

## University of Groningen

### Rationality in discovery

Bosch, Alexander Petrus Maria van den

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*

Publisher's PDF, also known as Version of record

*Publication date:*

2001

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Bosch, A. P. M. V. D. (2001). *Rationality in discovery: a study of logic, cognition, computation and neuropharmacology*. s.n.

**Copyright**

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

**Take-down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

## Chapter 9

---

# Discovery

### 9.1 Introduction

In Chapter 8 I reported on my own epistemological experiment, where I observed and inquired about a scientific practice, the process of discovery in neuropharmacology. Chapter 7 reported on a part of the theory that is used and developed in that practice. In this final chapter I analyze both the theory and practice, using the concepts from my theoretical discussion of discovery in Part II. The particular question that is answered in this chapter is: what is the rational use of theory and experiment in neuropharmacology? For my description of discovery in neuropharmacology I will pursue answers to the three specific questions of this thesis, *i.e.* 1) what is the structure of a scientific theory?; 2) what is the process of scientific reasoning?; and 3) what is the route between theory and experiment?

In answering these questions in this chapter I combine the theoretical approaches of logic as introduced in Chapter 4, and cognitive science, as discussed in Chapter 5. My main goal is to describe the practice of neuropharmacology. I will use the problem solving concepts from cognitive psychology to describe steps in the *process* of discovery, while I use the concepts of the logical approach to describe the *products* of that process. My aim is not to explain the particular directions of the search process that is described, by extracting and representing implicit knowledge as production rules. Those rules are dependent on the personal experiences of researchers and learned in a particular practice, as argued in Sections 5.7 and 5.8.

To analyze the structure of the DA theory I will first, in section 9.2, introduce a logical approach to represent the structure of theories in general, and dynamical systems in particular. Then, in section 9.3, I formally represent the theory of the basal ganglia as a qualitative differential equation, to answer the first question of this thesis for the case study. Before going into the second question, the third question is addressed in section 9.4, where I describe the route between theory and experiment in the problems faced in the practice of neuropharmacology. In section 9.5 I go into the process of reasoning in explanation, prediction and design. I will also discuss how a description of that process could be applied in that practice. Finally, in section 9.6 I end with a general conclusion, discussing the consequences of my observations and analysis of the case for the theory about discovery as discussed in Part II.

## 9.2 Models

The first question I will address is, how to understand the structure of the DA theory of Parkinson's disease. And secondly, how does it explain the effect of known treatments. In this section I will introduce a model theoretic approach to the structure of theories.

The structuralist approach in the philosophy of science characterizes a theory by its models, conceived as structures. (Th.A.F. Kuipers, 2000). A structure, in this context, is usually represented as an ordered set of variables, functions and constants. A structure is called a model of a theory if the theory, seen as a proposition about that structure, is true.

The core of a theory consists of a set of models  $M$  which is a subset of all conceptually possible models  $M_P$  given the vocabulary of the theory. The difference between  $M_P$  and  $M$  are all models that the theory excludes and is called the empirical content of a theory. It contains all the potential falsifiers of the theory. Given a domain  $D$  of application of the theory it is assumed that there is a subset of  $M_P$  that are the empirically possible models of that domain. A weak empirical claim states that all empirically possible models are models of the theory, a strong claim also asserts that they are equal.

For my exposition I will characterize a theory by its vocabulary of variables, the quantity spaces of those variables (a quantity space of a variable defines the range and type of values of a variable), and constraints on the values of those variables, given that they represent together the set of possible models and models of the theory. I will further make a distinction between a theory  $T$ , which is basically a set of definitions, and a hypothesis  $H$  which is a statement that asserts that the properties of phenomena in domain  $D$  can be characterized by the vocabulary  $V$  and by the models of theory  $T$ .

**Definition 1** *Theory*. The ordered set  $\langle V, Q, C \rangle$  of variables  $V$ , quantity spaces  $Q$  and constraints  $C$  represents a theory. The theory determines an ordered set  $\langle M_P, M_T \rangle$  that contains the conceptually possible models  $M_P$ , given  $V$  and  $Q$ , and the models of the theory  $M_T$ , given the constraints  $C$  on  $V$ .

**Definition 2** *Hypothesis*. The ordered set  $\langle V, Q, C, D \rangle$  represents a hypothesis where a theory is applied to a domain  $D$ . The hypothesis determines the ordered set  $\langle M_P, M_T, M_E \rangle$ , that contains the conceptually possible models  $M_P$  of a domain  $D$  given possible descriptions by variables  $V$  and quantity spaces  $Q$ ; the models  $M_T$  of the theory of the domain given constraints  $C$  on  $V$ ; and the empirically possible models  $M_E$  of the phenomena of domain  $D$ . The hypothesis asserts that the set of empirically possible models  $M_E$  is a subset of, or equal to, the set of models  $M_T$  of the theory.

A model of a phenomenon in a domain is a structure that represents certain aspects of that phenomenon in terms of a set of interpreted variables with particular quantities. The structures that are possible according to the constraints  $C$  from a theory are called the models  $M_T$  of that theory. The conceptually possible models  $M_P$  is the set

of all the models that are possible if you combine all possible variables from  $V$  with all their possible quantities from  $Q$ .

The relation between the conceptually possible models  $M_P$ , the models of the domain  $M_E$  and the models  $M_T$  of a theory in a hypothesis can be graphically represented as in Figure 9.1. The different intersections represent subsets of structures that constitute either a success, an anomaly or a problem for the theory. The goal of explanation is to find a hypothesis, such that a better hypothesis has less problems (subset 1) or anomalies (subset 3) than a competitor (Th. Kuipers, 1992, p.303).

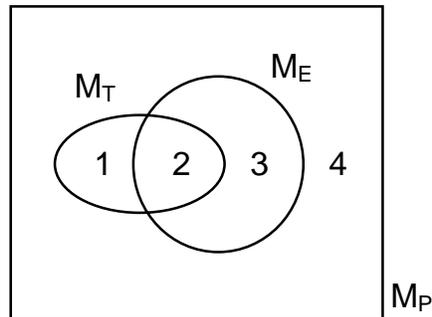


Figure 9.1: Models  $M_T$  of a hypothesis and empirically possible models  $M_E$  of the phenomena of a domain, both part of the conceptually possible models  $M_P$

Subset	$M_T$	$M_E$	
1	1	0	Explanatory problem
2	1	1	Explanatory success, confirming instance
3	0	1	Empirical anomaly, counter example
4	0	0	Explanatory success

Table 9.1: Subsets of conceptually possible models  $M_P$  of a domain

To understand the theory of Parkinson's disease we can understand it to be a hypothesis about the dynamical behavior of the brain. The theory asserts what kind of states and behaviors are possible. The set  $V$  and  $Q$  describe the known structural properties of the brain, and the constraints in  $C$  describe the assumed functional relations between those properties. A variable  $x$  of a structure is related to variable  $y$  if there is a functional constraint in  $C$ , such that  $y = f(x)$ .

### Disease and intervention

To understand the research problems in pharmacology we need to extend our vocabulary. Pharmaceutical research is not only interested in how to explain observations of a pathological biological system. It also aims to know how to treat it, and why a treatment works. For this we can introduce two extra subsets of  $M_P$ , the models of a biological system that is influenced by a (drug) intervention,  $M_I$ , and the models of phenomena that we wish to cause, the set  $M_W$ , see Figure 7.3.

Given a set of conceptually possible models of the behavior of a biological system a set of drug interventions can be assumed to cause behaviors represented by the set  $M_I$ , while the set  $M_W$  represents the set of wished for behaviors. Let  $M_E$  represent the

empirically possible behaviors of a living organism with a given biological structure. Hence if the assumptions are correct  $M_I$  should be a subset of  $M_E$ .

In Figure 9.2 subset 1 denotes an undesired behavior that is not treated by known interventions. Subset 2 contains unsuccessfully treated system behavior and unwanted side effects of a partially successful drug treatment, while subset 3 denotes behavior that is successfully treated. Subset 4 may be equal to health, given that  $W$  denotes health. Subset 5 can contain a behavior that is not possible given the biological structure of the organism, but can still be desired. Subset 6 equals the periphery of both possibility and interest.

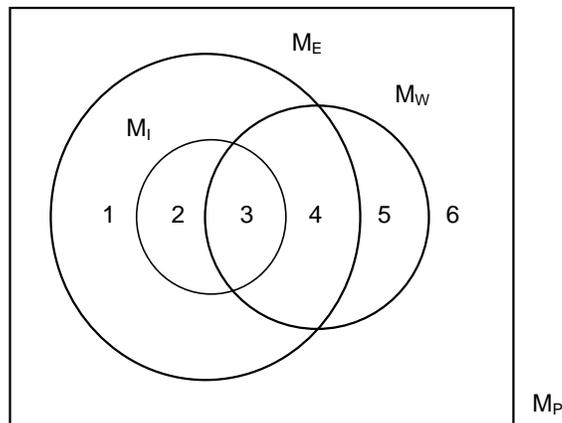


Figure 9.2: Empirically possible models  $M_E$  of a biological system, wished for models  $M_W$ , and models  $M_I$  of a system that is influenced by an intervention, all part of conceptually possible models  $M_P$  of a biological system

Subset	$M_E$	$M_I$	$M_W$	
1	1	0	0	Disease, untreated by known interventions
2	1	1	0	Disease, treated with side effects
3	1	1	1	Successfully treated
4	1	0	1	Health
5	0	0	1	Desired, but not empirically possible
6	0	0	0	periphery of interest and possibility

Table 9.2: Subsets of conceptually possible models  $M_P$  of a biological system

These three sets define the main goals of neuropharmacology. It is a goal to describe and explain  $M_E$ , what kinds of values of variables describing the brain and behavior of the organism are empirically possible, and why? It is also a goal to determine what kinds of states and behaviors  $M_W$  constitute health, or are desired for other reasons. And finally what kind of drug or other medical interventions cause those desired behaviors.

### Dynamical systems

In neurobiology the function of the brain is described and explained as a complex dynamical system. In physics, the most powerful tool to model a dynamical system is by making use of differential equations. Variables represent properties of the system

whose values can change over time. By defining the specific relations between those variables those values can be predicted, given an initial state of the system.

Empirical studies of both the brain and behavior in Parkinson research results in many quantitative data, correlating variables of the activation frequency of nuclei and neural pathways and local concentrations of different kinds of neurotransmitters. Yet those relations are not sufficiently known to define them as a quantitative equation. The relation is only known qualitatively. Many results of empirical studies of the brain amount to conclusions such as *e.g.* if the value of this variable changes in this direction, the change of the value of that variable in that direction is statistically significant. In this way the theory that explains Parkinson's disease can explain why the activation of the thalamus decreases, when the concentration of DA in the striatum significantly decreases.

While these results are insufficient to define a model with the aid of an ordinary differential equation, they can be represented by a more abstract qualitative differential equation (QDE), *cf.* (B. Kuipers, 1994). A QDE can be defined as follows:

**Definition 3** *Qualitative differential equation.* A Qualitative differential equation (QDE) represented by the ordered set  $\langle V, Q, C \rangle$  is an abstraction of an ordinary differential equation (ODE):

$$dV/dt = C(V)$$

where  $V$  is a set of variables each of which is a *reasonable function* over time, whose values are described in a finite set of qualitative landmark values belonging to  $Q$ , and  $C$  is a set of constraints between those variables.

**Definition 4** *Reasonable function*  $v \in V$ . Given an interval of the set of real numbers extended with  $\infty$  and  $-\infty$ ,  $[a, b] \subseteq \mathfrak{R}^*$ , the function  $v: [a, b] \rightarrow \mathfrak{R}^*$  is a reasonable function over  $[a, b]$  if

1.  $v$  is continuous over  $[a, b]$
2.  $v$  is continuously differentiable over  $(a, b)$
3.  $v$  has only finitely many critical points in any bounded interval,
4. the one-sided limits  $\lim_{t \rightarrow a} v'(t)$  and  $\lim_{t \rightarrow b} v'(t)$  exist in  $\mathfrak{R}^*$  and are defined as to be equal to  $v'(a)$  and  $v'(b)$ , respectively.

