Chapter 4
Mechanistic vs Statistical Extrapolation in Preclinical Research in Psychiatry: Challenging the Received View

Catherine Belzung, Etienne Billette de Villemeur, and Maël Lemoine

C. Belzung
University of Tours, INSERM U930, Tours, France e-mail: catherine.belzung@univ-tours.fr

E. Billette de Villemeur
University of Lille, LEM (UMR CNRS 9221), Lille, France e-mail: etienne.de-villemeur@univ-lille.fr

M. Lemoine
University of Bordeaux, Immunoconcept, Bordeaux, France e-mail: mael.lemoine@u-bordeaux.fr

 Uncertainty in Pharmacology, Boston Studies in the Philosophy and History of Science 338, https://doi.org/10.1007/978- 3- 030- 29179- 2_4

Abstract This chapter questions the received view that in medical research extrapolation from animal models mainly consists in establishing mechanisms of human pathological states in organisms, thanks to a step by step comparison of causal pathways. *Mechanistic extrapolation* takes the form: *(1)* cause C brings out effect E in animal through causal pathway M, *(2)* M is similar in animals and humans, *(3)* therefore C will likely bring out E in humans. As the example of psychiatric research shows, such mechanistic extrapolation may be replaced by *statistical extrapolation*, an inference of the form: *(1)* An animal model A has been successful in predicting the effects E of drugs $D_1 \ldots D_n$ of a certain class; *(2)* A will be successful again in predicting the effects of a new drug D_{n+1} of the same class. Statistical extrapolation relies on the predictive validity of a given animal model, without any knowledge of the mechanisms involved, on the sole ground of past successes of the model in predicting the effects of a class of drugs on their human target.

4.1 Introduction

In spite of growing opposition to their use in the general public and notwithstanding the emergence of alternatives such as *in silico* simulation, animal models still constitute the standard experimental system in preclinical studies. Evidence drawn from animal models is deemed necessary before it is acceptable to test a drug on humans for the first time. This chapter focuses on the way standard preclinical studies provide such evidence: how do they establish that a drug is, or should be, effective for humans afflicted with a disease? This has often been summarized by philosophers as an extrapolative inference of the sort we may call *mechanistic*. It may take the following form:

(1)'fluoxetine relieves depressive-like symptoms in observed rodents'

(2) 'the mechanisms of fluoxetine action in rodents are similar in humans'

(3) 'fluoxetine relieves depressive symptoms in humans'

Conclusion (3) is generally held to be a prediction. Assumption (2) is considered as a comparison based on the investigation of mechanisms, that is, mechanistic reasoning. A mechanism is an integral causal process or pathway leading from an initial state to a final state. It is constituted of entities. These entities interact with each other in a determined way (Machamer et al. 2000). They may be biological entities, such as proteins, receptors, enzymes, genes or small molecules, which engage in biological interactions such as 'binds with', 'upregulates', 'methylates', etc. They may also be cognitive constructs, such as 'memory', 'short term memory', 'long term memory', 'attention', and cognitive interactions such as 'inhibits', 'retrieves', 'activates', etc. They may at last be social factors, such as 'single parent', 'abused child', and social interactions, such as 'parental love' and 'parental abuse'. Mechanisms may also be biological, psychological and social at the same time. For example, major depression is characterized by alterations of some affective/cognitive processes including anhedonia, stress coping and overgeneralization (Belzung et al. 2015). This can be counteracted by chronic treatment with fluoxetine. The mechanisms underlying this action are complex: fluoxetine inhibits the serotonin transporter, which leads to an increase of serotonin in the synaptic cleft, inducing stimulation of serotoninergic receptors of various types and causing the activation of a second messenger cascade in the cytoplasm of the post-synaptic neuron. This ends by the activation of transcription factors within the nucleus of the cell, leading to the synthesis of neurotrophic factors such as BDNF (Brain Derived Neurotrophic Factor), which might stimulate adult neurogenesis in brain areas in which newborn neurons can be observed during adulthood, such as the dentate gyrus of the hippocampus. The dentate gyrus displays specific computational properties such as sparse coding: a given stimulus induces the activation of very few cells, so that the probability that two different stimuli activate the same neuronal population is very low. This property in turn justifies a distinction between two closely resembling stimuli, improving pattern separation. Therefore, the fluoxetine-induced increase in adult newborn neurons may decrease the over-generalization displayed by depressive patients. In parallel, depressive patients display a disruption of the regulation of stress hormones. Again, the hippocampus and more precisely the dentate gyrus participates in the regulation of stress hormones: therefore, the fluoxetine-induced increase in adult neurogenesis might counteract the stress hormone dysregulation seen in depressed patients (for more details, see Willner et al. 2013).

We call the view that focuses on such extrapolation of mechanisms in preclinical studies the 'received view'. In contrast, our contention in this chapter will be that, at least in the case of treatments, inference of the following sort is sometimes considered sufficient to establish a claim of interest:

(1) 'fluoxetine relieves depressive-like symptoms in observed rodent models'

(2) 'the reaction in humans to similar drugs is frequently the same as that observed in rodent models'

(3) 'fluoxetine relieves depressive symptoms in humans'

In this alternative version, step (2) is not based on mechanistic reasoning,

but on the observation of correlations between the effects observed in the model organism's, and the ones seen in patients. We call it *statistical* extrapolation. A correlation is an association found to be frequent enough between two events: if it is close to 0, it is likely that the two events are independent; if it is close to 1, it is likely that they are concomitant, and if it is close to -1, that they are exclusive (i.e. one cannot be observed when the other is). In the case at hand, the events are the exposure to a drug and the general state of the patient. The form of extrapolation under scrutiny involves a correlation with a high value, i.e. 'close to 1'. In other terms, it is based on statistical reasoning instead of mechanistic reasoning.

At first sight, all preclinical studies rely on statistical and mechanistic information. Yet the difference between statistical and mechanistic extrapolation depends on the justification of the inference. In mechanistic extrapolation, the observation that an effect is caused by a mechanism in the model justifies the prediction that a similar mechanism in the target will cause a similar effect. In statistical extrapolation, the observation of certain effects in a model which has consistently predicted effects of drugs of a certain class in humans, justifies the prediction that the effects will be observed in humans. Although knowledge of mechanisms and statistical reasoning may be present in both cases, they do not play the same role. The very existence of statistical extrapolation as we describe it implies a strong claim that it does not presuppose mechanistic extrapolation in the first place as a justification. However, this does not mean that mechanistic information has not been used in the first place to select the initial hypothesis. Finally, note that we explicitly focus on the case of animal models in psychiatry, where the lack of well-defined mechanisms is notable. This constitutes a good reason why it is more often resorted to in this domain. However, it remains an open question whether statistical extrapolation is also resorted to beyond this field.

This chapter thus suggests an important correction to the philosophical received view on animal studies, which misleadingly focuses on the role that the investigation of mechanisms plays in extrapolation through what Steel calls "comparative process tracing" (Steel 2008). Section 1 presents the received view. Section 2 presents a case study where statistical extrapolation is used in place of mechanistic extrapolation.

Section 3 raises objections to the existence of such statistical extrapolation and answers them.

