

Reports from the Department of Philosophy

Vol. 28

# **Will and Mechanisms**

**Mikael Melan**

University of Turku

Finland

Copyright © 2013 Mikael Melan

SERIES EDITORS:

Olli Koistinen

Juha Räikkä

Department of Philosophy

University of Turku

FIN-20014 Turku

Finland

ISSN 1457-9332

ISBN 978-951-29-5369-1 (print)

ISBN 978-951-29-5370-7 (pdf)

Juvenes Print – Suomen Yliopistopaino Oy, Turku 2013

## Acknowledgements

First and foremost, I want to thank my supervisor, Olli Koistinen, who merits as much credit for this work as I do. His belief in me was something that truly changed my life and only little by little I have begun to share his belief. Simply put, without his wisdom and encouragement this study would never have come about. Moreover, even with all the philosophical disputes we have had, since the beginning our discussions have been the foundation of my independent philosophical work. The pain of convincing him about my arguments has driven me to shape my thoughts to an extent I would never otherwise have reached.

These four years I have spent working in the Department of Philosophy at the University of Turku have been very happy ones, even with all the hard work it has demanded. This is mostly so because of my colleagues who have created a working environment which I believe is highly unique in its warmth. Thus, I want to express my gratitude to all those who have been around during these years. Workwise I need to thank especially Tapio Korte, Eerik Lagerspetz, Hemmo Laiho, Arto Repo and Valtteri Viljanen for their help in the process of shaping the original form of my dissertation. Additionally, besides giving valuable comments on my work and help in my struggles within the academic world, Rosa Rantanen and Susanne Uusitalo have been simply priceless in their support within all sectors of life. Without them, I would not have managed through all the stress, pressure and anxiety which I believe are, sad to say, the inevitable constituents of the life of a young researcher in this insecure line of work. Hanna-Mari Salonen also deserves my special gratitude, time after time giving her help in practical matters and bureaucracy, which seem to be there only to make the living of us theoreticians intolerable.

This dissertation is founded on my licentiate thesis which I defended at the end of 2011. My two opponents, namely Arto Laitinen (University of Jyväskylä) and Kari Lukka (Turku School of Economics), have both been of enormous help in their challenging questions and constructive words of advice. It is my privilege that the conversations have continued after the ceremony, and thereby both my opponents have played an important role in the process of improving my account. Arto Laitinen also served as a referee of this work with Kristina Rolin (University of Helsinki) and they have both provided highly valuable

comments and suggestions for which I am very obliged. Furthermore, Osmo Kivinen (University of Turku) has been of great assistance with the questions about social sciences. For language consultation, I would like to thank Tim Glogan.

In 2011 I spent the spring at the University of Southern California, which was both professionally and personally a valuable experience. For that, I am greatly indebted especially to Gary Watson who made it all happen. His kindness made my stay there pleasant as his exceptional philosophical awareness was of huge help in my work. Naturally, I also want to show my gratitude to the whole Department of Philosophy in USC. Finally, I must apologize to all those who I have not mentioned here but who have nonetheless had an influence on my work. I have been lucky to receive great help and clever comments from numerous people at various conferences and in private conversations all around, and since it would be impossible to acknowledge my gratitude to all of them here, I surely hope I remembered to do it at the time.

For making this study possible in the monetary sense I must thank the IASM (Institutions and Social Mechanism) and COREHES (Comparative Research on Higher Education and Science Institutions) research schools as well as the Turku University Foundation.

On a personal note, my deepest gratitude belongs to my mother Helena Tiheäsalo without whom I would not be here, not only because of the birth she gave to me, but of the continuous love and support she has given me even through the hardest times. I hope this achievement compensates for at least a little of the trouble she has had to go through because of me. Of course, my gratitude also goes to the rest of my family of whom I still want to highlight my Uncle Hannu and my already departed grandfather Toivo Tiheäsalo for always being there when needed. I am also blessed with an enormous number of supportive and tolerant friends of whom it is impossible to name any particular ones. To these Ladies and Gentlemen, to the ones who are still around as well as to the ones I have already had to let go, I want to express my deepest gratitude and love. While finishing this study I have been reading Ludwig Wittgenstein's diaries (12.10.31) where I found a few lines that have haunted me ever since. I simply cannot resist the temptation to repeat two of them here: "I am for example a petty, lying rogue & yet can talk about the grandest things./ And while I am doing that, I seem to myself perfectly detached from my pettiness. Yet I am not."