**Definition 5** *Quantity space*  $q \in Q$ . A quantity space  $q$  is a finite, totally ordered set of landmark symbols such as  $-\infty < l_1 < \dots < 0 < \dots < l_k < \infty$  that describe qualitatively important distinctions for a variable.

Important distinctions described by landmarks are values of variables that change in time and become steady or start to increase or decrease at a certain time-point. The qualitative value of a variable  $v$  at time-point  $t$  is expressed by a landmark from its quantity space and a direction of change.

**Definition 6** *Qualitative value at a time-point.* The qualitative value  $QV(v, t)$  of a variable  $v(t)$  with respect to a quantity space  $q = \langle l_1, \dots, l_n \rangle$  is defined by the tuple  $\langle qmag, qdir \rangle$  where,

$$\begin{aligned} qmag = & \quad l_i & \quad \text{if } v(t) = l_i, \\ & (l_i, l_{i+1}) & \quad \text{if } v(t) \in (l_i, l_{i+1}) \\ \\ qdir = & \quad \text{inc} & \quad \text{if } v'(t) > 0 \\ & \quad \text{std} & \quad \text{if } v'(t) = 0 \\ & \quad \text{dec} & \quad \text{if } v'(t) < 0 \end{aligned}$$

In a QDE the possible values of the variables in  $V$  are constrained by constraints in  $C$ . The constraints in  $C$  can consist of constraints corresponding to additions, multiplications, negations, derivatives, and incompletely known functions specified only as being part of a monotonicity class.

The last category is relevant for our case. We can know about a function  $f$  between two variables  $v_1(t)$  and  $v_2(t)$ ,  $v_1(t) = f(v_2(t))$ , that it belongs to either  $M^+$ , the class of monotonically increasing functions, or  $M^-$ , the class of monotonically decreasing functions. That is, for every  $f \in M^+$ ,  $f' > 0$ , and for every  $f \in M^-$ ,  $f' < 0$  over the domain of the function. These classes can be generalized to multivariate functions so that e.g.  $M^{+-}$  is the class of functions  $v_1(t) = f(v_2(t), v_3(t))$ , such that  $\partial f / \partial v_2 > 0$  and  $\partial f / \partial v_3 < 0$ .

The constraints  $C$  in a QDE define which qualitative states and behaviors are possible. So  $C$  amounts to a theory about a system. We can define the qualitative state of a dynamical system at a distinguished time-point, or on an interval between two distinguished time-points.

**Definition 7** *Qualitative state.* The qualitative state of a dynamical system described by  $m$  variables  $V$  at time-point  $t_i$  is an ordered set of individual qualitative values at a certain time-point or time interval from  $t_i$  to  $t_{i+1}$ :

$$\begin{aligned} QS(V, t_i) &= \langle QV(v_1, t_i), \dots, QV(v_m, t_i) \rangle \\ QS(V, t_i, t_{i+1}) &= \langle QV(v_1, t_i, t_{i+1}), \dots, QV(v_m, t_i, t_{i+1}) \rangle \end{aligned}$$

The qualitative behavior of a dynamical system can now be defined as an ordered set of qualitative states:

**Definition 8** *Qualitative behavior.* The qualitative behavior of a dynamical system with variables  $V$  on time interval  $[t_0 < \dots < t_n]$  is a sequence of qualitative states:

$$QB(V) = \langle QS(V, t_0), QS(V, t_0, t_1), QS(V, t_1), \dots, QS(V, t_n) \rangle$$

The possible states and behaviors of a system can be seen as models of the differential equation. Benjamin Kuipers developed a computer program called QSIM that can generate those models. It takes as input a QDE and an initial qualitative state description and produces a tree of possible state sequences. This can be seen as:

$$\text{QSIM}(\langle V, Q, C \rangle, \text{QS}(t_0)) = M$$

such that  $M$  is an ordered set  $\langle S, B \rangle$ , where  $S$  is a set of all possible qualitative states and  $B$  is a set of all possible qualitative behaviors, *i.e.* totally ordered sets of qualitative states, consistent with  $C$  (*cf.* Schultz and B. Kuipers, 1994).

It can be proved that given an ordinary differential equation (ODE) and its QDE abstraction, all abstractions of genuine behaviors of the ODE are generated by QSIM, but also some behaviors that are not an abstraction of a genuine ODE behavior. It can predict spurious behaviors, not predicted by a numerical solution of the ODE. It remains an open research problem whether qualitative solutions can be made complete or are inherently incomplete. But to put this problem in perspective, numeric algorithms also may produce non-sensical solutions due to round-off errors and careless simulation around singular points (De Jong, 1998).

In the next section I will use the QDE representation to explicate the structure of the dopamine theory of Parkinson's disease, and how it explains the function of known treatments.

### 9.3 Theory

Neurobiologists study the processes of the brain, *e.g.* by recording values of activation frequencies and concentrations of neurotransmitters in different locations of the brains of guinea pigs, Wistar rats, or monkeys. When the values of two variables  $v_1$  and  $v_2$  are consistent with a monotonic function in all trials of an experiment, a correlation could be proposed.

This is a simple style of descriptive induction, they are monotonically related in the sample, so they are monotonically related in all brains, of the sample organism or even in the human brain. It becomes an explanation if it is hypothesized what processes make it so that the variables act in that way.

In Parkinson research it is observed that the increase of symptoms is correlated with a substantial decrease of the availability of the neurotransmitter DA, which is due to a decay of the substantia nigra pars compacta (SNC). The model of the basal ganglia aims to explain why the decrease of DA can lead to these symptoms, by explaining why the activation of the SNR increases as a result of this decrease.

I will formally reconstruct this explanation by first representing the model of the basal ganglia as a qualitative differential equation. This equation serves as a hypothesis from which it can be deduced that given a decrease of DA, an increase of the SNR activation is a consequence.

I will also show how the activity of known treatments can be explained as well. This elaborate reconstruction will serve my analysis in section 9.4 about how the basal ganglia hypothesis itself is tested experimentally. The process of drug design will be discussed in section 9.5, where I will argue how such explicit models can be used to infer possible new interventions.

### Theory of the basal ganglia

The basal ganglia theory is a qualitative theory about a dynamical system, so we can represent it as a QDE. We defined a QDE as a tuple  $\langle V, Q, C \rangle$ , where  $V$  is a set of variables which are reasonable functions over time,  $Q$  is a set of quantity spaces for those variables, and  $C$  is a set of constraints. In the basal ganglia theory there are two basic variables describing firing rate ( $f$ ) of nerve cells in a cell group, nuclei or pathway, and the amount ( $a$ ) of a particular neurotransmitter released in the vicinity of a cell group, nuclei or neural pathway. The relation  $d/dt y = M^+ x$  is used to state that the change of values of  $y$  over time is monotonically related to the change of value of  $x$ . It is a matter of debate whether this relation represents a causal direction from  $x$  to  $y$ , for discussion see Iwasaki & Simon (1994).

I will represent the model of the basal ganglia as depicted in Figure 3.1, which was used by dr. Timmerman (1992). While this model could be further extended to include other influences, such as those of substance P and enkephalin, as depicted in Figure 7.4, the simpler model suffices for my analysis of the observed practice. The notation  $x$ -to- $y$  in the cell groups denotes the neural pathway from cell group  $x$  to cell group  $y$ . I further abbreviate SNR/Gpi to SNR. So, we can define the basal ganglia theory as follows:

**Definition 9** *Basal ganglia theory.*  $T_{BG} : \langle V, Q, C \rangle$  is a QDE such that:

#### 1. Variables in $V$

- Cell groups  $G$ , containing nuclei and neural pathways

$G$ : {striatum, GPe, STN, SNR, thalamus, brainstem, cortex-to-striatum, SNC-to-striatum, striatum-D1-to-SNR, striatum-D2-to-GPe, GPe-to-SNR, GPe-to-STN, STN-to-SNR, SNR-to-thalamus, SNR-to-brainstem}

- Set of neurotransmitters  $N$ : {Glu, DA, GABA}
- The firing rate  $f(g)$  of cell group  $g$  is a value of quantity space  $F$

$$f: G \rightarrow F$$

- Amount  $a(n, g)$  of neurotransmitter  $n$  in cell group  $g$  is a value of  $A$

$$a: N \times G \rightarrow A$$

#### 2. Quantity spaces in $Q$

- Boundaries of firing rates  $F$ : {0, MAX}
- Boundaries of amounts  $A$ : {0, MAX}

### 3. Constraints in C

- Firing rate of nuclei in the basal ganglia

**C.1**  $d/dt f(\text{striatum}) = M^+ a(\text{Glu, striatum})$   
**C.2**  $d/dt f(\text{GPe}) = M^- a(\text{GABA, GPe})$   
**C.3**  $d/dt f(\text{STN}) = M^- a(\text{GABA, STN})$   
**C.4**  $d/dt f(\text{SNR}) = M^- (a(\text{GABA, SNR}), a(\text{Glu, SNR}))$   
**C.5**  $d/dt f(\text{thalamus}) = M^- a(\text{GABA, thalamus})$   
**C.6**  $d/dt f(\text{brainstem}) = M^- a(\text{GABA, brainstem})$

- Firing rate of neural pathways between nuclei

**C.7**  $d/dt f(\text{cortex-to-striatum}) = M^+ f(\text{cortex})$   
**C.8**  $d/dt f(\text{SNC-to-striatum}) = M^+ f(\text{SNC})$   
**C.9**  $d/dt f(\text{striatum-D1-to-SNR/GPi}) = M^{++}(f(\text{striatum}), a(\text{DA, striatum}))$   
**C.10**  $d/dt f(\text{striatum-D2-to-GPe}) = M^{+-}(f(\text{striatum}), a(\text{DA, striatum}))$   
**C.11**  $d/dt f(\text{GPe-to-SNR}) = M^+ f(\text{GPe})$   
**C.12**  $d/dt f(\text{GPe-to-STN}) = M^+ f(\text{GPe})$   
**C.13**  $d/dt f(\text{STN-to-SNR}) = M^+ f(\text{STN})$   
**C.14**  $d/dt f(\text{SNR-to-thalamus}) = M^+ f(\text{SNR})$   
**C.15**  $d/dt f(\text{SNR-to-brainstem}) = M^+ f(\text{SNR})$

- Amount of released neurotransmitters in nuclei

**C.16**  $d/dt a(\text{DA, striatum}) = M^+ f(\text{SNC-to-striatum})$   
**C.17**  $d/dt a(\text{Glu, striatum}) = M^+ f(\text{cortex-to-striatum})$   
**C.18**  $d/dt a(\text{GABA, GPe}) = M^+ f(\text{striatum-D2-to-GPe})$   
**C.19**  $d/dt a(\text{GABA, STN}) = M^+ f(\text{GPe-to-STN})$   
**C.20**  $d/dt a(\text{GABA, SNR}) = M^{++}(f(\text{striatum-D1-to-SNR}), f(\text{GPe-to-SNR}))$   
**C.21**  $d/dt a(\text{Glu, SNR}) = M^+ f(\text{STN-to-SNR})$   
**C.22**  $d/dt a(\text{GABA, thalamus}) = M^+ f(\text{SNR-to-thalamus})$   
**C.23**  $d/dt a(\text{GABA, brainstem}) = M^+ f(\text{SNR-to-brainstem})$

- Metabolism of dopamine

**C.24**  $d/dt a(\text{DA, } x) = a(\text{L-dopa, } x) \times \text{Enzyme-ratio}$   
**C.25**  $\text{Enzyme-ratio} = a(\text{AADC, } x) / a(\text{MAO-B, } x)$

For brevity I include the assumptions about the metabolism of dopamine as part of the theory of the basal ganglia. The availability of dopamine outside the dopaminergic cell terminal is dependent on the activation of the cell by the neural pathway from the SNC. But DA can only be released by the vesicles of the terminal if the precursor L-dopa and the enzyme AADC is available. The enzyme MAO-B breaks down the excess of dopamine to DOPAC, see Figure 7.3.