4.2 The Received View on Evidence Based Upon Preclinical Studies

The received view claims that preclinical extrapolation consists in the attribution to humans, by analogy, of an extra entity or activity, unknown (or incompletely understood) in humans but observed in animals. An implicit claim follows that evidence of a mechanism is necessary to the extrapolation of any causal claim to humans. We will explicate the two claims in turn.

4.2.1 Inferences from Preclinical Studies Consist in Comparative Tracing

The received view consists in a certain picture of how extrapolation in preclinical studies works, which we dub 'mechanistic extrapolation'. It is necessary to describe it in order to understand when evidence of a mechanism is required in preclinical studies, and to contrast mechanistic extrapolation with statistical extrapolation.

Mechanistic extrapolation is analogical reasoning, whereby the presence of a property in the target is inferred from its presence in the model. Some knowledge of the mechanisms the model and target share is necessary for it to be plausible. Yet we do not by definition know all of it, and we have to assess the extent to which relevant differences impinge on extrapolation. More specifically, it consists in the following line of reasoning about the mechanism, that is, the causal chain linking cause (C) to effect (E):

1) C* causes E* (in model)

2) C (in target) is analogous to C* (in model)

3) E(in target) is analogous to E* (in model)

4) C will cause E (in target).

In her classic book about models in science, Mary Hesse emphasized the role that unknown factors may play in causal analogies. She suggested that based on both "*positive"* and "*negative"* analogies, respectively, what is known to be and not to be common in model and target, analogical reasoning on a model consists in investigating "neutral" analogies, i.e., what remains unknown (Hesse 1970). In other terms, the problem is:

> An analogue *x* containing the characters *ABD* is a model for an *explicandum y* containing *BC* but not *A*, where *B* represents the total known similarity between *x* and *y*. If *A* and *B* together are known to be causally connected to *D* in the model, are there any rational grounds for expecting or inferring the occurrence of *D* with *B* and *C* in the absence of *A*? (Hesse 1970)

Two other philosophical positions frequently cited when it comes to animal extrapolation in particular, LaFollette and Shanks (1995) and Steel (2008), phrase the problem in essentially the same terms as Hesse. Yet they solve it in opposing ways. First, LaFollette and Shanks state that for the inference to be valid,

> *there must be no causally relevant disanalogies between the model and the thing modeled* (LaFollette and Shanks 1995).

On the contrary, Steel concludes that a total absence of causally relevant disanalogies is not required for extrapolating (Steel 2008). Whatever the answer, analogy of mechanisms plays a central role in that sort of extrapolation, according to the received view.

A further claim is that mechanistic extrapolation consists in an incremental comparison of mechanisms. The potential difference under investigation between mechanisms in model and target should be limited to one

entity (or process) only. This might be called the 'increment condition'. Indeed, mechanisms are manipulated in experiments. An ideal intervention is supposed to act upon *one* variable of interest and isolate it from the rest of the mechanism, which in turn is supposed to go on working as it did before. To put all this briefly, we extrapolate causal claims of the sort '*X* causes *Y*', *X* and *Y* being variables. The causal claim itself is justified insofar as it conforms to the conditions of an "ideal intervention" *I*, that is,

> *I* eliminates all other influences on *X*;
> *I* does not change anything to related variables except *X*;
> *I* does not depend on any variable related to *X*. (Steel 2008)

Then the mechanisms in the model, i.e. the manipulated mechanisms, are *compared* to the mechanisms in the target, i.e. the mechanisms of interest. More precisely, enough must be explicitly similar in both mechanisms for a further part of it to be extrapolated. Steel has proposed a view that rapidly became accepted, that this is achieved through a mode of inference he calls "comparative process tracing" (Steel 2008). Process tracing consists in trying to fill in a schema of a causal pathway in the causal structure under study. This in turn is achieved through tracing causal pathways, backward or forward, from a known starting point. Now *comparative* process tracing consists in establishing parallel processes in both the model and the target, then drawing the conclusion that the next causal link observed in the model, but not yet in the target, is likely to be found there. Steel also relaxes the condition from identity to relative similarity between known causal pathways in the target and in the model:

> If significant differences between the model and the target are likely to be restricted to a relatively small number of stages of the mechanism, then comparisons at those stages may provide good grounds for the suitability of the model. (Steel 2008)

That said, evidence of a mechanism in humans does not always rely on an incremental comparison. It sometimes relies on mere generalization of observations made on a single species to all living organisms, without any step-by-step tracing. However, in the health science in particular, such generalization is generally used as background biological knowledge – for instance about the mechanisms of gene expression, methylation, etc. –, but they are generally not intended to provide evidence about diseases in humans.

The investigation of mechanisms justifies the prediction about humans. It does not involve that statistics are absent from preclinical studies. Most philosophers agree with the difference, emphasized by (Illari 2011), between "mechanistic evidence" and "evidence of mechanisms" on the one hand, and "statistical evidence" and "evidence of difference making" on the other hand. It is a difference between object and method. Mechanistic methods involve manipulation, tracing steps in processes, and they certainly are dominant at the preclinical level. On the other hand, statistical methods are prominent at the clinical level. But it is easy to show that in the experimental method, evidence of correlation is critical to evidence that a mechanism exists (Campaner 2011).

4.2.2 Evidence of a Mechanism Is the Main Point of Preclinical Studies

The focus on mechanistic extrapolation has consequences as to the role played by evidence of mechanisms in preclinical studies. A widely accepted view in the philosophy of medicine is that evidence for causal claims in the health science is of two sorts, evidence of a correlation and evidence of a mechanism. Evidence of a correlation is evidence that a cause C makes a difference to an effect E (generally: the presence of C increases the probability of E: see Salmon 2006). Evidence of a mechanism is evidence that there is a possible continuous chain of events, generally called a 'pathway', linking the occurrence of C to the occurrence of E.

Among philosophers, the main debate about evidence of causal claims in the health science is about what is called the Russo-Williamson Thesis (RWT), which states that evidence that C causes E requires *(1)* statistical evidence that C makes a difference to E, and *(2)* evidence of at least one mechanism or pathway from C to E (Russo and Williamson 2007, criticized in Howick 2011; Broadbent 2011; see also LaCaze 2011; Worrall 2010). Are these two conditions necessary, or can either one suffice to establish a causal claim? This debate parallels the debate in the health science itself, about the status of evidence of mechanisms generally obtained in the so-called 'basic science' as compared to evidence of correlation, generally obtained in either epidemiological studies or clinical trials.

Let us ask the RWT one question: where does evidence of mechanisms come from in the health science? Obviously, if implicitly, it comes from basic science, in which animal models play a major role. The RWT does not explicitly refer to animal experiments in the original paper, but further developments of this view do (Illari 2011; Clarke et al. 2014). Now let us ask the reverse question: what sort of evidence does basic science in general, and studies on animal models in particular, provide? Whereas clinical studies are "aimed at establishing difference making", preclinical studies are "clearly aimed at establishing knowledge of mechanisms" (Clarke et al. 2014). Strictly speaking, this does not discard the possibility that there is statistical extrapolation in preclinical studies, but it leaves it in the dark.