For FC TPS,  
my dearest

## Contents:

<b>Introduction: A Man on a Bus -Part I .....</b>	<b>9</b>
---	----------

---

### **PART I: THE FOUNDATIONS**

<b>1. Explaining with Mechanisms .....</b>	<b>16</b>
1.1 Hempel's Idea of Scientific Explanation .....	16
1.2 About Causality.....	21
1.2.1 Hume's Problem.....	21
1.2.2 Living in the World of Properties.....	26
1.2.3 A Place for Causality.....	33
1.2.4 Causality and Context.....	37
1.2.5 Causality and Explanation.....	43
1.3 The Need for Mechanisms .....	45
1.3.1 The Problematic Nature of Laws.....	45
1.3.2 The Concept of Mechanism.....	50
1.3.3 Explaining with Mechanisms .....	53
<b>2. The Multitude of Actions.....</b>	<b>58</b>
2.1 What Actions Are? .....	58
2.2 Voluntary Actions .....	61

2.2.1 Aristotle and the Origins of Action .....	61
2.2.2 Explaining Actions with Desires and Belief .....	65
2.2.3 The Obscure Nature of Mental States .....	70
2.2.4 Unorthodox Agents .....	77
2.3 Intentional Actions .....	83
2.3.1 Aristotle and the Teleological Side of Actions.....	83
2.3.2 Two Aspects of Reason for Action .....	88
2.3.3 About Normativity in Intentional Actions.....	95
2.3.4 The Concept of Will .....	101

## **PART II: THE APPLICATION**

<b>3. Actions in Scientific Mechanisms .....</b>	<b>107</b>
3.1 Two Types of Mechanisms .....	107
3.2 The Mechanisms of Natural Sciences .....	110
3.2.1 The Idea of Mechanical Philosophy .....	110
3.2.2 Determinism <i>versus</i> Indeterminism.....	116
3.2.3 Actions and Agents in the Physical Domain .....	122
3.3 The Mechanisms of Social Sciences .....	129
3.3.1 The Role of Social Sciences .....	129
3.3.2 About Social Mechanisms.....	135
3.3.3 Actions in Social Mechanisms .....	141
3.3.4 The Two Senses of Intentionality.....	145
3.3.5 Selves and Wills .....	152
3.3.6 Changing the Perspective .....	161

---

<b>Conclusions: A Man on a Bus -Part II.....</b>	<b>168</b>
<b>Bibliography .....</b>	<b>177</b>



## **Introduction:**

### **A Man on a Bus -Part I**

This whole study is rooted in one of the discussions I had with my supervisor, Professor Olli Koistinen, a few years back. For some reason, we were having this conversation about scientific explanations and predictions for human action. At some point, Koistinen started a thought experiment about a man sitting on a bus. To build the case, we imagined a bunch of scientists from whatever field making a perfectly accurate prediction about the phenomenon. Now, the question was what it means, in the end, that this perfectly accurate prediction could be given. Clearly, the fact that there is such a prediction does not remove the fact that this man nonetheless has to make his way to the bus. In other words, the prediction does not in some magical way carry this man to the bus in order to fulfill itself. Quite the contrary, with or without any predictions this man still needs to take himself to the bus so that the prediction can be called accurate. The same goes for explanations. Let us imagine there being a complete causal explanation for the same phenomenon so that this explanation does not refer to any such factors as, say, an agent or decision *per se*. For instance, the explanation in question may be a neurophysiological one. Now, even if this explanation is complete in a sense that there is nothing more needed to explain the phenomenon, it still does not remove the facts that the man must make the decision of going to the bus and, without that very decision, he would not be sitting there.

This story has stuck in my mind ever since. The seemingly innocent thought experiment started to bother me little by little more and more, and after a while I was finally able to see the whole depth of the issue. I understood that there was easily enough work on this question for one doctoral dissertation. As should be the case with all philosophical problems, the basic issue behind the little story about the man sitting on the bus keeping the scientists busy is rather simple. It is merely a collision between two contradictory images of the same phenomenon. Nevertheless, it eventually took me almost 200 pages to unravel it in a somewhat satisfactory way. The problem seemed to be one especially attached to human action and thus it does not arise to the same extent while we have our eyes on events taking

place in non-human nature. To put it simply, I realized that the source of the problem is our ability of reflection. If we are, for instance, investigating an earthquake, the basic structure of the process is quite unproblematic. After all, we merely study the phenomenon from outside, trying to make as much sense of it as possible. However, if we are now to shift our interest to some action that we ourselves are performing, we end up with a bipartite situation. At the same time, we are the one who is studying and the one who is the object of the study. It should be clear that this kind of tension is far from trouble-free.