### Explanation of Parkinson's Disease

The theory of the basal ganglia can be applied to explain observations in Parkinson's disease research. The hypothesis of the basal ganglia states that the empirically possible states E of the basal ganglia, given the empirical study of the basal ganglia D, are part of the theoretically possible states M.

**Definition 10** *Basal ganglia hypothesis.*  $H_{BG} : \langle V, Q, C, D \rangle$  represents a hypothesis about the basal ganglia brain structure where V, Q, C are part of the  $T_{BG}$  and D is the set of instances of the basal ganglia, the domain of application of the theory.

We saw that the symptoms of Parkinson's disease are assumed to be caused by an increase of activation of the SNR, which on its turn is explained by a steep decrease of DA in the striatum due to the decay of dopaminergic nerve cells. One question in this chain, how the observed decrease of DA causes the assumed increase of SNR activation, is explained by the theory about the basal ganglia. I will show how this proposition can be deduced from the basal ganglia theory by programs like QSIM. In this proposition and proof I will reduce the values of the variables to just their qualitative direction, abstracting from time and qualitative magnitude.

From  $d/dt y = f(x)$  where  $f \in M^+$  we know that x and y both increase or decrease together, while if  $f \in M^-$ , y increases when x decreases, and vice versa. If  $dz/dt = f(x, y)$  and  $f \in M^{++}$ , the direction of change of z is unknown if x increases and y decreases, since we do not now their magnitude, cf. Table 9.3. This is similar for  $f \in M^{+-}$ , when both variables increase or decrease in value

y \ x	inc	std	dec
inc	inc	inc	?
std	inc	std	dec
dec	?	dec	dec

Table 9.3: Derivative values for z if  $dz/dt = f(x,y)$  and  $f \in M^{++}$

As background assumptions we assume that the amount of dopamine in the striatum decreases and the firing rate of the striatum is steady. I will use the notation  $v = qdir$  as shorthand for  $\exists y \exists t QV(v, t) = \langle y, qdir \rangle$ .

**Theorem 1**  $H_{BG} \cup B : \{a(DA, striatum) = dec, f(striatum) = std\} \vdash$   
 $P : \{f(SNR) = inc\}$

**Proof:** As a proof I deduce the conclusion P from the premises B by applying the constraints C from the basal ganglia hypothesis  $H_{BG}$ .

$$a(DA, striatum) = dec \wedge f(striatum) = std \\ \Rightarrow f(striatum-D1-to-SNR) = dec \wedge f(striatum-D2-to-GPe) = inc \text{ (C.9, C.10)}$$

$$f(striatum-D2-to-GPe) = inc \\ \Rightarrow a(GABA, GPe) = inc \text{ (C.18)} \\ \Rightarrow f(GPe) = dec \text{ (C.2)}$$

$$\Rightarrow f(\text{GPe-to-SNR}) = \text{dec} \wedge f(\text{GPe-to-STN}) = \text{dec} \text{ (C.11, C.12)}$$

$$\begin{aligned} f(\text{GPe-to-STN}) &= \text{dec} \\ \Rightarrow a(\text{GABA, STN}) &= \text{dec} \text{ (C.19)} \\ \Rightarrow f(\text{STN}) &= \text{inc} \text{ (C.3)} \\ \Rightarrow f(\text{STN-to-SNR}) &= \text{inc} \text{ (C.13)} \\ \Rightarrow a(\text{Glu, SNR}) &= \text{inc} \text{ (C.21)} \end{aligned}$$

$$\begin{aligned} f(\text{GPe-to-SNR}) &= \text{dec} \wedge f(\text{striatum-D1-to-SNR}) = \text{dec} \\ \Rightarrow a(\text{GABA, SNR}) &= \text{dec} \text{ (C.20)} \end{aligned}$$

$$\begin{aligned} a(\text{Glu, SNR}) &= \text{inc} \wedge a(\text{GABA, SNR}) = \text{dec} \\ \Rightarrow f(\text{SNR}) &= \text{inc} \text{ (C.4)} \end{aligned} \quad (\text{Q.E.D.})$$

### Explanation of known treatments

I will first introduce a new set in my terminology. Next to a hypothesis H, background assumptions B, and propositions P that are explained or need to be explained, we also have a set of interventions I. This set contains propositions that describe a property of the world, usually a value of a particular variable, that can be set by a manipulation. All consequences of that manipulation hold for all the structures in the set  $M_I$ .

A theory can explain why a particular intervention has a particular consequence. With  $H_{BG}$  we have a hypothesis that explains the symptoms of Parkinson's disease by linking them to the observed decrease of DA. The hypothesis also explains the function of metabolites like L-dopa, MAO-B and AADC. These metabolites can serve as an artificial intervention by changing their amount with the aid of a drug. Parkinson drugs all serve to increase the amount of dopamine, which according to the theory would decrease the overactivation of the SNR, reducing the behavioral symptoms. In the theorems below I demonstrate how the basal ganglia hypothesis explains the activity of known drug interventions for Parkinson's disease. All these drugs aim to influence the amount of dopamine, so I first pose the following theorem:

**Theorem 2**  $H_{BG} \cup B: \{f(\text{striatum}) = \text{std}\} \vdash$   
 $P: \{a(\text{DA, striatum}) = \text{inc} \rightarrow f(\text{SNR}) = \text{dec}\}$

From  $H_{BG}$  it can be deduced in theorem 2 that an increase of DA implies a decrease of the firing rate of the SNR output nuclei of the basal ganglia. The proof follows similar lines as the proof of theorem 1.

Theorem 3 states that an increase of L-dopa in the striatum will increase DA in the striatum, which is a consequence of C.24, and given that the enzyme ratio does not increase.

**Theorem 3**  $H_{BG} \cup I: \{a(\text{L-dopa, striatum}) = \text{inc}\} \vdash P: \{a(\text{DA, striatum}) = \text{inc}\}$

But to increase L-dopa by a drug intervention, which is taken up in the bloodstream, means that L-dopa is increased in the whole body, causing side effects. A decrease of

the amount of AADC in the periphery by also administering an inhibitor that can not cross the blood brain barrier, will cause DA to increase in the brain, but to be relatively steady in the periphery. This theorem (4) is a consequence of C.24 and C.25, given the assumption that the amount of MAO-B does not increase in the periphery.

**Theorem 4**  $H_{BG} \cup I: \{a(\text{L-dopa, body}) = \text{inc}, a(\text{AADC, periphery}) = \text{dec}\} \vdash$   
 $P: \{a(\text{DA, striatum}) = \text{inc}, a(\text{DA, periphery}) = ?\}$

By C.25 and C. 25 one can also prove theorem 5, which states that decreasing the enzyme that breaks up DA will increase the amount of DA, assuming that both the amount of AADC and L-dopa in the striatum do not increase:

**Theorem 5**  $H_{BG} \cup I: \{a(\text{MAO-B, striatum}) = \text{dec}\} \vdash P: \{a(\text{DA, striatum}) = \text{inc}\}$

The function and activity of these treatments can be explained by the theory of the basal ganglia, but another question is if the hypothesis is true. That is, are all the states that are empirically possible also states allowed by the theory?

In Section 8.4 we saw how experiments showed that the background assumption in theorem 2 about the activation of the striatum turned out to be incorrect, by testing the predicted effect of selective dopamine receptor agonists. In the next section I will go into that problem.

## 9.4 Practice

In the second part of this thesis I discussed several ways of understanding rationality in the process of scientific discovery. In this discussion it was assumed that the main goal of scientific discovery is to gain knowledge about natural phenomena. In order to do so I made a distinction between five basic types of tasks as problems with different sub-goals: observation, description, explanation, prediction and intervention. These tasks, repeatedly executed in that order, could lead to an increasingly better knowledge of the natural world.

In this section I extensively analyze the actual problems and tasks that are tackled in the practice of neuropharmacology that I observed. To describe a problem I will make use of the distinctions made in Chapter 5. That is, pursuing a problem is characterized by the following constituents:

### Problem

Start:	propositions about the initial situation
Goal:	the conditions for a problem to be solved
Result:	the propositions describing the result
Process:	the kind of action that is used to pursue the goal.

To abstractly distinguish different contents of propositions I will make use of different sets of propositions that describe:

**Propositions**

O:	observations
I:	interventions
H:	hypotheses
W:	wished for properties
P:	predictions

A question-mark after the name of a set, *e.g.* H?, will designate the set or property that is the goal of the problem. A star after the name of a set, *e.g.* H\*, or a star in a set, *e.g.* H: {\*}, will mean that the set contains propositions, for which the truth-value is unknown, that describe new information relative to the initial situation. Processes that can lead to achieving a goal are distinguished as:

**Processes**

Intervention:	manipulating a property of a natural process
Observation:	observing a property of a natural process
Description:	describing a property of a natural process
Explanation:	finding an explanation for the initial situation
Prediction:	deducing a consequence of a initial situation
Design:	creating a property given a specification

Design as a process is added because it is needed to describe some problems of neuropharmacology involving wished for properties of a drug or a treatment. I will make a further distinction between a focused and broad kind of process, meaning that the activity is directed to a small or large set of properties.

In Chapter 5 these terms were used to designate particular problems and processes with a particular initial and goal situation, see table 5.10. I will now use them to describe (parts of) larger scale problems. In the next section I will go deeper into the process of reasoning and compare the theoretical archetypes with my observations.

In describing a problem I will first model the observed examples from the practice. I will loosely follow the order of the report of my interviews in Chapter 8. A problem never comes by itself, so I will describe different situations where the result of a particular problem leads to new problems that are addressed in that situation.

In Chapter 8 I reported on several situations that lead to different kinds of discoveries: new functions of known drugs are discovered in the clinic and further investigated in the lab; it is investigated what a desired function of a new drug should be; given that wished for function new drugs are designed, searched, predicted, created and tested; new drugs are tested and investigated in the clinic; they are also used to explore biological systems in the lab; given that exploration, new treatments are designed and tested; and the function of a drug is explored and explained. Below I analyze the structure of those problems. The specific problem of exploring and explaining the function of DA is analyzed in detail. I summarize and generalize the examples in Table 9.4 at the end of this section.

### Discovering a new drug effect in the clinic

In Section 8.2 It is reported that chlorpromazine was specifically administered to treat particular diseases in the clinic (*cf.* Sec. 8.2, Par. 2). It was noted that another aspect than the one that was to be treated was influenced as well. This was followed by lab investigations that aimed to first observe and to explain the drug effect on another level. So we can distinguish a sequence of three kinds of problems: clinical drug treatment; lab investigation of the observed drug effect; explanation of the observed drug effect. This practice can be described as follows (*cf.* Table 9.1a):

#### Clinical drug treatment

Start: O: {particular disease}  
 Goal: Clinic I: {chlorpromazine} → W: {treatment}  
 Result: I → Clinic O\*: {parkinsonism}  
 Process: Focused intervention, broad observation

#### Lab investigation of observed drug effect

Start: Clinic I: {chlorpromazine} → O  
 Goal: Lab I: {chlorpromazine} → O?  
 Result: Lab O\*: {amount of DA release}  
 Process: Focused intervention, focused observation

#### Explanation of observed drug effect

Goal: H? ⊨ I: {chlorpromazine} → O\*: {amount of DA}  
 Result: H: {chlorpromazine = DA-antagonist,  
                   DA is related to Parkinson' disease}  
 Process: Focused explanation

### Searching a desired drug effect

In the above situation a potential treatment for a disease was discovered while it was initially not the goal of the activity that led to the discovery to study the function of that drug for that disease. Given that a desired effect was observed it was further investigated what the specific effect was on a lower biological level (*cf.* Sec. 8.2, Par.2).