Now evidence of any sort can be put to particular uses, namely, explanation on the one hand, to which evidence of mechanisms is critical; prediction and control on the other hand, to which evidence of difference-making is critical:

> Evidence of difference making is required because causal claims are used for prediction and control, and one can only predict an effect on the basis of the cause, or control the effect by manipulating the cause, if the cause makes a difference to the effect. Evidence that there is an underlying mechanism is required because causal claims are used to explain, but in order to explain some phenomenon one needs to point to the (functioning of the) mechanism responsible for it; so invoking a cause as an explanation for an effect is only successful to the extent that the cause is a part of the mechanism responsible for the effect. (Russo and Williamson 2012)

Evidence of any mechanism between C and E is therefore necessary to the causal claim, *at the level of demand required by explanation*. Along with evidence of difference making, it is also sufficient. Some, as Broadbent, have agreed that it is necessary to *explain*, but not to *warrant* the causal claim (Broadbent 2011) – to which evidence of difference making can suffice.

Our claim is that preclinical extrapolation of a certain kind, i.e. statistical extrapolation, does not aim to investigate mechanisms, but only to establish evidence of difference-making. Its goal is not to explain, but to predict and control. Our correction to the RWT is therefore more a development than a criticism.

That said, the complete received view on animal extrapolation may be phrased as follows:

1. Most preclinical studies consist in the extrapolation of the existence of an incremental part of a given mechanism as observed in the model, to the mechanism in the target, thanks to ideal intervention.
2. Preclinical studies provide evidence that a mechanism M explains why C causes E in animals, thereby indirectly predicts, in a weaker way than evidence of difference-making in humans would, that C will cause E in humans as well.

We will now challenge this view by examining how preclinical studies may also establish evidence of difference-making and thus predict a little more reliably than by establishing mechanisms.

4.3 Challenging the Received View on Preclinical Evidence

A challenge to this received view is introduced here through a case study of animal models of a major class of antidepressant drugs, selective serotonin reuptake inhibitors (2.1). Even if the logic of discovery was mainly based on mechanistic reasoning indeed in this case, the logic of justification is rather based on statistical reasoning (2.2).

4.3.1 The Effects of Selective Serotonin Reuptake Inhibitors Have Not Been Established by Mechanistic Extrapolation: A Case Study

Serotonin reuptake inhibitors are considered the standard treatment of major depression. Specific animal models have become standard in preclinical testing of this class of molecules. In this section, we first sum up the history of these models and molecules, then highlight how extrapolation has been used in this literature.

A first part of the context and background knowledge is constituted by animal models of depression. Models of learned helplessness, which consist in conditioning helplessness behavior in animals and had been developed in the mid-sixties (Seligman and Maier 1967; Overmier and Seligman 1967), are generally considered to have started the practice of testing drugs for mood disorders on animal models. However, these models were soon criticized because they relied on the imposition of major stressors, like electric foot

shock, a situation which was not deemed relevant for depression. The first study on one of the currently best validated models for depression, the chronic mild stress model (CMS), which consists in submitting animals to mild stressors for a long duration, was published in 1981 (Katz et al. 1981) while the first test, the forced swimming test (FST), which consists in measuring how long a rodent tries to escape water, was published in 1977 (Porsolt et al. 1977). Another animal model of depression, the olfactory bulbectomy in rats, which consists in removing a part of their brains important in the perception of smell, was first described in 1977 (van Riezen et al. 1977; Cairncross et al. 1977). CMS and FST are currently the most frequently used models in this area.

A second part of the context and background knowledge is constituted by available drugs. During the same years, a new class of antidepressants was developed: selective serotonin reuptake inhibitors (SSRIs). SSRIs have been the most popular treatments of major depression ever since. They stimulate neurotransmission in serotonergic circuits of the brain by increasing the availability of synaptic serotonin. This is achieved by downregulating the reuptake of serotonin in the post-synaptic cleft. When the first reports on these drugs were published in the mid-seventies, major depression was treated with inhibitors of the monoamine oxidase enzyme, which degrades monoamines such as serotonin or noradrenaline, or with molecules that blocked their uptake in the presynaptic neuron (tricyclics).

The first publication regarding the effects of SSRIs was a report by Wong and colleagues in 1974 (Wong et al. 1974, 1975a, b) showing that fluoxetine, a molecule of this class, had a highly potent effect on serotonin reuptake in nerve endings of drug-treated rats (*ex vivo*). This effect was confirmed using *in vivo* preparations (Fuller et al. 1974). In 1976, clinical studies started (for a review see Wong et al. 2005) and in 1978 Lemberger and colleagues published a paper in *Science* showing that fluoxetine is able to act as a SSRI in healthy human subjects as well, and they concluded that the drug could be effective in treating "mental diseases," but without providing any evidence of such effects, in animals or in humans (Lemberger et al. 1978). One year later, in 1979, Meltzer and colleagues administered the drug to 3 patients with severe depression, reported the effects on 28 others, and tried to explain the paradoxical effects on dopaminergic activity in only one of them (Meltzer et al. 1979). Finally, the drug appeared on the European market in 1986 and received Food and Drug Administration (FDA) approval one year later.

In the same period, other SSRIs have been tested, such as citalopram. The effects of this molecule were first described in Hyttel (1977). Three years later, the results of the first Phase II study were published (Gottlieb et al. 1980), showing an antidepressant efficacy of citalopram. A similar story repeats with paroxetine as, in a first paper, authors insist on the serotonergic effects of paroxetine (Lassen 1978); one year later a tolerance study is published in human subjects and in 1982 the results of the Phase II Open trial are published, again with no reference to preliminary data using animal models. In 1983, Koe and colleagues showed that sertraline is a very potent SSRI modifying brain serotonin content, and that this is sometimes associated to antidepressant effects in standard animal tests such as the FST (Koe et al. 1983). Moreover, they showed that it acts also on noradrenergic variables. They concluded that the drugs could have

"potential antidepressant activity". This was indeed shown in a first clinical trial published in 1988 (Reimherr et al. 1988; Burrows et al. 1988).

A third and last part of the context and background knowledge is constituted by more recent treatments. Few new antidepressants have been introduced in clinical practice in the last years. Vilazodone is one of them. It was approved by the FDA in 2011. The first report is an unavailable communication at a meeting, suggesting antidepressant-like effects. One year later, authors report that the compound acts as a SSRI with selective presynaptic 5-HT1A receptor agonistic properties (Bartoszyk et al. 1997). As 5-HT1A ligands are known to reduce anxiety behavior, such effect is investigated in this article and in another one (Treit et al. 2001). Both show that the drug reduces anxiety behavior in some tests but not in others. One year later, a study reported antidepressant-like effects of this compound on animals (Page et al. 2002). A first randomized, double-blind, placebo-controlled trial was published in 2009 (Rickels et al. 2009), indicating that the treatment induces relief after 1 week. The drug received FDA approval in January 2011.

A brief chronology of the essential facts is presented in Fig. 4.1. Let us focus first on Hyttel's study on citalopram. The main result is to establish that this then new compound has a stronger and more selective effect on the level of serotonin available in the synaptic cleft in animal models, through inhibition of the reuptake. It draws from pre-existing knowledge of general mechanisms of monoamine, serotonin, uptake, etc., but predates standard animal models and tests of depression. It is a typical investigation of one part of the mechanism of action of a potential new drug, in comparison with the mechanism of action of other, known drugs (namely, tricyclics). However, note that there is not direct extrapolation to humans here. Hyttel's conclusion is that if the effect of citalopram was confirmed in humans with depression, it might help investigate the etiology of the disease. For this reason, it is not a typical example of "comparative process tracing", but a simpler investigation of possible mechanisms with a remote analogy with humans in mind.