If we are to study our own actions, it seems that we are able to do it from two different perspectives. Firstly, we may do it from the subjective one, namely that of the agent. This is the view we have while performing an action. Secondly, we are able to, figuratively, step out of ourselves, by which I mean that we may take the objective standpoint towards our own actions as well. Now, the problem is not in our unique skill of applying different kinds of perspectives *per se* but in the images they sketch. From the first perspective, I seem to be a free agent making genuine decisions but, from the second one, I seem to be only a mere part of a world where every event is a part of some larger course of events. For example, let us imagine an alcoholic who decides once again to go to the bar down the street. Evidently, this poor man considers himself a free agent while he puts on his shoes after a long while of contemplating what to do. Either way, it is easy to imagine a further situation in which, after many days of drinking, this same man explains his actions to his friend by crying that he has this ulcer which he cannot control and which drives him to the bar every night. Now, what has happened here is that the man has changed his perspective from the subjective to the objective, even though he is picturing the same action all the time. From the first perspective, he is a free agent performing an action after a meaningful contemplation whereas from another one he is merely driven by his urge to drink, which is there because of his drinking habits of which the causes may again be traced to various antecedent events and conditions. The tension between the two perspectives is there to see with ease. Anyway, the problem does not lie solely in situations in which we contemplate our own actions. If we are to return to our original example about the man on the bus, we are able to see that both these two perspectives can be applied even from outside. We are undeniably able to consider this man as a free agent deciding to take a bus as well as we are able to see him through the eyes of the scientists trying to explain his action by referring to underlying causal factors of some different kind.

Of course, there is nothing novel in the basic philosophical problem of there being two conflicting images for human action. For instance, St. Augustine of Hippo (354–430 A.D.) (*On Free Choice*) already struggled with the tension between free will and divine foreknowledge which both were indispensable, yet seemingly contradictory parts of the Christian teachings. After the scientific revolution which happened around the 16th and 17th centuries, the deterministic world view became the counterpart to divine foreknowledge. Evidently, the tension between determinism and free will has been one of the big issues in philosophy as, for ages, philosophers have been trying to deny the existence of one or another or, alternatively, to find a way for their harmonious coexistence. However, what I will argue in this thesis is that, with or without determinism, there is a problem between two conflicting images, both of which are indispensable parts of our experience of the world but which still really are contradictory to the extent that there cannot be any possibility of fitting them into the same picture. After all, we are able to see ourselves both as mere parts of ever continuous causal chains in the world and as genuine agents making a true impact on the world, and these two images evidently contradict one another. Now, in my view, there is nothing wrong in the contradictory images the perspectives provide as long as we understand to keep them separate, and thus one important aim for us as philosophers is to find justification for them and, most importantly, for the application of the different perspectives. Interestingly enough, Immanuel Kant (1724–1804) (*Idea*) already saw things in a rather similar way. Perhaps he was the first one to admit that neither one of the images could be denied, yet that both cannot coexist at the same time in the same bigger picture. In my view, his extremely complicated ontological structure may be taken as one solution to the contradiction, but in all its abstruseness and forbidding ontological commitments, it is hard to take as an ideal one even with its uncontested merits. Later Thomas Nagel (1986) did highly valuable work in redefining the problem of free will, among others, as a problem of two perspectives. Actually, reading his writings truly opened my eyes to the matter. In the end, I guess I will only be following his footsteps as I believe that the importance of his findings is far from fully appreciated. There is a lot of work to be done to make the most of the discovery. The problem between free will and determinism is only a minor piece of a much bigger whole.