When it was discovered that dopamine was related to Parkinson's disease it was investigated what the specific effect of dopamine is, to determine what a drug that should act as dopamine should do. So part of the result of the investigation is a profile for a new drug. The activity of a new drug such as a specific DA receptor agonist can then be compared to the desired activity of dopamine. The problem goes as follows (*cf.* Table 9.1b):

#### Lab exploration of wished drug effects

Start: I: {in vivo/ vitro DA}  
 Goal: I → O?: {receptor activity?}  
 Result: I → W\*: {amount of C-Amp release}  
 Process: Focused intervention, focused observation

**Lab testing for wished drug effect**

Start: I: {in vivo/ vitro, specific DA-agonists}  
 Goal:  $I \rightarrow W?$ : {receptor activity?}  
 Result:  $I \rightarrow O^*$ : {amount of C-Amp release}  
 Process: Focused intervention, focused observation

**Designing a new drug**

In Groningen professor Horn aimed to design a variant of dopamine that had a similar activity and metabolism as dopamine itself, but had also effects that made it more useful as a drug, such as specific receptor selectivity and lipophilicity (*cf.* Sec. 8.2, Par. 4-5). The suggestion for variants that Prof. Horn considered were based on his experience and knowledge of the field (Par. 6). While that knowledge may not always be explicit, it could still imply the suggestions. Success of a suggestion is hard to predict given the partly uncertain and incomplete nature of knowledge at a scientific frontier.

Professor Horn used his knowledge to design variants of ADTN. He could explain how a propyl and hydroxy group could respectively aid the lipophilicity and metabolism of the variant. So we can describe the process as designing, testing and explaining a new drug (*cf.* Table 9.1c):

**Rational drug design**

Start: I: {ADTN}  
 Goal:  $H \models I? \rightarrow W$ : {activity, lipophilicity, metabolism, selectivity}  
 Result:  $I^*$ : {ADTN-variant}  
 Process: Focused design

**Lab testing for wished drug effect**

Start: I: {ADTN-variant}  
 Goal:  $I \rightarrow W?$ : {activity?, lipophilicity?, metabolism?, selectivity?}  
 Result:  $I \rightarrow O^*$   
 Process: Focused intervention, focused observation

**Explanation of drug effect**

Start: I: {ADTN-variant}  $\rightarrow$  O: {lipophilicity, metabolism}  
 Goal:  $H? \models I \rightarrow O$   
 Result: H: {propyl group  $\rightarrow$  lipophilicity, hydroxy group  $\rightarrow$  metabolism }  
 Process: Focused explanation

**Searching a new drug effect in a drug library**

For a pharmaceutical company the results of the process designing new drugs leads to a library of novel compounds that are created with a specific goal, a given set of criteria (*cf.* Sec. 8.2, Par. 8). Often these criteria include the selectivity for a particular known receptor. A new drug treatment can be discovered by testing those drugs on other receptors by trial and error. In this process the drugs are given, and only mas-

sively tested on one criterion. A compound that is found to be active can be the lead for a new drug to target the new receptor, (*cf.* Table 9.1d).

#### Lab testing for wished drug effect

Start: I: {in vitro, all compounds of company on new receptor}  
 Goal:  $I \rightarrow W?$ : {receptor activity?}  
 Result:  $I \rightarrow O^*$ : {amount of C-Amp release}  
 Process: Broad intervention, focused observation

#### Searching a new drug by combinatorial chemistry

A still broader approach is taken when a drug lead is not specifically varied, based on *fingerspizengefühl* and personal experience, but by techniques from combinatorial chemistry (*cf.* Sec. 8.2, Par. 9). In this approach many variants are created at once. In this process the combinations are made and massively tested for activity. If activity is measured, the responsible variant is retrieved and further explored for its structure, (*cf.* Table 9.1e):

#### Combinatorial drug design

Start: I: {drug lead}  
 Goal:  $I? \rightarrow W$   
 Result:  $I^*$ : {many variants by combinatorial chemistry}  
 Process: Broad design

#### Lab testing for wished drug effect

Start: I: {in vitro, all variants on a receptor}  
 Goal:  $I \rightarrow W?$ : {receptor activity?}  
 Result:  $I \rightarrow O^*$ : {C-Amp release}  
 Process: Broad intervention, focused observation

#### Lab exploration of drug effect

Start: I: {retrieved drug}  $\rightarrow$  O: {high receptor activity}  
 Goal:  $O?$   
 Result:  $O^*$ : {chemical structure}  
 Process: Focused intervention, focused observation

#### Searching a new drug by computational modeling

While the trial and error approach in combinatorial chemistry is very expensive, the cheaper knowledge based approach by computer modeling is less successful. In this approach one starts with a computer model of the structure of a receptor and a drug (*cf.* Sec. 8.2, Par. 10). The goal is to predict by a simulation how a drug will dock (interact with a receptor), or how the receptor will fold, (*cf.* Table 9.1f).

#### Computational drug design

Start: I, H  
 Goal:  $H \models I? \rightarrow W$   
 Result:  $H \models I^* \rightarrow P^*$ : {drug docking, protein folding}

**Testing a new drug**

When a promising new drug or drug function is found and explored in the lab, it will leave the lab for further tests. As reported there are three different test phases (*cf.* Sec. 8.2, Par. 11). In the first phase the drug is tested for its possibly toxic effects on a specifically selected group of animals and volunteers. In phase two the focus of intervention changes to selected patients, where therapeutic effects are tested. In phase three this group is further extended, and the drug is used in hospitals and will undergo double blind tests, (*cf.* Table 9.1g).

**Clinical testing Phase I**

Start: I: {new drug}  
 Goal: I: {drug on animals, volunteers} → O?: {toxicity?}  
 Result: I → O\*  
 Process: Focused intervention, broad observation

**Clinical testing Phase II**

Goal: I: {volunteer patients} → W?: {therapeutic effect?}  
 Process: Focused intervention, focused observation

**Clinical testing Phase III**

Goal: I: {double blind, hospitals} → W?: {therapeutic effect?}  
 Process: Broad intervention, focused observation

**Exploring a biological system**

Highly selective drugs are also being used to explore and chart biological systems and to find out the function of specific drugs (*cf.* Sec. 8.2, Par. 12). The goal of Dr. Timmerman is to chart a system like the basal ganglia, using a broad range of selective interventions to explore it (*cf.* Sec. 8.3, Par. 17-18, 20).

**Lab exploration of a biological system**

Start: I: {different selective agonists, antagonists}  
 Goal: I → O?: {behavior?, local transmitter response?, electric activity?}  
 Result: I → O\*: {stereotypical beh., amounts of transm., firing frequency}  
 Process: Broad intervention, broad observation

This is achieved by focusing on the effects of particular drugs. This exploration is also undertaken in the case of a pathological system. For Parkinson's disease the function of dopamine in the basal ganglia is being studied (*cf.* Sec. 8.3, Par. 19). (This case will be analyzed in detail further below). The pathological situation is studied in rats whose dopamine cells are lesioned (*cf.* Sec. 8.3, Par. 35).

**Lab exploration of a drug effect**

Start: I: {normal rat/lesioned rat, local infusion of selective drug}  
 Goal: I → O?: {behavior?, local transmitter response?, electric activity?}  
 Result: I → O\*: {stereotypical beh., amounts of transm., firing frequency}  
 Process: Focused intervention, broad observation

Information about the observed difference between a healthy and pathological system can then be used to understand compensation mechanisms and in the design of new treatments (*cf.* Table 9.1h).

### Designing a drug treatment

When an observed difference between a healthy and pathological system is explained on a biochemical level, as in the case of Parkinson's disease, this difference can be used to rationally design a treatment. The goal is to intervene in such a way that the difference is minimized. In Parkinson's disease the goal is to restore the function of dopamine to normal (*cf.* Sec. 8.3, Par. 24-26, 34-35).

#### Rational drug treatment design

Start:  $H \models O$ : {depletion of DA  $\rightarrow$  Parkinson's disease symptoms}  
 Goal:  $H \models I? \rightarrow W$ : {restored DA function, best effect on symptoms }  
 Result:  $I^*$ : {DA-agonists?, selective D1? and/or D2?, NMDA-antagonists?}

In the case of Parkinson's disease it is not clear how best to restore the function of dopamine. Different interventions are designed and tested on their effect on disease symptoms. But all give rise to different (side) effects (*cf.* Sec. 8.3, Par. 27-30).

#### Clinical testing of a drug treatment

Start:  $I$   
 Goal:  $I \rightarrow W?$ : {effect on symptoms?}  
 Result:  $O$ : {hyper or poor response, effects in brain periphery, loss of effect}

It is also a problem to know what it means to restore a function to normal. One way to tackle this problems is to vary a dose by trial and error to search for a desired response. Yet targeting the dopamine receptor via the bloodstream also induces nausea via the extra stimulation of DA-receptors in the periphery. Hence part of designing a treatment of disease symptoms is designing treatment for the side effects of that treatment (*cf.* Table 9.1i).

#### Design of a treatment of side effects

Start:  $H \models I$ : {DA-agonists in bloodstream}  $\rightarrow$   $O$ : {hyper or poor response}  
 Goal:  $H \models I? \rightarrow W$ : {desired response}  
 Result:  $I^*$ : {vary doses DA-agonists by trial & error}

#### Design of a treatment of side effects

Start:  $H \models I$ : {DA-agonists in blood stream}  $\rightarrow$   $O$ : {nausea}  
 Goal:  $H \models I? \rightarrow W$ : {DA not in periphery}  
 Result:  $I^*$ : {peripheral DA blockers}

**Exploring a drug effect**

It was discovered that acetylcholine also has an effect on Parkinson symptoms. As an explanation of this effect it was proposed that there might exist a brain mechanism that maintains a balance between acetylcholine and dopamine. This explanation was later confirmed by experimental research that discovered that acetylcholine cells respond to D2-agonists (*cf.* Sec. 8.3, Par. 31-33).

**Explanation of a drug effect**

Start: I: {acetylcholine antagonist} →  
           O: {effect on Parkinson's disease symptoms}  
 Goal: H?  $\models$  I → O  
 Result: H: {acetylcholine-dopamine balance}

**Lab exploration of drug effect**

Start: I: {D2-agonist, acetylcholine cell}  
 Goal: I → O?: {activity?}  
 Result: I → O\*: {inhibition acetylcholine cell}

Dopamine is assumed to be related to the activity of the substantia nigra reticulata (SNR) in the basal ganglia (*cf.* Sec. 8.3, Par. 13-16). The model of the basal ganglia implies that dopamine would act as a modulator of GABA activity, inhibiting the activity of the SNR (*cf.* Sec 9.3).

**Explanation of a drug effect**

Start: I: {low amount of DA} → O: {high SNR activity}  
 Goal: H?  $\models$  I → O?  
 Result: H<sub>BG</sub>: {DA is modulator of GABA activity in SNR}

To test this claim the interaction between dopamine and the SNR was further explored in Groningen by Dr. Timmerman. She set the problem to investigate the effects of dopamine specifically in the striatum (*cf.* Sec. 8.3, Par. 36). I analyze her approach to this problem in detail, following my report of her experiments, and making use of the QDE formalism of Section 9.3. The problems are summarized in Table 9.4j. The general problem goes as follows:

**Exploration of a drug effect**

Start: H<sub>BG</sub>  $\models$  I → P  
 Goal: I: {dopamine in striatum} → O?  
 Result: I → O  
 Process: Focused prediction, focused observation

The model of the basal ganglia  $H_{BG}$  predicts that D1-agonists will excitate the direct pathway to the SNR, while D2-agonists will inhibit the indirect pathway (*cf.* Sec. 8.3, Par. 37-40).

### Prediction of a drug effect

Start:  $H_{BG}: \{DA \text{ is modulator of GABA activity in SNR}\} \models I \rightarrow O$   
 Goal:  $H_{BG} \models I: \{D1\text{-agonist, D2-agonist}\} \rightarrow P?$   
 Result:  $P^*: \{D1 \text{ excitation of direct pathway, D2 inhibition of indirect pathway}\}$

The problem is now to test whether the implications of the model are correct. First Dr. Timmerman explored three different predicted effects. She locally intervened the amounts of glutamate, a D1-agonist and a D2-agonist in the striatum under basal conditions, and observed the activity of the SNR (*cf.* Sec. 8.4, Par. 41-42). The predicted effects can be deduced from the axioms of  $H_{BG}$  in section 9.3.