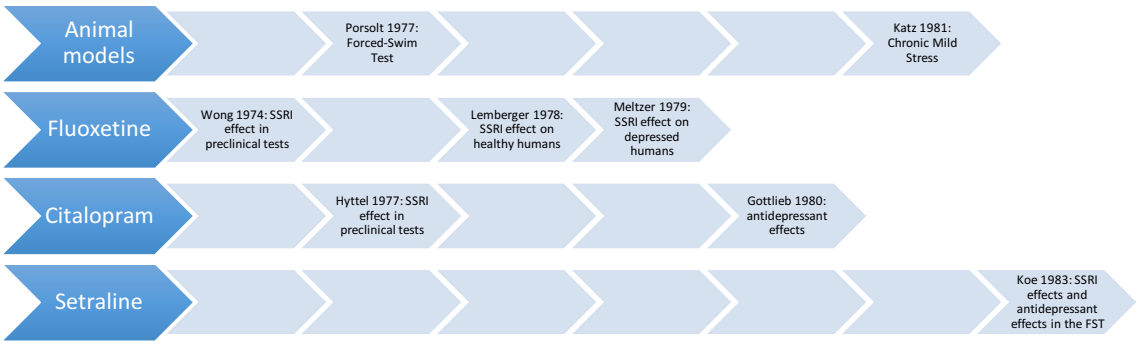

Fig. 4.1 A chronology of essential facts in preclinical studies of SSRIs

Let us now turn to a later study. Hyttel's paper could not rely on any standard model for depression, there being none at the time. It could only rely on the consistency of the result in various experiments. The argument is very different in Koe's paper on sertraline in 1983:

1. Depressed patients have low levels of serotonin and compounds with a strong serotonin component are effective in treating depressed patients;
2. Sertraline selectively inhibits serotonin reuptake;
3. Sertraline has a significant effect on depressive behavior in the standard FST;

---

4. Sertraline will have a significant effect on depressive behavior in humans.

Koe's paper does not meet the conditions of the received view. First, its point is not to establish a causal pathway between selective inhibition of serotonin reuptake and relief of depressive symptoms. As we have evidence that there probably is one, it is worth investigating the effects of a new molecule that we class in the same category. The goal is not to explain *how* sertraline would produce such an effect. Nor is such an explanation required for prediction. That we know that similar drugs have an antidepressant effect on standard models, is sufficient to predict that this drug will have the same effect. This is mere correlation. As in Hyttel's paper, this seems to presuppose that the mechanism is similar. However, the extrapolation itself is statistical: because the FST should be trusted in general, and not because the effect of sertraline in the mechanism involved in this test is well known or comparatively traced in both humans and animals, sertraline should be effective on human depression. For this reason, it is difficult to apply the standard view here. Indeed, this sort of animal extrapolation is of a different nature than that emphasized by the standard view. We call it 'statistical', and claim that it is actually often resorted to, not only in the case of 'me-too' drugs, as one could expect, but also when innovative drugs are under investigation. Let us take a third example. Around 2000, basic research suggested that methylation of genes providing a protective effect against stress could play a role in vulnerability to depression. It became an interesting idea to investigate whether molecules with a demethylation effect could have a therapeutic effect. A standard strain of a rodent known to be responsive to standard childhood stress protocol was chosen to test this hypothesis, because if there were any behavioral alteration, it would be both obvious and predictive of what would happen in human subjects. Some mechanisms are indeed known in this standard model, but precisely not those hypothesized on the basis of basic research: behavioral effects would be strong piece of evidence that they exist. The basis of evidence about humans in this case is the well-established correlation between animal and human behaviors, in other terms, the existence of a standard model. Note however that not all innovative drugs can be studied in standard models. When basic research had suggested for instance that a family of molecules called neuropeptides intervened in serotonergic neurotransmission as a catalyst in case of very intense stimulation, these effects could not be investigated on standard animal models. Indeed, some of the pharmacology was different and the distribution of the receptors of these neuropeptides was not homologous between humans and the standard rodent species used in this kind of research. Researchers had to find different, non-standard species, with entirely different behavioral patterns, responsive to different pathogenic protocols (Salomé et al. 2006; Griebel and Beeské 2012).
This brief overview of the history of antidepressant since the 1970s emphasizes how problematic the application of the received view to typical cases of preclinical extrapolation may be. However, we also found

that statistical extrapolation, although different from mechanistic extrapolation according to the received view, seems to also rely on the presupposition that there is a shared mechanism. It seems impossible to conclude that drug D will have an effect on humans H on the basis that it has on animal A, if the mechanism linking D to H is not supposed to be the same as the one linking D to A, even if we do not know this mechanism. In other terms, it seems that statistical extrapolation in preclinical studies is not independent from mechanistic extrapolation. The rest of the paper investigates this question and clarifies the nature of statistical extrapolation.

4.3.2 The Effects of Selective Serotonin Reuptake Inhibitors Were Established by Statistical Extrapolation

These papers, and specially Koe's, show instances of the sort of animal extrapolation we dub 'statistical'. In several ways, they look like they involve mechanistic reasoning, at least partially. We do not deny it. But it is important to specify what role mechanistic reasoning plays in such studies. First, for any study of animal models to be potentially published in a scientific journal, either a mechanistic or a statistical rationale is needed. By 'rationale', what is meant is a piece of evidence of either a working mechanism, about which we want to know whether it underlies a set of cases, or of a correlation we know exists and want to use as a signal of the effect of an intervention on a given mechanism. In Koe's study, there is evidence that sertraline binds with serotonin transporters and prevents them from reuptake by the presynaptic neuron, leading to an increase of the quantity of serotonin available in the synapse. But such evidence establishes the relevance of the hypothesis of a study, not its result. To put it in a nutshell, mechanistic knowledge here does not play a direct role in the logic of *justification* of a hypothesis, but only in the logic of *discovery* of this hypothesis. What justifies the result, is the trust researchers place in the model based on the basis of its passed successes (statistical evidence). Yet, the accumulation of successes, i.e. the very fact that a model becomes standard, means that it supposedly involves the same mechanism. It is indeed correct that the model may not be relevant anymore when a significant difference in the mechanism is involved, as was the case for neuropeptides. But results from this model are deemed conclusive even when a different mechanism is known to be involved, as was the case for demethylation drugs referred to above. Moreover, researchers sometimes incorrectly focus on one mechanistic effect of a drug and misinterpret it as the explanation for its efficacy, possibly without any effect on the reliability of the model.