Naturally, the title of my work, “Will and Mechanisms”, refers to the very tension between the two perspectives. Here the collision is seen, not between free will and determinism, but between true agency and explanatory mechanisms for human action. As emerged in the example of the man on the bus, they are after all the explanations for actions

which contradict with each other. For instance, explaining something with causally effective neural factors is something very different to referring to agent's decisions. When it comes to predictions on the other hand, they again require explanations in order to meet the terms of science. A prediction without a scientifically indissoluble foundation is nothing more than a mere prophecy. Either way, as was the case with divine foreknowledge, workable predictions seem to contradict free choices. So, all in all, what is at the core of the problem is the possibility to explain actions in different and even contradicting ways. Of all kinds of scientific explanations there are various reasons why I have picked mechanisms at stake. First, for a while explanatory mechanisms have been the subject of huge interest in the philosophy of science. 'Mechanism' means an explanatory model which can be used both in natural as well as in social sciences. As they are at the top of the discussion about adequate models for scientific explaining, I believe that all the research done in this very field is undoubtedly beneficial. Besides, taking into account all the work done around theoretical issues in my study, it is highly worthwhile binding the results to something more concrete and more practical as well. Second, as the concept of mechanism suggests, there we have a perfect antithesis for the idea of freedom. Even if scientific mechanisms are usually much more flexible structures than the mechanical devices we have, like locks and clocks, the word gives us an impression of a more or less strict construction which works in the way of a machine. The whole idea of mechanical motion seems to be quite far from fitting into our picture of ourselves as autonomous agents, regardless of whether the mechanism is associated with something which produces our actions or something where we act as a part. Anyway, my purpose with this study is, nevertheless, not only to figure out whether there is a possibility for true agency to coexist with the mechanism-based view of human action. On the contrary, as I believe that this problem has its roots deeper in the tension between the two perspectives, my main aim is to achieve a proper understanding of these perspectives and to strip down the images they provide. What is important is to justify and be aware of these perspectives as well as to clarify the concepts we are using while applying them. After all, it comes across as various philosophical problems arise only from our confusion about the changing perspectives and of the varying meanings the concepts share while we shift from one to another. The only appropriate method of avoiding and clearing up these confusions is to carry out a profound analysis of the possible ways of understanding action and agency as well as of the roles they play in changing perspectives.

My study consists of two parts so that, crudely, in the first one my aim is to lay the theoretical foundation for the main questions answered in the second one. As the two chapters of the first part are divided in such a way that the first one is mainly about the philosophy of science and the second one about the philosophy of action, in the third chapter, which is the same as the whole second part, I will combine these teachings and see where we have arrived. Next I will introduce the contents of these chapters in a little more detail.

As this study is a great deal about scientific mechanisms, the aim of the first chapter is to demonstrate their idea as well as their importance in the modern philosophy of science. The best way to do this is to return to the origins of the discussion about the ideal scientific explanation, since only through the understanding of the faults of the preceding alternatives are we able to see the merits of mechanisms. However, during the process we will learn that for a proper understanding it is impossible to escape even the most fundamental questions in the philosophy of science. For example, the challenge David Hume (1711–1776) (*EHU*, VII, I, 50) posed for the idea of causality must be met in order to achieve a watertight account of causal explaining. Again the problem takes us further to the question about the basic process of experiencing the world as I believe that it can be solved only by advancing deeper and deeper into the most fundamental philosophical issues. Nevertheless, by the same token I come to lay the groundwork for the later discussion about the possibility of applying different perspectives towards actions. I will demonstrate how we necessarily conceive the world and phenomena in it through their properties among which events and causal relations also belong. This means that we never explain or even talk about any phenomenon in itself as all the talk is about their properties. In this simple thought lies the foundation of the idea according to which one can find various events and causal structures in one and the same phenomenon. Causality must always be attached to certain properties, and naturally different properties include different causal stories. Similarly, as we are able to contemplate phenomena through different perspectives, we are also able to find actions of different kinds requiring different types of explanations. All in all, we never explain phenomena in themselves but only properties dependent on the perspective applied.

As was said above, the second chapter of my study is solely about actions. I will try to reach all the aspects there are in acting, and through this will end up distinguishing three categories of action, namely purely physiological, voluntary and intentional. Naturally, in the context of human action the latter two are of most interest. The conceptual basis for them can be found in the history of the philosophy of action which definitely has its most influential

character in Aristotle (384–322 B.C.). Thus, for instance, Aristotle plays a crucial role in this chapter. Anyway, while making the distinctions, I need also to analyze the common concepts related to the theme, like ‘reason for action’ and ‘intention’. In my view, the distinctions between the types of action by the same token define the meaning of the concepts above. For example, the answer to the traditional question of the nature of reasons for action depends on which category of actions we are talking about. As I will argue, the idea of a suitable explanation changes according to the category, and again these categories are defined by the perspective chosen.