### Lab testing of a predicted drug effect (in vivo)

Start:  $B: \{f(\text{striatum}) = \langle 0, \text{std} \rangle\}$   
 Goal:  $I \rightarrow O?$   
 Process: Focused intervention of glutamate, D1 and D2 receptors respectively  
 Focused observation of SNR activity

Start:  $B \cup H_{BG} \models I: \{a(\text{glutamate, striatum}) = \text{inc}\} \rightarrow P: \{f(\text{SNR}) = \text{dec}\}$   
 Result:  $I \rightarrow O^*: \{f(\text{SNR}) = \text{dec}\}$

Start:  $B \cup H_{BG} \models I: \{a(\text{D1-agonist, striatum}) = \text{inc}\} \rightarrow P: \{f(\text{SNR}) = \text{dec}\}$   
 Result:  $I \rightarrow O^*: \{f(\text{SNR}) = \text{dec, slight}\}$

Start:  $B \cup H_{BG} \models I: \{a(\text{D2-agonist, striatum}) = \text{inc}\} \rightarrow P: \{f(\text{SNR}) = \text{dec}\}$   
 Result:  $I \rightarrow O^*: \{f(\text{SNR}) = \text{inc, slight}\}$

Glutamate produced the predicted effect. The intervention with the D1-agonist only produced a very slight effect in the predicted direction, and the D2-agonist produced a slight effect against the predicted direction, but both were not significant. When the D1-agonist is tested in vitro a different effect than the one predicted is observed as well (*cf.* Sec. 8.4, Par. 43).

### Lab testing of a predicted drug effect (in vitro)

Start:  $B \cup H_{BG} \models I: \{a(\text{D1-agonist, striatum}) = \text{inc}\} \rightarrow P: \{f(\text{striatum-D1-to-SNR}) = \text{dec}\}$   
 Goal:  $I: \{a(\text{D1, slices striatum}) = \text{inc}\} \rightarrow O?$   
 Result:  $I \rightarrow O^*: \{a(\text{striatum-D1-to-SNR}) = \text{inc}\}$   
 Process: Focused intervention, focused observation

The explanation that was proposed to account for the observation of the slight effect of the selective dopamine agonists was that under basal conditions, there is no GABA activity to modulate (*cf.* Sec. 8.4, Par. 44). So a starting condition with a higher activation of the striatum should show an effect of a dopamine-agonist infusion.

### Prediction of a drug effect

Start:  $B: \{f(\text{striatum}) = \langle 0, \text{std} \rangle\}, H_{BG}$   
 Goal:  $B \cup H_{BG} \models I: \{f(\text{striatum}) = \langle +, \text{std} \rangle, a(\text{DA}, \text{striatum}) = \text{inc}\} \rightarrow P^*?$   
 Result:  $P^*: \{f(\text{SNR}) = \text{inc}\}$

The problem now for exploring the effect of the dopamine agonist is that it is important to maintain a steady activity of the striatum. If two variables vary then it is difficult to explain the effect on a third variable by pointing to only one of those two. So the goal is first to find an intervention that causes a desired effect that is needed as an initial condition for the experiment that tests another intervention (*cf.* Sec. 8.4, Par. 45-47).

### Rational design of an experimental condition

Start:  $B: \{f(\text{striatum}) = \langle 0, \text{std} \rangle\}, H_{BG}$   
 Goal:  $B \cup H_{BG} \models I? \rightarrow W: \{f(\text{striatum}) = \langle +, \text{std} \rangle\}$   
 Result:  $I^*$   
 Process: Focused design

### Lab exploration of a predicted drug effect

Start:  $B: \{f(\text{striatum}) = \langle 0, \text{std} \rangle\} \cup H_{BG} \models I \rightarrow P$   
 Goal:  $I \rightarrow O?$   
 Process: Focused intervention, focused observation

For instance:

Start:  $B \cup H_{BG} \models I: \{a(\text{glu}, \text{striatum}) = \langle +, \text{std} \rangle\} \rightarrow P: \{f(\text{striatum}) = \text{std}\}$   
 Result:  $I \rightarrow O^*: \{f(\text{striatum}) \neq \text{std}\}$

Start:  $B \cup H_{BG} \models I: \{a(\text{glu}, \text{cortex}) = \langle +, \text{std} \rangle\} \rightarrow P: \{f(\text{striatum}) = \text{std}\}$   
 Result:  $I \rightarrow O^*: \{f(\text{striatum}) \neq \text{std}\}$

Start:  $B \cup H_{BG} \models I: \{a(\text{glu}, \text{thalamus}) = \langle +, \text{std} \rangle\} \rightarrow P: \{f(\text{striatum}) = \text{std}\}$   
 Result:  $I \rightarrow O^*: \{f(\text{striatum}) \neq \text{std}\}$

Given the model it is predicted that the striatum can be activated directly with glutamate in the striatum, or indirectly with glutamate in the cortex or thalamus (*cf.* Sec. 8.4, Par. 48). However, when tested it turns out that it is difficult to maintain a steady activation. This can be attributed to incorrect assumptions of the model but also to compensation mechanisms that are not included in it.



**Exploration of drug effect**Start:  $H_{BG}$ Goal: I:  $\{a(\text{DA}, \text{striatum}) \rightarrow O?\}$ Results: B:  $\{f(\text{striatum}) = \text{std}\} \cup H_{BG} \models$   
I:  $\{a(\text{D1}, \text{striatum}) = \text{inc}\} \rightarrow O: \{f(\text{SNR}) = \text{std}\}$   
I:  $\{a(\text{D2}, \text{striatum}) = \text{inc}\} \rightarrow O: \{f(\text{SNR}) = \text{std}\}$ B:  $\{a(\text{DA-agonist}, \text{striatum}) = \text{inc}\} \cup H_{BG} \models$ I:  $\{a(\text{glutamate-agonist}, \text{striatum}) = \text{inc}\} \rightarrow O: \{f(\text{SNR}) = \text{dec}\}$ **Summary**

In the last subsections I extensively analyzed the structure of example problems in the process of discovery neuropharmacological research. In Table 9.5 I summarize and generalize these examples. This practice shows that testing a new prediction of a hypothesis that explains an earlier observation is only one of many ways of making a discovery. All the different problems I discussed lead to different empirical and conceptual discoveries. In the next section I will go deeper into the process of reasoning in explanation and design.

Problem	Start	Goal	Result	Process
Clinical drug treatment	a. Discovering a new drug effect in the clinic			
	O: {pathologic}	I $\rightarrow$ W: {normal}	I $\rightarrow$ O*: {side effects}	Focused intervention Broad observation
	I $\rightarrow$ O: {side effects}	I $\rightarrow$ O?: {biochemistry}	I $\rightarrow$ O*: {mechanism}	Focused intervention Focused observation
Explanation of an observed drug effect	I $\rightarrow$ O	H? $\models$ I $\rightarrow$ O	H* $\models$ I $\rightarrow$ O	Focused explanation
Lab exploration of wished drug function	b. Searching a desired drug effect			
	I: {transmitter}	I $\rightarrow$ O?: {effect}	I $\rightarrow$ W*: {effect}	Focused intervention Focused observation
	I: {drug}	I $\rightarrow$ W?: {effect?}	I $\rightarrow$ O*: {effect}	Focused intervention Focused observation
Rational drug design	c. Designing a new drug			
	I: {drug lead}	H $\models$ I? $\rightarrow$ W: {effect}	I*: {variant}	Focused design
	I: {drug}	I $\rightarrow$ W?: {effect?}	I $\rightarrow$ O*: {effect}	Focused intervention Focused observation
Explanation of drug function	I $\rightarrow$ O	H? $\models$ I $\rightarrow$ O	H* $\models$ I $\rightarrow$ O	Focused explanation
Lab testing for wished drug effect	d. Searching a new drug effect in a drug library			
	I: {given drugs}	I $\rightarrow$ W?: {effect?}	I $\rightarrow$ O*: {effect}	Broad intervention Focused observation
Combinatorial drug design	e. Searching a new drug by combinatorial chemistry			
	I: {drug lead}	I? $\rightarrow$ W: {effect}	I*: {variants}	Broad design

Lab testing for wished drug effect	$I : \{ \text{variants} \}$	$I \rightarrow W? : \{ \text{effect?} \}$	$I \rightarrow O^* : \{ \text{effect} \}$	Broad intervention Focused observation
Lab exploration of drug structure	$I \rightarrow O : \{ \text{effect} \}$	$I \rightarrow O? : \{ \text{structure?} \}$	$I \rightarrow O^* : \{ \text{structure} \}$	Focused intervention Focused observation
Computational drug design	f. Searching a new drug by computational modeling $I, H$ $H \models I? \rightarrow W$ $H \models I^* \rightarrow P^*$			Focused prediction Focused design
Clinical drug testing, phase I	g. Testing a new drug $I$ $I \rightarrow O?$ $I \rightarrow O^*$			Focused intervention Broad observation
Clinical drug testing, phase II	$I$	$I \rightarrow W?$	$I \rightarrow O^*$	Focused intervention Focused observation
Clinical drug testing, phase III	$I$	$I \rightarrow W?$	$I \rightarrow O^*$	Focused intervention Broad observation
Lab exploration of a biological system	h. Exploring a biological system $I$ $I \rightarrow O?$ $I \rightarrow O^* : \{ \text{path., normal} \}$			Broad intervention Broad observation
Lab exploration of a drug effect	$I$	$I \rightarrow O? :$	$I \rightarrow O^* : \{ \text{path., normal} \}$	Focused intervention Broad observation
Rational drug treatment design	i. Designing a drug treatment $H \models O : \{ \text{pathologic} \}$ $H \models I? \rightarrow W : \{ \text{normal} \}$ $I^*$			Focused design
Clinical drug testing	$I$	$I \rightarrow W?$	$I \rightarrow O^* : \{ \text{side effects} \}$	Focused intervention Broad observation
Design of a treatment of side effect	$H \models I \rightarrow O : \{ \text{side effects} \}$	$H \models I? \rightarrow W$	$I^*$	Focused design
Explanation of a drug effect	j. Exploring a drug effect $I \rightarrow O$ $H? \models I \rightarrow O$ $H^* \models I \rightarrow O$			Focused explanation
Exploration of a drug effect	$H \models I \rightarrow P$	$I \rightarrow O?$	$I \rightarrow O$	Focused prediction Focused observation
Prediction of a drug effect	$H \models I \rightarrow O$	$H \models I \rightarrow P?$	$H \models I \rightarrow P^*$	Focused prediction
Rational design of an experiment condition	$H \models I \rightarrow P$	$H \models I? \rightarrow W$	$I^*$	Focused design
Lab testing of a predicted drug effect	$H \models I \rightarrow P : \{ \text{pred. effect} \}$	$I \rightarrow O? : \{ \text{pred. effect?} \}$	$I \rightarrow O^*$	Focused intervention Focused observation
Data interpretation	$B \cup H \models I \rightarrow O$	$B \cup H \models I \rightarrow O$	$B^*, I^*, O^*$	Focused description

Table 9.4: Overview of the structure of discussed problems in drug research.

## 9.5 Reasoning

In the last section I described the structure of problems in neuropharmacology. We saw that different research activities lead to different kinds of discoveries. Intervention and observation can lead to new empirical discoveries about the natural world. We also saw that reasons to do a particular intervention or observation in a particular way or in a particular location are often suggested by conceptual discoveries that are the result of explanation, prediction and design.

I will now take a closer look at these three kinds of reasoning processes in neuropharmacology. I will discuss logical and computational models of those processes in problem solving that do not specifically aim to explain the cognitive processes that are involved when humans solve these problems, such as is aimed at with ACT-R models. That is a modeling task that requires a different approach. However both descriptive and normative models can share the initial assumptions (start), the goal condition and sometimes the result. In science, these can ideally all be described symbolically. Yet a psychological model will usually differ, compared to a normative computational model, in its modeling of the *process* of solving a problem. The description of the discovery cases in the last section could be a basis for both a psychological ACT-R model, as well as a problem solving model with other constraints.