Second, it seems that no direct statistical extrapolation is possible from a given signal in an animal model to an effect on a human disease, if there is no assumption that the same mechanism links treatment to signal in the animal and treatment to effect on the human disease. We found that such assumptions were likely to be held in the studies we analyzed in the previous section. Thus, it seems that statistical extrapolation is always mixed with at least a bit of mechanistic extrapolation, be it a hypothetical mechanistic extrapolation. We suggest however that a distinction has to be made between the assumption that a specific mechanism links cause to effect in both model and target, and the assumption that cause and effect are mechanistically linked in both model and target. Suppose two scientists agree that a study reliably predicts the effects *E* of drug *D* in humans, but hypothesize different mechanisms *M* and *M'* to explain why. Or suppose one scientist

who does not know whether *D* elicits *E* through *M*, *M'*, or *M"*, or sometimes *M*, sometimes *M'*, sometimes *M"*, or always a mix of *M, M'*, *M"*, etc. It would matter if the question was to produce an explanation. It does not if the question is to produce a prediction. A prediction assumes that there are mechanisms in the world that allow it to be reliable, but only an explanation needs that one is picked up, even if it still is unspecified. The only mechanistic assumption in a statistical extrapolation from animal models, is therefore that interventions and effects in both models and targets are mechanistically linked, even though a specific mechanism is very often provided. This hardly makes it more than a general *credo* about causality in the world.

The rationale of statistical extrapolation would therefore not be such a general assumption. It is rather that the model, in Koe's paper, the FST, is considered a reliable, predictive test of such drugs. In other terms, there is a correlation between the effects of such drugs in such animal models, and the effects of the same drugs on their human targets.

We propose to define such *statistical extrapolation* as a prediction based on the past successes of the model to predict what will happen in the target. Koe's paper, we claim, is an example of an extrapolation of this latter sort. Indeed, sertraline should have a significant effect on depressive behaviors in humans because it has this effect in the FST and because the FST is a reliably predictive test of the effects of monoaminergic compounds on human behavior.

As compared to mechanistic extrapolation, statistical extrapolation is an alternative way to escape what Steel (2008) calls the 'extrapolator's circle': as we do not know everything about model and target systems, there may still be significant differences that impair the former's predictive validity as regards effects in the latter. On the other hand, in order to know that the model system is identical to the target system in every relevant aspect, we would need to know everything about them, and experiments would be unnecessary. One way out of the circle, he justly says, is "comparative process tracing". Another way out of the circle, we claim, relies in establishing the predictive validity of a model system, the precise mechanisms of which may remain unknown. Indeed, a model is predictive insofar as a mechanism is shared, but one does not need to know what the mechanism is to know whether the model is predictive. In other terms, statistical extrapolation is incremental in a different sense that mechanistic extrapolation is: it consists in induction to the next case, based on former cases, and escape the extrapolator's circle because it is not causal knowledge that is searched, but only prediction.

The rule of statistical extrapolation may be summed up like this: the more a model of a disease has successfully predicted the effects of treatments of a certain class on human subjects, the more it can be considered predictive of the action of new molecules of the same class. For lack of space, a formalized expression of this sort of inductive inference is not provided here. Suffice it to say that the model is primarily proposed on the basis of its resemblance to known phenomena involved in the human disease, from etiology to symptoms through pathophysiology. It is then put to test. If a drug does produce on humans the effects

predicted by the animal model, then the latter is corroborated and finally becomes standard. Statistical extrapolation may be summarized thus:

1. Animal model of a disease M has successfully predicted the effects of drugs D1, . . . Dn of a similar class C on human targets T

2. Drug Dn+1 belongs to C

3. Drug Dn+1 has effect E on M

4. Drug Dn+1 will have effect E on T

Note here that 'class' does not necessarily mean a mechanistically homogeneous family of molecules (e.g. SSRIs). It may be more or less extensive – for instance, it could be the class of 'antidepressant medication' as well as the class of 'antidepressant medication except SSRIs', etc.

4.4 Objections to the Existence of Statistical Extrapolation

In this last section, we raise, and respond to, 4 objections to the view that statistical extrapolation exists and is distinct from mechanistic extrapolation. The first objection is that standardization, i.e. the frequent success of animal models in predicting effects in humans, is not evidence in and by itself. To this objection we reply that at least in the sense of 'used to represent the proper construct', standardization provides evidence in and by itself. The second objection is that statistical extrapolation is rare: on the contrary, we argue that it is the most frequently resorted to, at least, in preclinical studies of mental disorders. The third objection is that statistical extrapolation must in principle rely on mechanistic knowledge, an argument that ignores the possibility of statistical clustering. To the fourth objection, namely, that statistical extrapolation is typical of applied research and eventually relies on basic research, we respond that it does rely on basic research in the logic of discovery but not in the logic of justification, that they are strategically interdependent, but that predictive validity and mechanistic validity are theoretically independent: it is one thing to assess whether a drug is effective and another to discover why it is.

4.4.1 Standardization Is Not Evidence

'Standard model' is a phrased often resorted to in the scientific literature. Philosophers have also discussed the implications of standardized models (Meunier 2012). Note that what is at stake, at least in psychiatry, is not so much a model organism (Ankeny and Leonelli 2011) than a standard situation for a given organism, say, mice under chronic mild stress. At least five different meanings of this term must be mentioned:

1) 'arabidopsis thaliana is a standard model' may mean that it is 'frequently resorted to', independently from what is to be established;

2) 'Skinner box #1 at Columbus is a standard model' means that the experimental apparatus is 'set up with a normalized protocol', i.e., one can trust the reproducibility of the procedure;

3) 'the standard protocol resorted to in learned helplessness' means that it is an 'agreed on protocol', i.e. everybody agrees on what the procedure *should* consist in, although confounding factors in the *actual* setup in one investigation may have influenced the result (e.g. the perfume worn by an investigator may influence the outcome of the experiment);

4) 'UCMS is a standard model' means that it is 'robust', i.e., the model always gives the same results under the same circumstances (which is not strictly implied by the reproducibility of the procedure per se);

5) 'latent inhibition is a standard model of attention deficit in schizophrenia' means that this experimental process, consisting in teaching animals not to take a stimulus into account, has been 'used to represent the proper construct', i.e., the model allows to establish something about the construct this model is supposed to represent.

We are especially focused on the fifth meaning of the term standard here, notwithstanding the importance of the first four sense to the success of statistical extrapolation. Indeed, that a model is 'used to represent the proper construct' cannot mean anything else than it 'has frequently been successful in predicting causal claims about a certain construct'. It could not mean anything along the lines of 'it involves the right mechanism', because the fact that a mechanism is the right one, that is, instantiates all instances of a disease and only them, cannot be proved mechanistically and has to be established statistically.

Many scientific papers are refused because they justify their use of a given model by its being standard in any of the other four meanings, but fail to convince that it actually represents the proper construct. Yet the results of some articles are accepted as valid although they do not use standard models in this sense. A striking example is a study by Cryan and colleagues (Cryan et al. 2005). Since its introduction in 1977, the FST did not always detect the effects of SSRIs on rats. Cryan introduced two changes in the test, a slightly deeper tank so that rats cannot reach the ground, and a camera instead of the observer's presence, so that rats cannot see humans. This non-standard test detected effects of SSRIs, and has become the standard FST ever since, precisely because it then became successful in predicting causal claims about effects of SSRIs.

Now, the objection goes, the fact that a model has been standardized in the sense of 'validated as the proper model of a given construct', is not in itself evidence of any other claim that those already proven.

