal for Philosophy of Chemistry*, **23**, 81-104.

Maienschein, J.: 2014, *Embryos Under the Microscope: The Diverging Meanings of Life*, Cambridge Mass.: Harvard University Press.

Martinez-Arias lab: 2016, 'In Defence of Embryonic Stem Cells as a New Model System for Developmental Biology', Department of Genetics, University of Cambridge [available online: https://amapress.gen.cam.ac.uk/?p=2102, accessed 14th June 2019].

National Research Council: 2008, *Toxicity Testing in the 21st Century: A Vision and a Strategy*, Washington, DC: The National Academies Press.

National Academies of Sciences, Engineering, and Medicine.: 2017, *Using 21st Century Science to Improve Risk-Related Evaluations*, Washington, DC: The National Academies Press.

Pamies, D.; Barrera, P.; Block, K.; Makri, G.; Kumar, A.; Wiersma, D.; Smirnova, L; Zhang, C.; Bressler, J.; Christian, K,M.; Harris, G.; Ming, G.-L.; Berlinicke, C.J.; Kyro, K.; Song, H.; Pardo, C.A.; Hartung, T. & Hogberg, H.Y.: 2017, 'A Human Brain Microphysiological System Derived from Induced Pluripotent

Stem Cells to Study Neurological Diseases and Toxicity', *Alternatives to Animal Experimentation*, **34**, 362-376.

Ratti, E.: 2020, 'Models of' and 'Models for': On the Relation between Mechanistic Models and Experimental Strategies in Molecular Biology', *British Journal for the Philosophy of Science*, **71**, 773-797.

Russell, W.M.S & Burch, R.L.: 1959, *The Principles of Humane Experimental Technique*, London: Methuen.

Ruthenberg, K.: 2016, 'About the Futile Dream of an Entirely Riskless and Fully Effective Remedy: Thalidomide', *Hyle: International Journal for Philosophy of Chemistry*, **22**, 55-77.

Sahu, S.C.: 2017, *Stem Cells in Toxicology and Medicine*, Chichester UK, John Wiley & Sons.

Schummer, J.: 2001, 'Ethics of Chemical Synthesis', *Hyle: International Journal for Philosophy of Chemistry*, **7**(2), 103-124.

Seiler, A. & Spielmann, H.: 2011, 'The Validated Embryonic Stem Cell Test to Predict Embryotoxicity *in vitro*', *Nature Protocols*, **6**(7), 961-977.

Shrader-Frechette, K.: 2008, 'Evidentiary Standards and Animal Data', *Environmental Justice*, **1**(3), 1-6.

Speilmann, H. & Liebsch, M.: 2001, 'Lessons Learned from Validation of *In Vitro* Toxicity Test: From Failure to Acceptance into Regulatory Practice', *Toxicology In Vitro*, **15**, 585-590.

Steel, D.: 2008, *Across the Boundaries: Extrapolating in Biology and Social Science*, Oxford: Oxford University Press.

Takahashi, K. & Yamanaka, S.: 2006, 'Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors', *Cell*, **126**, 663-676.

Takahashi, K.; Tanabe, K.; Ohnuki, M.; Narita, M.; Ichisaka, T.; Tomoda, K. & Yamanaka, S.: 2007, 'Induction of Pluripotent Stem Cells From Adult Human Fibroblasts by Defined Factors', *Cell*, **131**, 861-872.

United States Department of Health and Human Services, National Institutes of Health: 2019, 'About Tissue Chip' [available online: https://ncats.nih.gov/tissuechip/about, accessed 30th December, 2019].

United States Environmental Protection Agency Chemical Substance Inventory [online available at: https://www.epa.gov/tsca-inventory/about-tsca-chemical-substance-inventory, accessed 14th June 2019].

Weber, M.: 2014, 'Experimental Modeling in Biology: In Vivo Representations and Stand-Ins as Modeling Strategies', *Philosophy of Science*, **81**(5), 756-769.

Yao, X.; Yin, N. & Faiola, F.: 2016, 'Stem Cell Toxicology: A Powerful Tool to Assess Pollution Effects on Human Health', *National Science Review*, **3**, 430-450.

Grant Fisher:
Graduate School of Science and Technology Policy, Korea Advanced Institute of Science and Technology (KAIST), Republic of Korea;
fisher@kaist.ac.kr