In this section I will discuss several models of the reasoning processes in scientific problem solving, with a normative pretension. I will argue how these computational models of qualitative explanation, prediction and design could be used to aid the problems in the domain of neuropharmacology, by discussing examples based on the case study.

### Explanation and prediction in biology

To formally describe the process of reasoning in explanation and prediction in neuropharmacology I first discuss two computer models of Peter Karp's that model those processes. Karp investigated the development of knowledge about the biological process of attenuation (Karp, 1992; Karp, 1993).

He encoded intermediate states of knowledge about biological objects and processes so that his genetic simulator program GENSIM could use it to simulate experiments and compute predictions. The HYPGENE program takes these predictions as input and compares it with given observations made during an experiment. If there is a discrepancy, HYPGENE modifies assumptions about the initial objects or the processes to explain the difference between GENSIM's prediction and the observation.

Karp considers the process of hypothesis formation as employed by HYPGENE to be a design problem. In this way a hypothesis is an artifact to be synthesized and is subject to design constraints, such as among others consistency with the data, predictive success and simplicity. HYPGENE modifies a theory to satisfy constraints by implementing design rules. Karp derived these from his historical study of knowledge about attenuation, which on its turn provided a test bed for the development of HYPGENE and GENSIM.

Peter Karp's research goal was to model the competence of biologists, not their performance, by identifying reasoning mechanisms that are sufficient to solve hypothesis formation problems in biology, regardless whether they are valid psycho-

logically (Karp, 1992). To implement GENSIM and HYPGENE, he made use of effective tools for search control and an assumption based truth maintenance system.

The program GENSIM can make qualitative predictions about biochemical processes. Types of chemical objects are represented in a taxonomic hierarchy in a frame based Class Knowledge Base (C). Theories about biological processes like chemical reactions are represented as production rules in a Process Knowledge Base (T). The process rules define what classes of objects participate in a reaction and what conditions must be true for the reaction to occur. Process rules further specify what new objects are created if the rule's conditions are met.

GENSIM can predict the outcome of an experiment by applying the process rules to the given specified objects at the start. It is assumed that no objects are entirely consumed during a reaction, and therefore GENSIM only *adds* new objects monotonically to the initial ones. In this way GENSIM can predict what objects should result if the assumptions about initial conditions and the processes are correct. Because GENSIM implements a qualitative chemistry it does not make predictions concerning concentrations or reaction rates. Yet, it can predict increasing and decreasing quantities of chemical compounds.

The HYPGENE program can make qualitative explanations to account for the difference between predictions from GENSIM and observations. As input HYPGENE takes GENSIM's initial conditions  $I_a$ , *i.e.* the set of statements about objects that are assumed to be initially present in an experiment that is named "a", the predicted outcome  $P_a$ , *i.e.* the set of statements about the objects after the experiment, plus dependency information that records how  $P_a$  was computed from  $I_a$ , the prediction error  $Error_a$ , *i.e.* the difference between the prediction and the observation, and also access to all elements in the class knowledge base C and the process knowledge base T. In short, the input contains all elements from the tuple  $\langle I_a, P_a, Error_a, T, C \rangle$ .  $P_a$  is entailed by the initial conditions  $I_a$  and the processes in T:

$$I_a \cup T \models P_a$$

An experiment is anomalous when  $O_a$ , *i.e.* the set of statements about observation made in an empirical experiment, is not equal to the predicted observation  $P_a$ . The prediction error  $Error_a$  is by definition  $P_a \Delta O_a$ , *i.e.* the symmetric difference between prediction  $P_a$  and observation  $O_a$  (see also Figure 9.3):

$$P_a \Delta O_a := (P_a - O_a) \cup (O_a - P_a)$$

The main goal of HYPGENE is to eliminate  $Error_a$ . To achieve that goal HYPGENE reasons backwards from the difference between  $P_a$  and  $O_a$ . Its sub-goals become to remove statements about objects from  $P_a$  not in  $O_a$ , to modify assumptions about properties of objects in  $P_a$ , to modify assumptions about the quantity of objects in  $P_a$ , and to add assumptions about objects from  $O_a$  that were not in  $P_a$ . To achieve these sub-goals two main types of design operator are employed, those that redesign statements in  $I_a$  to  $I_a^*$ , and those that modify statements in T to  $T^*$  in such a way that:

$$I_a^* \cup T^* \models P_a^* \text{ and } P_a^* = O_a$$

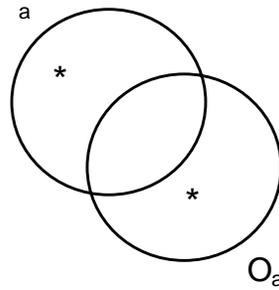


Figure 9.3: The symmetric difference between  $P_a$ , the set of statements about objects as predicted, and  $O_a$ , the set of statements about the observed objects.

HYPGENE examines the outstanding goals one by one, choosing operators that may satisfy it. For example regard the following simplified GENSIM prediction. For short a statement such as “x” means that there are objects of kind x present in the set-up of experiment  $a$ . Furthermore we have the following sets:  $I_a: \{x, y\}$ ,  $P_a: \{x, y, z\}$  and  $O_a: \{x, y, v\}$ . Say we have process rules  $r_1, r_2, r_3$  as part of  $T$  such that:

$$\begin{aligned} r_1: &= x \ \& \ y \rightarrow z \\ r_2: &= x \ \& \ z \rightarrow v \\ r_3: &= x \ \& \ w \rightarrow v \end{aligned}$$

For example, to remove the assumption about object  $z$  from  $P_a$  which is not in  $O_a$  a design rule can disable process  $R_1$  that causes the metabolism to an object of kind  $z$  by modifying its input, assuming that *e.g.*  $y$  was actually not an element of  $I_a$  and therefore also not in  $O_a$ . To explain that a statement about the observation of object  $v$  was present in  $O_a$  a design rule can modify process rule  $r_2$  assuming that object kind  $z$  is not necessary to cause  $v$ , or another operator can assume that  $w$  was also an element of  $I_a$ . So the problem becomes as follows:

#### GENSIM prediction

Start:  $I_a: \{x, y\}, T: \{r_1, r_2, r_3\}$

Goal:  $I_a \cup T \models P_a?$

Result:  $P_a: \{x, y, z, v\}$

#### HYPGENE explanation

Start:  $I_a: \{x, y\} \cup T \models P_a: \{x, y, z\}, O_a: \{x, v\}$

Goal:  $I_a? \cup T? \models O_a$

Result:  $I_a^*: \{x\} \cup T^*: \{r_1, r_2^*: = x \rightarrow v, r_3\} \models O_a$

The representations of objects in  $C$ , and the conditions and actions of the processes in  $T$ , are much more complex. For the details of the hierarchy of different design rules see (Karp, 1992). There are also operators that modify assumptions about quantity and about the structure of classes  $C$ .

It may seem odd to change  $I_a$ , but  $I_a$  only represents assumptions about what objects are present during an experiment. In biological practice knowledge of initial conditions is often uncertain because of the complexity of objects under study and the sometimes unpredictable effects of laboratory techniques. Karp found that it is normal practice in biology to first take a closer look at the assumed initial conditions, before changing hard earned theories. This practice was confirmed in the case of testing the basal ganglia model. Yet it is possible to slightly revise the model to explain the observed effect of dopamine agonists.

### Explanation of a drug effect

The HYPGENE and GENSIM programs model qualitative reasoning in biochemistry. So, they are able to model the process of explanation and prediction about transmitter metabolism in the brain. Yet, reasoning in neuropharmacology is also about increasing and decreasing values of variables. Reasoning about these aspects is better modeled by the QSIM program.

As we saw in section 9.3, given a qualitative differential equation, QSIM can make qualitative predictions about the behavior of a dynamical system. Richards et al (1994) devised a program that does the opposite. Given a qualitative description of the behavior of a system the program MISQ infers a qualitative differential equation that implies that behavior. I shall apply the techniques of this program to an example of the Parkinson case.

In the detailed discussion of the exploration of the effect of dopamine we saw that the basal ganglia model predicted an inhibitory effect in the SNR after a dopamine-agonists intervention in the striatum.

### QSIM prediction

Start:        B:  $\{f(\text{striatum}) = \langle 0, \text{std} \rangle\}, H_{BG}$   
 Goal?        B  $\cup H_{BG} \models I: \{a(\text{DA-agonist, striatum}) = \text{inc}\} \rightarrow P?$   
 Result:      P:  $\{f(\text{SNR}) = \text{dec}\}$

Yet under basal activation of the striatum the effect was not significant. The prediction error can be traced to constraints C.9 and C.10 of  $H_{BG}$ .

$$\text{C.9 } d/dt f(\text{striatum-D1-to-SNR}) = M^{++}(f(\text{striatum}), a(\text{DA, striatum}))$$

$$\text{C.10 } d/dt f(\text{striatum-D2-to-GPe}) = M^{+-}(f(\text{striatum}), a(\text{DA, striatum}))$$

In C.9 the activity of the direct pathway  $f(\text{striatum-D1-to-SNR})$  is positively dependent on the activity of the striatum  $f(\text{striatum})$  and the amount of dopamine  $a(\text{DA, striatum})$  that can act on the D1-receptor. So an increase of dopamine will cause the same amount of increase of activation of the direct pathway, regardless of the activation of the striatum. The same goes for the inhibition of the indirect pathway as defined in C.9. A program like MISQ is able to suggest different constraints that imply the observed values. The observed effects can be accounted for with a revision of C.9 and C.10 to C.9\* and C.10\*:

**MISQ explanation**

Start: B:  $\{f(\text{striatum}) = \langle 0, \text{std} \rangle\}$ ,  $H_{BG}$   
 I:  $\{a(\text{DA-agonist, striatum}) = \text{inc}\} \rightarrow O: \{f(\text{SNR}) = \text{std}\}$   
 Goal:  $H_{BG} \models I \rightarrow O$   
 Result:  $\{C.9^*, C.10^*\} \in H_{BG}^*$

C.9\*  $d/dt f(\text{striatum-D1-to-SNR}) = f(\text{striatum}) \times a(\text{DA, striatum})$

C.10\*  $d/dt f(\text{striatum-D2-to-GPe}) = f(\text{striatum}) / a(\text{DA, striatum})$

Now if there is only low basal activity of the striatum then DA will have a lot less effect than when the activity of the striatum is higher. Based on the revised hypothesis a new prediction can be deduced. An increase of activation of the striatum together with an increase of dopamine agonists now implies a stronger effect.

**Prediction of a drug effect**

A formal description of qualitative theories such as the basal ganglia model can also be useful in the research practice itself. The problem of the basal ganglia model, as noted in Chapter 3, is that it is too simple to be real and becomes too complex to work with when it would be extended to incorporate details.

The advantage of a formal description is that you can add more kinds of details, while you can still easily explore predictions by making use of a computer program like QSIM that easily computes the consequences for the variables you are interested in. As an example I list a number of computable predictions of different effects on the SNR after intervening in the direct and indirect pathways of the basal ganglia with selective dopaminergic agonists. Comparing these kinds of predictions with lab observations can result into more detailed and accurate models of biological structures such as of the basal ganglia.