We claim that it is. First, statistical extrapolation is not just an instance of induction in the classical sense, in the sense that it does not simply rely on the reproduction of the same results in the same circumstances. It is not an instance of 'expect more of the same': remember that different drugs ($D_1 \ldots D_n$) of the same class C must be tested for the model to become 'standard' in that sense. Second, what matters to a standard model of a construct, is not the number of articles having established the same claim about the same dimension of

the construct, it is on the contrary the number of different dimensions of the construct the model has been shown to instantiate. Indeed, scientists make a difference between a test or a bioassay and a model. Whereas tests and bioassays are valid representation of one dimension of a construct, models are valid representations of different dimensions of a construct. In this field, 'model' is frequently used in cases where animals have been subjected to manipulations inducing a pathological state (although, strictly speaking, many symptoms, such as deficit attention in schizophrenia or anhedonia in depression, are also constructs rather than dimensions). The tail suspension test is not a standard model of depression properly speaking, but rather a standard *test* for (one dimension of) depression. On the other hand, it could be considered a standard model of resignation, as a dimension of depression, although the term is generally not used because resignation is not a disease. Hence, the more dimensions of a construct a given model instantiates, the more predictive of the target's behavior it can be considered. This includes both known and unknown aspects of the target's behavior. Statistical extrapolation relies on the multidimensional resemblance of a model to its target to make new predictive claims that have not formally been proven. This does not mean that the mechanistic investigation of the model as well as that of the disease cannot play a key role here in guiding researchers in the selection of the relevant dimensions of the model.

4.4.2 Statistical Extrapolation Is Rare

A second objection is that if statistical extrapolation exists, it is rarely resorted to, and that standard reasoning in preclinical studies is mainly mechanistic. Here we exclusively focus on the literature in psychiatry. Although we have reasons to think that statistical extrapolation is also frequent elsewhere, some may challenge our views by claiming that research in neuroscience is a particular case.

The pharmacologist Guy Griebel described the major trends over the past 50 years in the search for new anxiolytic drugs (Griebel and Holmes 2013), carefully analyzing and summarizing 10 000 preclinical experiments, on approximately 1500 novel drugs.

It can be clearly seen that preclinical research in this field has developed progressively during the 1980s, peaking in the 1990s, with a number of published experiments/year ranging between 400 and 800. These papers always mention the neurotransmission system that is targeted: half of them focused on the serotoninergic system, a third of them focused on different kind of neuropeptides, and, more recently, a trend toward an increase in studies investigating the effects of ligands of the glutamatergic or of the endocannabinoid system appeared. These studies thus described quite precisely the molecular target that was investigated (a given neurotransmission system, or more precisely a subtype of a receptor) as data exist on the molecular substrate to which these compounds are binding. In some cases, these studies eventually used genetically-engineered mice to show that in the absence of the molecular target, the pharmacological effects were absent: in this case, they really provide data on the primary target of the putative drug. However, they less frequently describe the neural circuit that was targeted. For example, on the 10 000 experiments that are listed in this paper, only 2740 consisted in the local administration of the drug in a given brain area belonging to the fear circuit: the other studies did not target a specific sub-system.

Thus, these investigations focus on the primary molecular target, not on more sophisticated underlying mechanisms.

Indeed, a difference is to be made between the definition of a class of molecules, and the mechanism involved in a disease. The former can, and actually is, specified without the latter in most studies. This is strong evidence that in fact, evidence in preclinical studies is, in psychiatry at least, mostly based on statistical extrapolation.

Note that the main shortcoming of preclinical studies may be blamed on statistical extrapolation based on poor knowledge of the mechanisms involved. As a result, the clinical outcome of most studies cited above was really poor: for example, among the serotoninergic compounds tested, only one made it to go to clinical use (the 5-HT1A agonist tandospirone, and only in Japan and China). In fact, in their conclusions, the authors recommend to shift from a target validation strategy, to an advanced strategy, focusing on the underlying mechanisms: "*Anxiolytic drug discovery (...) will be greatly facilitated by concerted efforts to elucidate the underlying neurobiology of anxiety*". They precisely make this recommendation because this strategy was poorly used up to recently, which indicates that most of the studies only rely on target-related drug discovery strategy.

4.4.3 Statistical Extrapolation Must Rely on Mechanistic Thinking

Models turn into standard models when their validity becomes public knowledge. Thus, one may argue that although their validity might appear to some as statistically proven, their widespread use actually relies upon prior work, based upon mechanistic thinking.

The function of animal models in scientific reasoning, and in particular when assessing the effect of a new molecule, can be summarized as follows. A drug D_n is known to have an effect upon both animal model A and human target H. Mechanistic reasoning allows to establish that this can be explained by the same mechanisms being at work on both sides. A is thus said to be an animal model of the action of D_n upon H. When a new molecule D_{n+1} is under scrutiny, mechanistic reasoning again suggests that its possible effects upon A are to be explained by the same mechanisms that explain the impact of D_n on M. When D_{n+1} indeed displays an impact upon A, the very fact that A is an animal model of the action of D_n upon H allows to predict, by transitivity, that D_{n+1} should also have an impact upon H.

The very existence of the animal model allows to dispense with that part of the evidence which consists in tracing the (mechanistic) link between the action of D_{n+1} upon A and the action of D_{n+1} upon H. Yet, whatever the number of molecules whose effects have been assessed, everything started with, and still relies on, mechanistic thinking.

We do not deny the accuracy of this description of animal model use in scientific reasoning, but we claim that the adjective 'mechanistic' is dispensable at every stage. It is possible to establish statistically that whenever a molecule D_n triggers an effect upon A, it does trigger that same effect upon H, without having

the slightest idea of what could explain that effect, on either side (A or H). It is also possible to group molecules according to their effects, without having any clue of what can explain them. And the use of transitivity is not restricted to chain mechanistically justified logical propositions, but is legitimate piece of any rational reasoning.

"Statistical clustering" by which objects are grouped by properties that make them similar is nowadays a common technique of statistical ("model free") data analysis. Correlations are just an instance of statistical measures, which do have a recognized scientific status. Moreover, relying upon the results of others to extend the existing knowledge is at the basis of modern science. Although mechanics certainly played an important role in the development of science, the latter need not be based upon mechanistic reasoning, not even partially nor initially.

4.4.4 Statistical Extrapolation Is Applied Research in a Specific Case But Eventually Relies on Basic, Mechanistic Research on General Processes

It is one thing to show that in particular cases, no mechanistic knowledge is used, but it is quite another to claim that no mechanistic knowledge in general is involved in the justification of a causal claim. One may argue that even if no mechanistic extrapolation has been established in the first place for a particular study, general mechanistic knowledge, i.e. background knowledge gathered from basic research, is necessary to statistical extrapolation as an instance of applied research.

Applied research in preclinical studies focuses on *predicting* what the effects of a drug on a human disease will be, whereas basic research focuses on *describing* what the mechanisms of a disease are. The prediction of how effective a drug will be on humans is as reliable as the effects of similar drugs tested on animal models are correlated to the effects of these latter drugs on humans. This line of reasoning is generally adopted by pharmaceutical companies. Prediction is blind to *how* the disease is caused, maintained, cured, and pays attention only to *whether* effects of clinical interest will happen. In other words, it is a direct extrapolation from a sample of animals (humans not included) to a broader set of animals (including humans). It is atheoretical – it does not presuppose any theoretical model of how this all works.