**QSIM prediction**

Start: B:  $\{f(\text{striatum}) = \langle +, \text{std} \rangle\}$ ,  $H_{BG}$ , I  
 Goal:  $H_{BG} \models I \rightarrow P?$

**D1- agonists**

Start: I:  $\{f(\text{striatum-D1-to-SNR}) = \text{inc}, f(\text{striatum-D2-to-GPe}) = \text{inc}\}$   
 Result: P:  $\{a(\text{GABA, SNR}) = ?, a(\text{Glu, SNR}) = \text{inc}, f(\text{SNR}) = \text{inc}?\}$

Start: I:  $\{f(\text{striatum-D1-to-SNR}) = \text{std}, f(\text{striatum-D2-to-GPe}) = \text{inc}\}$   
 Result: P:  $\{a(\text{GABA, SNR}) = \text{dec}, a(\text{Glu, SNR}) = \text{inc}, f(\text{SNR}) = \text{inc}\}$

**D2-agonists**

Start: I:  $\{f(\text{striatum-D1-to-SNR}) = \text{dec}, f(\text{striatum-D2-to-GPe}) = \text{dec}\}$   
 Result: P:  $\{a(\text{GABA, SNR}) = ?, a(\text{Glu, SNR}) = \text{dec}, f(\text{SNR}) = ?\text{dec}\}$

Start: I:  $\{f(\text{striatum-D1-to-SNR}) = \text{dec}, f(\text{striatum-D2-to-GPe}) = \text{std}\}$   
 Result: P:  $\{a(\text{GABA, SNR}) = \text{dec}, a(\text{Glu, SNR}) = \text{std}, f(\text{SNR}) = \text{inc}\}$

**Ratios of D1 and D2 agonists:**

Start: I:  $\{f(\text{striatum-D1-to-SNR}) = \text{std}, f(\text{striatum-D2-to-GPe}) = \text{dec}\}$

Result: P:  $\{a(\text{GABA}, \text{SNR}) = \text{dec}, a(\text{Glu}, \text{SNR}) = \text{dec}, f(\text{SNR}) = ?\}$

Start: I:  $\{f(\text{striatum-D1-to-SNR}) = \text{inc}, f(\text{striatum-D2-to-GPe}) = \text{std}\}$

Result: P:  $\{a(\text{GABA}, \text{SNR}) = \text{inc}, a(\text{Glu}, \text{SNR}) = \text{std}, f(\text{SNR}) = \text{dec}\}$

**Rational drug design**

Another process that is an important part of pharmacology is design. Vos and Kuipers (1992) proposed that the development of design research in general can best be described as a more or less systematic attempt to bring together the properties of available materials and the demands derived from intended applications. They proposed a set-theoretic model of this process. In this model there is a set RP of all relevant properties for a product to be developed. A subset W of RP includes the wished-for properties of the intended product, so RP-W is the set of unwanted properties. For each possible prototype x that is created there is an operational profile, consisting of a set of operational properties O(x) that are part of RP, see Figure 9.4.

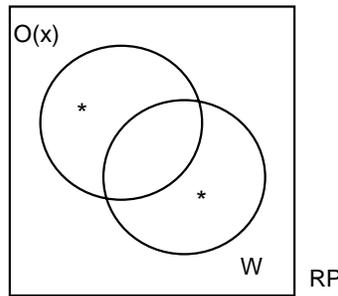


Figure 9.4: The symmetric difference between operational properties O(x) of prototype x, and the set of wished-for properties W, both part of relevant properties RP for the object to be developed.

A problem-state during development can now be described as the symmetric difference  $W \Delta O(x)$ , defined by the set of unrealized wanted properties together with the set of realized properties that are not wanted, *i.e.*:

$$W \Delta O(x) := (W - O(x)) \cup (O(x) - W)$$

$W \Delta O(x)$  denotes the set of problems, *i.e.* the qualitative deviation of O(x) from W. The number of problems to be solved is defined as  $|W \Delta O(x)|$ , indicating the quantitative deviation.

The goal of design research is to develop a better product  $x^*$  such that ideally  $O(x^*)$  is closer to W. Kuipers & Vos propose a descriptive model of the transitions of problem states. This model defines a proper assessment criterion for the improvement of problem state transitions. Prototype  $x^*$  is an improvement of x in view of W iff:

$$O(x^*) \Delta W \text{ is a proper subset of } O(x) \Delta W$$

So a new prototype is an improvement only if it has at least one wished-for property more or at least one unwished-for property less.

For most design research it is possible to divide the set of relevant properties into two complementary sets of structural and functional properties  $S$  and  $F$ . Often first a functional profile  $WF$  is determined of what the product is supposed to do. The next question is how this can be realized. Yet a product is not uniquely determined by  $WF$ , often functional equivalents are possible, so the set of looked for structural properties is called an appropriate structural profile  $AS$  for  $WF$  if it causally implies the desired functional properties  $WF$ .

In drug research the determination of the desired functionality  $WF$  is normally guided by known characteristics of a disease. These can be explicated as part of the set of conceivable characteristics of potential applications  $C(A)$  of a drug, see Figure 9.5. We can say that for a given disease  $y$  its profile  $C(y)$  uniquely determines the desired functional profile  $WF$ , while the reverse is not the case. A drug for a given characteristic can be useful for each disease containing that characteristic. An improvement of a drug's structure and functionality can be defined analogous to the definition above.

For example, in the case of Parkinson's disease, the characteristics of the pathological condition  $C(\text{Parkinson's disease})$  includes a degeneration of dopaminergic neurons in the substantia nigra, and a subsequent depletion of the neurotransmitter dopamine in that area. For symptomatic treatment it is suggested that a drug is found that replaces the function of dopamine in that area ( $WF$ ). This should be an effect caused by a drug with an appropriate structure ( $AS$ ). How the properties  $AS$  of a drug might be suggested by  $WF$  is extensively discussed in Vos (1991).

In the next subsection I will discuss how the functional properties  $WF$  for a drug intervention can be rationally inferred given characteristics of a disease  $C(y)$  and (a formal description of) an available biological theory.

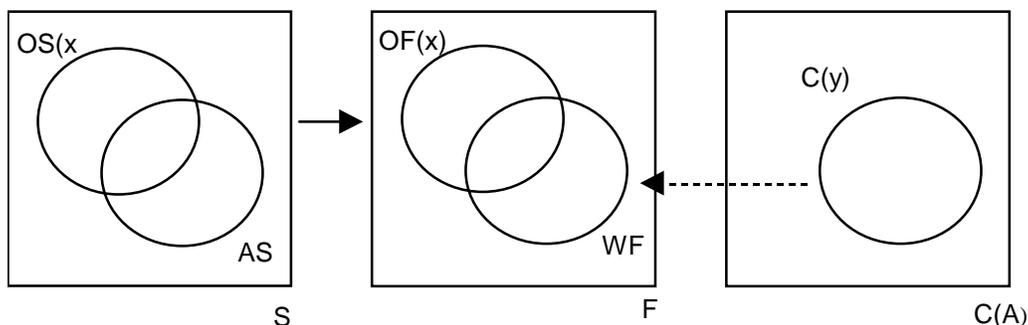


Figure 9.5: A problem state in the S/FA model of drug design research. The disease characteristics  $C(y)$  determine the wished-for biochemical effect  $WF$  that is caused by a looked for drug with appropriate structure  $AS$ .  $OS(x)$  and  $OF(x)$  are the operational structural and functional profile of prototype drug  $x$ .

### Rational drug treatment design

In section 9.4 we saw that in the rational design of a drug treatment knowledge of biological processes is used to infer the effect of a drug intervention. The suggested intervention can either contain a description of the desired local influence of a drug on the system, or a description of a drug that is known to have the needed functional properties. These desired properties of the drug should cause a decrease in disease symptoms, and are called a drug lead (Vos, 1991). The rational search for a drug lead can be understood as a problem of qualitative reasoning. Knowledge of qualitative relations between variables describing properties of a pathological biological system can be sufficient to find variables that can influence that system.

The search involved is structurally similar to that of explanatory reasoning, but has a different search goal. Instead of finding a simple hypothesis that explains an observed behavior, the task is to find a minimal intervention that has a desired effect on properties such as the behavior of the system, with minimal side effects. So, analogously to inference to the best explanation, this process can be called “inference to the best intervention”.

The object of drug treatment design does not initially concern the properties of a compound as in drug design, but the properties of a biological system, an organism. In the latter the goal is to create a drug so that it has given desired properties, in the former the goal is to create the behavior of a biological system so that it has given desired properties. These properties can also be divided in structural and functional properties. A disease is a set of unwanted properties of a biological system. These can be compared with wished for properties of a system. So we can define the characteristics of a disease as follows.

**Definition 11** *Disease characteristics.* Given the operational properties  $O(x)$  of a pathological system  $x$  and the wished for properties  $W$ , the characteristics  $C(y)$  of a disease  $y$  can be defined as the symmetric difference between  $O(x)$  and  $W$ :

$$C(y) := W \Delta O(x)$$

The set  $O(x)$  contains all the considered properties of a system  $x$ , not only the pathological properties. So the set  $W \cap O(x)$  is not empty. The goal of drug treatment is to change the properties  $O(x)$  of system  $x$  to  $O^*(x)$  such that both  $O^*(x) - W$  and  $W - O^*(x)$  are minimized

Rational drug treatment design involves finding a drug treatment for a given pathological condition of a system by maximally employing known theories and knowledge about biological processes. A proper theory about a disease should be able to explain the pathological properties.

So, let a set  $H$  of theories about biological processes be given as well as background assumptions  $B(x)$  involved in the explanation of the observed properties among the properties  $O(x)$  of a pathological system  $x$ . The problem of the design of a drug treatment of the pathological properties  $O(x) \Delta W$  is to cause only wished for properties from  $W$  by a drug intervention  $I(x)$  of the system, *i.e.*  $H \cup B(x) \models I(x) \rightarrow W$ . If we can explain the pathological condition, then we can use that knowledge to infer a suitable intervention.

**Rational drug treatment design**

Start :  $H \cup B(x) \models O(x)$

Goal :  $H \cup B(x) \models I? \rightarrow W$

Result :  $I^*(x)$

The search goal is to find, by reasoning about processes in  $H$ , a proper drug intervention that influences processes that cause the desired properties  $W$ , but not those from  $O(x) - W$ . That is, the goal is to eliminate the difference between  $W$  and  $O(x)$ . The result of the search is the suggestion of a manipulation of a local biochemical property that can be affected by a drug. A drug that has this wished for functional effect (WF) can be searched for in the set of known drugs, or pose a new problem for rational drug design.

Of course it would be ideal, given the known  $H$  and the nature of the disease, to infer a suggestion for a drug intervention  $I$  that only causes  $W$ . A drug usually also causes side effects, often creating undesired effects that are not part of the disease that is targeted. Therefore we need a gradual evaluation criterion for the improvement of suggestions (*cf.* T.A.F. Kuipers, Vos en Sie 1992). Let us say that the moderated design goal is to find the suggestion  $I$  such that its (predicted) consequence for a system  $H \cup B(x) \models I(x) \rightarrow P(x)$  resembles the desired condition  $W$  more than the pathological condition  $O(x)$ , *i.e.* that:

$P(x) \Delta W$  is a proper subset of  $O(x) \Delta W$

That is, roughly, the drug should not have more unwanted consequences than accomplished desired consequences, *cf.* Figure 9.6.

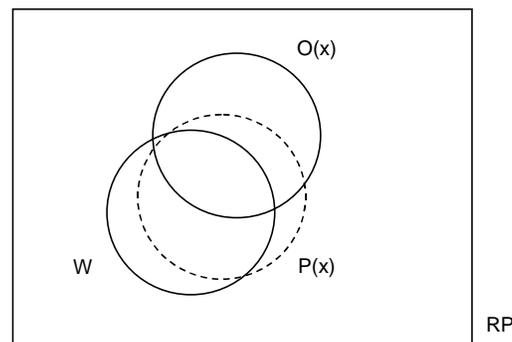


Figure 9.6: Problem state in searching an intervention with effect  $P(x)$  that most resembles desired properties  $W$  in treating a pathological system  $x$  with operational properties  $O(x)$ .

The evaluation of improvement of more than one drug suggestion can follow the same lines. A drug intervention  $I^*$  of  $x$  is better than an intervention  $I$  if the properties of consequence  $P^*$  resemble  $W$  more than those of  $P$ :

$P^*(x) \Delta W$  is a proper subset of  $P(x) \Delta W$

However, this is only an evaluation of properties that is neutral to the different kinds of undesired properties. In this way an intervention could be inferred that treats most of the symptoms, but causes a symptom that is worse than the disease that is treated. This could be remedied by a ordering of the undesired properties, together with a quantitative measure of deviation.

The resulting suggestion for a drug intervention can on its turn be used to test the theories used to find the suggestion. Given an inferred drug intervention  $I(x)$ , an experiment can be done and its resulting observation of the altered operational properties  $O(x)$  of  $x$  can be compared with the predicted properties  $P(x)$ . A discrepancy can be used to redesign  $H$ , or the assumptions about  $B(x)$  or  $I(x)$ .

### Testing of predicted drug effect

Start:  $H \cup B(x) \models I(x) \rightarrow P(x)$

Goal:  $I \rightarrow O?$

Result:  $I \rightarrow O^*$

The same kind of design and testing is found in designing experimental conditions for focused testing of hypotheses, as we saw at the end of section 9.4.

### Rational design of an experimental condition

Start :  $H \cup B \models I \rightarrow P$

Goal :  $H \cup B \models I? \rightarrow W$

Result :  $I^*$

### Computational drug lead discovery

Next to rational drug design in the lab, we also saw in section 9.4 that there is a sub-discipline in pharmacology called computational drug design. This discipline is concerned with the rational design and exploration of drug structures and drug function, making use of computational models of those structures. This is usually a quantitative approach, making massive computations in quantum mechanics to predict *e.g.* the folding of the protein structure of a receptor in reaction to a drug structure.

To search for a drug treatment there are many kinds of computer programs that can help to diagnose a particular disease and suggest a drug treatment. These programs make use of explicitly known established assumptions about pathology and medicine. These kinds of tools are less known in the practice of basic research.

An exception is the ARROWSMITH program of Swanson and Smallheimer (1997). This program searches for unknown relations between research findings in the literature. One research group the members of which know each other's writings may establish that there is a connection between biological properties  $A$  and  $B$ , while another group in a slightly different field could have established a link between  $B$  and  $C$ , but may be unaware of the other group's results. If  $C$  is related to a pathological property then  $A$  might be a lead candidate for drug treatment. The ARROWSMITH program searches for these implicit links in the published texts that describe results, making use of different statistical techniques.

A search by ARROWSMITH discovered a link between fish oil ( $A$ ) and Reynaud's disease ( $C$ ). Both are related to properties of blood viscosity, platelet aggregation, and vascular reactivity ( $B$ ). The program also discovered a relation be-

tween magnesium and migraine, they share 11 related properties. Weeber *et al* (2000b) develop techniques for the same kind of problems in order to find out whether published side effects of drugs may be beneficial for the treatment of other diseases.

### Textual drug lead discovery

Start:        ...,  $A \rightarrow B$ , ...,  $B \rightarrow C$ , ...  
 Goal:         $? \rightarrow C$   
 Result:       $A^*$

This search for a novel treatment is conducted in the enormous amount of published results, represented by the dots in the above scheme. A discovered relation is a discovery of an implicit conceptual relation in explicitly known results. These results contain explicit descriptions of interventions, observations, explanations and predictions. Such searches are fruitful, but they can not find implicit consequences of proposed explanations and theories. This is a hard problem because a description of a theory in natural or informal language, as is common in medicine, is difficult to analyze semantically. It is possible to cluster words that are related in meaning, but with current techniques it is not possible to computationally infer logical consequences from sentences in natural language. These techniques can assist in the discovery of textual relations, but they still have to be interpreted by a knowledgeable scientist.

A more formal description of both qualitative and quantitative results can result in computational discoveries of new interesting consequences, for both basic research and treatment.

### Logical drug lead discovery

Start:         $H \models C$   
 Goal:         $H \models ? \rightarrow C$   
 Result:       $H \models A^* \rightarrow C$

For example the formal description of the basal ganglia, incorporating more research details can be used to computationally explore interesting predictions, as we saw in this section, and to search interventions with desired properties in detail. The desired state  $W$  that should be caused by an intervention aimed to combat Parkinson's disease includes  $\{f(\text{SNR}) = \text{dec}\}$ . A search program can infer that this can be caused by an increase of GABA or a decrease of glutamate:  $\{a(\text{GABA}, \text{SNR}) = \text{inc}, a(\text{Glu}, \text{SNR}) = \text{dec}\}$ . These variables can on their turn be influenced by different interventions with selective dopamine agonist. The effects of those were predicted earlier in this section. We already saw in section 9.3 how the model implied the traditional treatment of targeting the metabolism of dopamine.

### Computational drug lead discovery

Start:         $H_{\text{BG}}$   
 Goal:         $\text{QSIM}(H, I?) = W$   
 Result:       $I^*$

In this search not only one looks for which variables are related, but also one tries to find out how the values of these variables influence each other. It is one thing to know that dopamine and Parkinson symptoms are related, but it is a more specific hypothesis that the decrease of the amount of dopamine is related to the increase of symptoms. With this knowledge available in QDEs one may better evaluate possible interventions in a system. If an intervention causes a variable to be steady while you wished that it would increase, then that result is not as bad as when it would decrease, *cf.* Table 9.5. This evaluation can be extended with specific weights for particular properties, depending on the disease and the importance of particular properties.

P \ W	dec	std	inc
dec	1	0	-1
std	0	1	0
inc	-1	0	1

Table 9.5: Quantitative evaluation of a predicted qualitative value of a variable from the set of predicted properties P, compared with the desired value of the same variable in the set of wished-for properties W.

In this way all variables under consideration can be evaluated to find the best intervention, for which the sum of the evaluations of all value comparisons should be maximal. This evaluation can also be applied to the other structurally similar rational design problems I discussed.

## 9.6 Conclusion

So, what is the rational use of theory and experiment in the process of scientific discovery in the practice of drug research for Parkinson's disease, compared to theory as discussed in part II? In Part II we saw that in the theoretical conception of scientific discovery the rational use of a theory in scientific discovery is to explain observations with simple additional hypotheses that can predict properties of phenomena. The rational use of experiment is to test predictions of those explanations. This use of theory and experiment is considered to be what makes the process of scientific discovery rational.

After analyzing a practice of scientific discovery in detail it is apparent that both theory and experiments are used for many more reasons, with different goals and results, all leading to different kinds of scientific discoveries, as summarized in Table 9.4. Theories are used to explain observations, and to make predictions that test the theory. But they are also used to rationally design experimental conditions and treatments, and to explore phenomena. Experiments are used to test theories by observing and intervening in properties of phenomena that are predicted by that theory. But they are also used in treatment and in many different kinds of exploration, often intervening in ways and having one look in to directions that are not suggested by theory, but just by curiosity. In this case the rational to use experiment to explore those areas is that there *is* no theory or expectation about it, giving ample room for new empirical discoveries.

Yet devising and testing theory remains an important part of science, both in theory and in practice. The use and nature of theories in scientific practice is grounded in primary and secondary cognitive mechanisms as explicated in Chapter 5. Natural language and informal diagrams are the vehicles of choice to represent assumptions in the scientific discipline I analyzed. Yet in this way it is not always possible to fully oversee the consequences of known theories and results. To better understand these, epistemologists devise formal theories about theories. These can be used to represent a theory more explicitly. In the case study, I showed how this can be done for the theoretical model of the basal ganglia.

So, it is now time to answer the specific questions of this thesis for my case study:

**Question 1** The theory of the basal ganglia consists of qualitative relations between variables of chemical and electrical neural activity in nuclei. The structure of this theory can be represented as a qualitative differential equation. In a structuralist approach the theory can be defined by its models, given a set of constraints on conceptually possible models defined by a set of variables and possible values.

**Question 2** When a drug intervention is observed together with a certain change in a property of a biological system, a conditional dependency can be inferred. Clusters of observed or assumed relations between variables that describe the domain can together become a theory that explains other relations. A goal in neuropharmacology is to infer a hypothesis  $H$  that best explains how observed properties  $O$  of a biological system are conditionally dependent on an intervention  $I$ , *i.e.* to infer the best explanation (IBE). This hypothesis can be used to rationally design a drug treatment or a condition for an experiment, to cause wanted properties  $W$ , *i.e.* to infer the best intervention (IBI). Given a hypothesis and an intervention new consequences  $P$  can be predicted. If the goal is to test the theory, the goal of the reasoning process is to infer the best prediction (IBP), see Table 9.6.

Problem	Start	Background	Process	Goal	Goal properties
Explanation	$I \rightarrow O$	$B, V, Q, D$	IBE	$H^*$	$B \cup H^* \models I \rightarrow O$
Design	$W$	$B, H: \langle V, Q, C, D \rangle$	IBI	$I^*$	$B \cup H \models I^* \rightarrow W$
Prediction	$I$	$B, H: \langle V, Q, C, D \rangle$	IBP	$P^*$	$B \cup H \models I \rightarrow P^*$

Table 9.6: Main processes of reasoning discussed in this part.

Finding a hypothesis  $H^*$ , an intervention  $I^*$  or a prediction  $P^*$  that is optimal given the assumed conditions, can be called a conceptual discovery and is often no trivial problem. It may require an exhaustive search in a problem space that is defined out of known concepts in  $V$  and  $Q$ . Within that problem space a hypothesis  $H$  may be found that is denoted by the set of constraints  $C$  on all the possible models of a domain, as determined by  $V$ ,  $Q$  and  $D$ . An intervention and prediction are rationally searched for within the models allowed by the constraints of the hypothesis. Finding a proper hypothesis for a domain may also require a conceptual revision of the problem space, by revising the variables in  $V$  and quantity spaces in  $Q$ . In contrast to a conceptual discovery, making an experimental intervention and observation can lead to an empirical discovery, when new properties or phenomena are actually created or observed.

**Question 3** In this chapter I have discussed many different routes between theory and experiment. Particular interventions and observations can lead to new empirical discoveries when the observed properties are not expected, or prove an expectation to be wrong. Such an empirical discovery is able to logically refute a theory, but in practice it will not be deserted. A false theory can remain a fruitful pointer to directions for new interventions and observations that can lead to new empirical discoveries. In biological practice explanations are revised to fit observations, looking first at the assumptions about the interventions and observations and in the background.

In contrast to the discussed diversity in the discovery process in the practice of neuropharmacology, I end my discussion of the questions of this thesis with a formal summary of the textbook example of the process of discovery in drug research for Parkinson's disease:

1. Observe phenomenon p:  $p_i, \dots, p_j$  (parts of the basal ganglia)
2. Describe p:  $I \rightarrow O$   
 $I: \{a(\text{DA, striatum}) = \text{dec}\}$   
 $I \rightarrow O: \{f(\text{SNR}) = \text{inc}\}$
3. Explain p:  $B \cup H? \models I \rightarrow P$   
 $H_{\text{BG}}^*: \langle V, Q, C, D \rangle$
4. Predict p:  $B \cup H_{\text{BG}} \models I \rightarrow P?$   
 $I: \{a(\text{DA, striatum}) = \text{inc}\} \rightarrow P^*: \{f(\text{SNR}) = \text{dec}\}$   
 Design p:  $B \cup H_{\text{BG}} \models I? \rightarrow W$   
 $I^*: \{a(\text{D}_{1/2}\text{-agonist, striatum}) = \text{inc}\} \rightarrow W: \{f(\text{SNR}) = \text{dec}\}$
5. Intervene p: do I
6. Observe p: see P?

In step 1. processes and properties of the basal ganglia are observed. It is described how a decrease of dopamine in the striatum by an intervention results into an increase of activity of the SNR. A model of the basal ganglia is proposed that implies the observation in step 3. In step 4. this model is used to predict that an increase of dopamine in the striatum will cause a decrease of activation of the SNR. Given the decrease of SNR activation as a wished-for property, the model also implies other possible interventions, such as agonists for a receptor-subtype. These suggestion can be experimentally tested in steps 5. and 6.

This process can be aided in both theory and practice with the use of computer modeling tools that can assist in finding descriptions, explanations, predictions, and new designs. However, the bigger problem to make these tools useful is the availability of biological theory in a formal representation. It would be ideal if scientists in biology would publish their results both in natural language and in a formal format. To this end, Peter Karp started an internet database that invites biologists to add their results in a provided formal format. This database is used to test new methods that can aid the process of discovery in science, aiding on its turn the process of understanding rationality in discovery.