Fundamental research on the other hand is more interested in the investigation of mechanisms. It seeks after an exact theoretical model of what happens. The theoretical model represents a part of what happens in the human disease, supposedly, the essential part of it. For that reason, it cannot be said that what happens in one model organism (animal) represents what happens in the human disease, but it only assesses one causal relation of possible interest to the understanding of the human disease.

We agree with this general claim that basic research is necessary to applied research, with qualifications. First, it is difficult indeed to make a conceptually clear distinction between basic research and applied research in the field of preclinical studies. The clear line between academic research and R&D in pharmaceutical companies is relevant, but insufficient. From a very general perspective, there is no essential difference between basic and applied research studies, for they both come down to exactly the

same problem of assessing a causal claim: does *C* (etiological factor, treatment) have an effect *E* (elicits, relieves) on disease *D*?

Second, if a distinction is to be made, it must rely precisely on the theoretical independence of explanation and prediction. Applied research resorts to models with high *predictive validity*, that is, "the resemblance of the apparent impact of the etiological factors and of the treatment on the observable effects", two of us wrote in a previous review, whereas basic research resorts to models with high *mechanistic validity*, that is, "the similarity of the mechanism we suppose or know is working in the animal disease to the mechanism that is or is presumed to be working in the human disease." (Belzung and Lemoine 2011). We emphasized the difference thus:

The fact that a therapeutic agent has a dramatic impact on a biological system does not imply that it dramatically reduces the symptoms. The reverse is also true of the dramatic action of an agent or a factor on the symptoms (or on the biological markers). The predictive validity of a model is assessed without looking into the mechanism which is really at work in the animal: generally it is assessed from a macro-observational point of view or through peripheral biological measurements (biomarkers). (Belzung and Lemoine 2011)

Again, there has been many mistaken claims about *why* many treatments were effective on given diseases, assorted with justified claims *that* these treatments were effective on these diseases. As *(1)* a treatment was effective on a disease and as *(2)* a mechanistic effect of the treatment was observed (say the increase of available serotonin in the synapse), the conclusion was mistakenly drawn that *(3)* the treatment was effective on the disease because of this mechanistic effect. In spite of that error, the model was predictively valid.

Third, what does rely on general mechanistic knowledge, is the formulation of a likely hypothesis (logic of discovery), not its predictive validity (logic of justification), which can only be assessed as a measure of predictive success in past attempts. It is a good strategy to rely on mechanistic knowledge to choose a model to put to test. Note that the reverse is also true: as a predictively valid model is likely to be mechanistically valid as well, predictive validity is a pragmatic measure of mechanistic validity. In other terms, it is a good strategy, and one frequently resorted to as well, to investigate the mechanisms of a given behavior in models with high predictive value of what happens in humans.

4.5 Conclusion

The upshot of this chapter is to question, at least in psychiatry, the prominence of the received view that extrapolation from animal models in medical research aims to establish mechanisms of human diseases and does so by investigating them in a step by step comparison with animal models of the same disease. We called this form of extrapolation 'mechanistic extrapolation'. We challenged its alleged prominence by describing another form of extrapolation from animal models, we called 'statistical extrapolation'. Statistical extrapolation relies on the predictive validity of animal models, a form of validity that can be established

without any knowledge of mechanisms, nor of any resemblance of mechanisms. It relies on past successes of the model to predict the effect of a class of drugs on their human target.

Two caveats are important here. First, we think that our views should not be opposed to the RWT, but that the latter should be revised only on a minor point, namely, its implicit acceptance of the idea that preclinical studies are dedicated to establishing mechanisms. We do not assume that predictive success of preclinical studies, i.e. statistical extrapolation, is a sufficient condition to establish a causal claim, and thereby do not question the core of the RWT.

The second caveat regards mechanistic hypotheses. A rapid reading of the chapter may convey the impression that we think preclinical research may dispense with hypotheses on what the mechanisms of a disease may be. The fact that mechanistic extrapolation is not ubiquitous, and the fact that mechanisms mostly play a heuristic role, should not lead to the conclusion that preclinical studies, even in applied research, is not essential. The fact that mechanistic evidence can be dispensed with does not mean that it is a good thing to do so when it is possible to do otherwise. Whereas being *blind* to mechanisms would involve purely random testing of hypotheses, being *blinded* only involves selective ignorance of vast areas of possibly relevant mechanisms. Even drug screening is selective. Statistical extrapolation is therefore blinded, but not blind to mechanisms.

Alongside mechanistic extrapolation, which consists in the incremental establishment of a mechanism by comparing its instances in a model and in a target, comes statistical extrapolation, which consists in the incremental establishment that a model supports prediction of the target's behavior, whatever the underlying mechanisms maybe. Statistical extrapolation is frequent in the field of animal models, at least in psychiatry, and possibly elsewhere. It does not essentially depend on mechanistic evidence, investigation or even hypothesis. For this reason, it deserves more philosophical attention.

References

Ankeny, R. A., & Leonelli, S. (2011). What's so special about model organisms? *Studies in History and Philosophy of Science Part A, 42*(2), 313–323.

Bartoszyk, G. D., Hegenbart, R., & Ziegler, H. (1997). EMD 68843, a serotonin reuptake inhibitor with selective presynaptic 5-HT1A receptor agonistic properties. *European Journal of Pharmacology, 322*(2–3), 147–153.

Belzung, C., & Lemoine, M. (2011). Criteria of validity for animal models of psychiatric disorders: focus on anxiety disorders and depression. *Biology of Mood & Anxiety Disorders, 1*(1), 9.

Belzung, C., Willner, P., & Philippot, P. (2015). Depression: From psychopathology to pathophysiology. *Current Opinion in Neurobiology, 30*, 24–30.

Broadbent, A. 2011. Inferring causation in epidemiology: Mechanisms, black boxes, and contrasts. In P. M. Illari, F. Russo, & J. Williamson, éd. Causality in the sciences. Oxford Oxford University Press, p. 45–69.

Burrows, G. D., et al. (1988). Clinical effects of serotonin reuptake inhibitors in the treatment of depressive illness. *The Journal of Clinical Psychiatry, 49*(Suppl), 18–22.

Cairncross, K. D., et al. (1977). The olfactory bulbectomized rat: A simple model for detecting drugs with antidepressant potential [proceedings]. *British Journal of Pharmacology, 61*(3), 497. Campaner, R. (2011). Understanding Mechanisms in the Health Sciences. *Theoretical Medicine and Bioethics, 32*(1), 5–17.

Clarke, B., et al. (2014). Mechanisms and the evidence hierarchy. *Topoi, 33*(2), 339–360.

Cryan, J. F., Valentino, R. J., & Lucki, I. (2005). Assessing substrates underlying the behavioral effects of antidepressants using the modified rat forced swimming test. *Neuroscience and Biobehavioral Reviews, 29*(4-5), 547–569.

Fuller, R. W., Perry, K. W., & Molloy, B. B. (1974). Effect of an uptake inhibitor on serotonin metabolism in rat brain: Studies with 3-(p-trifluoromethylphenoxy)-N-methyl-3- phenylpropylamine (Lilly 110140). *Life Sciences, 15*(6), 1161–1171.

Gottlieb, P., Wandall, T., & Overø, K. F. (1980). Initial, clinical trial of a new, specific 5-HT reuptake inhibitor, citalopram (Lu 10-171). *Acta Psychiatrica Scandinavica, 62*(3), 236–244.

Griebel, G., & Beeské, S. (2012). Is there still a future for neurokinin 3 receptor antagonists as potential drugs for the treatment of psychiatric diseases? *Pharmacology & Therapeutics, 133*(1), 116–123.

Griebel, G., & Holmes, A. (2013). 50 years of hurdles and hope in anxiolytic drug discovery. *Nature Reviews Drug Discovery, 12*(9), 667–687.

Hesse, M. B. (1970). *Models and analogies in science* (New edition). University of Notre Dame Press.

Howick, J. (2011). *The philosophy of evidence-based medicine*. Wiley-Blackwell, Bmj Books.

Hyttel, J. (1977). Neurochemical characterization of a new potent and selective serotonin uptake inhibitor: Lu 10-171. *Psychopharmacology, 51*(3), 225–233.

Illari, P. M. (2011). Mechanistic evidence: Disambiguating the Russo-Williamson thesis. *International Studies in the Philosophy of Science, 25*(2), 139–157.

Katz, R. J., Roth, K. A., & Carroll, B. J. (1981). Acute and chronic stress effects on open field activity in the rat: Implications for a model of depression. *Neuroscience and Biobehavioral Reviews, 5*(2), 247–251.

Koe, B. K., et al. (1983). Sertraline, 1S,4S-N-methyl-4-(3,4-dichlorophenyl)-1,2,3,4-tetrahydro-1-naphthylamine, a new uptake inhibitor with selectivity for serotonin. *The Journal of Pharmacology and Experimental Therapeutics, 226*(3), 686–700.

LaCaze, A. (2011). The Role of Basic Science in Evidence-Based Medicine. *Biology and Philosophy, 26*(1), 81–98.

LaFollette, H., & Shanks, N. (1995). Two models of models in biomedical research. *Philosophical Quarterly, 45*(179), 141–160.

Lassen, J. B. (1978). Influence of the new 5-HT-uptake inhibitor paroxetine on hypermotility in rats produced by p-chloroamphetamine (PCA) and 4,alpha-dimethyl-7-tyramine (H 77/77). *Psychopharmacology, 57*(2), 151–153.

Lemberger, L., et al. (1978). Pharmacologic effects in man of a specific serotonin-reuptake inhibitor. *Science (New York, N.Y.), 199*(4327), 436–437.

Machamer, P. K., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science, 67*(1), 1–25.

Meltzer, H. Y., et al. (1979). Extrapyramidal side effects and increased serum prolactin following fluoxetine, a new antidepressant. *Journal of Neural Transmission, 45*(2), 165–175.

Meunier, R. (2012). Stages in the development of a model organism as a platform for mechanistic models in developmental biology: Zebrafish, 1970–2000. *Studies in History and Philosophy of Science Part C, 43*(2), 522–531.

Overmier, J. B., & Seligman, M. E. (1967). Effects of inescapable shock upon subsequent escape and avoidance responding. *Journal of Comparative and Physiological Psychology, 63*(1), 28– 33.

Page, M. E., et al. (2002). Behavioral and neurochemical effects of 5-(4-[4-(5-Cyano-3-indolyl)- butyl-butyl]-1-piperazinyl)-benzofuran-2-carboxamide (EMD 68843): A combined selective inhibitor of serotonin reuptake and 5-hydroxytryptamine(1A) receptor partial agonist. *The Journal of Pharmacology and Experimental Therapeutics, 302*(3), 1220–1227.

Porsolt, R. D., Le Pichon, M., & Jalfre, M. (1977). Depression: A new animal model sensitive to antidepressant treatments. *Nature, 266*(5604), 730–732.

Reimherr, F. W., et al. (1988). Sertraline, a selective inhibitor of serotonin uptake, for the treatment of outpatients with major depressive disorder. *Psychopharmacology Bulletin, 24*(1), 200–205.

Rickels, K., et al. (2009). Evidence for efficacy and tolerability of vilazodone in the treatment of major depressive disorder: A randomized, double-blind, placebo-controlled trial. *The Journal of Clinical Psychiatry, 70*(3), 326–333.

Russo, F., & Williamson, J. (2007). Interpreting causality in the health sciences. *International Studies in the Philosophy of Science, 21*(2), 157–170.

Russo, F., & Williamson, J. (2012). EnviroGenomarkers: The interplay between mechanisms and difference making in establishing causal claims. *Medicine Studies, 3*(4), 249–262.

Salmon, W. C. (2006). *Four decades of scientific explanation* (Édition: 1). University of Pittsburgh Press.

Salomé, N., et al. (2006). Selective blockade of NK2 or NK3 receptors produces anxiolytic- and antidepressant-like effects in gerbils. *Pharmacology, Biochemistry, and Behavior, 83*(4), 533– 539.

Seligman, M. E., & Maier, S. F. (1967). Failure to escape traumatic shock. *Journal of Experimental Psychology, 74*(1), 1–9.

Steel, D. (2008). *Across the boundaries: Extrapolation in biology and social science*. Oxford: Oxford University Press.

Treit, D., et al. (2001). Systemic EMD 68843 injections reduce anxiety in the shock-probe, but not the plus-maze test. *European Journal of Pharmacology, 414*(2–3), 245–248.

van Riezen, H., Schnieden, H., & Wren, A. F. (1977). Olfactory bulb ablation in the rat: Behavioural changes and their reversal by antidepressant drugs. *British Journal of Pharmacology, 60*(4), 521–528.

Willner, P., Scheel-Krüger, J., & Belzung, C. (2013). The neurobiology of depression and antidepressant action. *Neuroscience and Biobehavioral Reviews, 37*(10 Pt 1), 2331–2371.

Wong, D. T., et al. (1974). A selective inhibitor of serotonin uptake: Lilly 110140, 3-(p-trifluoromethylphenoxy)-N-methyl-3-phenylpropylamine. *Life Sciences, 15*(3), 471–479.

Wong, D. T., et al. (1975a). A new selective inhibitor for uptake of serotonin into synaptosomes of rat brain: 3-(p-trifluoromethylphenoxy). N-methyl-3-phenylpropylamine. *The Journal of Pharmacology and Experimental Therapeutics, 193*(3), 804–811.

Wong, D. T., et al. (1975b). A new selective inhibitor for uptake of serotonin into synaptosomes of rat brain: 3-(p-trifluoromethylphenoxy). N-methyl-3-phenylpropylamine. *The Journal of Pharmacology and Experimental Therapeutics, 193*(3), 804–811.

Wong, D. T., Perry, K. W., & Bymaster, F. P. (2005). Case history: The discovery of fluoxetine hydrochloride (Prozac). *Nature Reviews Drug Discovery, 4*(9), 764–774.

Worrall, J. (2010). Evidence: Philosophy of science meets medicine. *Journal of Evaluation in Clinical Practice, 16*(2), 356–362.