The second part of my work concentrates on combining the first two chapters together through which I will be able to answer the most fundamental question of my work, namely that of the conflicting ways of explaining actions. What I will do is first contemplate the role of agents and actions in the mechanisms of natural sciences after which I will do the same with those of social sciences. I will argue that, in the case of the first type of mechanisms we must view actions as purely physiological whereas in the case of the second they should be seen as voluntary. In my view, intentional actions are not needed in these explanatory mechanisms. Behind all, my primary aim is to release the tension between the ideas of true agency and mechanical explaining, by demonstrating that they can both exist yet only in different images. Simply said, it is perfectly permissible to concentrate on different aspects in actions, and thus to refer to different causal stories while explaining them. After all, the justification for these altering images was laid in the first chapter where I demonstrated how there is nothing wrong in contemplating one and the same phenomenon in different ways. Nevertheless, at the same time, I will concern myself more and more with the question of what it means, in the end, that we have these perspectives and how we actually apply them in practice. However, not before the concluding section of my work, I will take perhaps the most crucial question under scrutiny, namely that of what it is to talk about perspectives in the first place; in other words, what I am saying about the world in itself while I am talking about how we can contemplate it through different perspectives? Am I saying anything at all and, if not, what is the relation between my words and the truth? The answer I propose is that even if there are different kinds of actions and explanations to be found in the world, it does not imply that all the actions and explanations we may find are valid. Quite the contrary, there are always the rules governing our discourse about the world even if these rules are not as strict as many philosophers hope. Anyway, we have a long run ahead of us before there comes the time for fundamental ontological issues. I hope we all have the patience needed.

**PART I:**  
**THE FOUNDATIONS**

# 1. Explaining with Mechanisms

## 1.1 Hempel's Idea of Scientific Explanation

It is already a very strong tradition to begin an introduction to the theory of scientific explanation with the ideas of Carl G. Hempel, and neither do I see any reason for making an exception to it. Certainly, the rule has not developed by accident. Even though there has been a lot of interest in the question during the history of science, only since the presentation of Hempel's view the discussion started to boom at full blast.<sup>1</sup> All in all, my aim in this study is not to give a full report on this discourse from the beginning to the modern day. On the contrary, in this first chapter of my work, I will focus mainly upon the idea of scientific mechanism. Nonetheless, the reasons for starting the chapter with the Hempelian model for explanation are quite clear. It is because, as I see it, understanding the importance of mechanistic explanations requires some basic knowledge about the foundation of the discussion laid by Hempel and Paul Oppenheim. At the very end, it was through the faults of their model that people started to endorse other kinds of explanations.

Before taking Hempel's ideas under closer scrutiny, we should have a quick glance at the concept of scientific explanation. As James Woodward does in his review "Scientific Explanation" (2–4), we must notice that there are two separate aspects to it. Firstly, with the noun 'explanation' the concept seems to make a distinction between explanations and other kinds of accounts in science. For example, quite often we are satisfied with information which is "merely" descriptive. The problem is that the concept of explanation is used in so many ways in common language that it is impossible to give any precise definition of it that would cover all the different senses. For instance, one may explain an English word as well as explain how to patch a broken bicycle tire. As we can see, although one is willing to increase the knowledge of others in both situations, there is not much else common in the two examples. In any case, even though there is no one definition to be found for the concept of

---

<sup>1</sup> The boom in question did not begin immediately, in any case. According to Wesley C. Salmon (1989, 11), it took surprisingly long time before the effect started to take place.



explanation, in science as well as in this study the word is used in a specific sense. As Bas C. van Fraassen (1980, 134) states, an explanation should be taken to be an answer to a why-question.<sup>2</sup> That is to say that to give an explanation for a phenomenon is to explain *why* it happened. Another seemingly similar type of question is the how-question. Often we seem to give explanations to questions about how this or that thing works. However, to be precise, there is a significant difference between why- and how-questions. Whereas with why-questions we seek reasons, with how-questions we are after *descriptions* of processes. This, in any case, does not diminish the fact that a description may contain a great amount of answers to why-questions. Anyway, undoubtedly not all answers to why-questions are scientific explanations. Secondly, the very adjective ‘scientific’ is there to distinguish scientific explanations from other kinds of explanations. For example, explanations for our everyday actions are often fairly unscientific. But what then makes an account scientific? It is hard to give a precise answer to the question, but most likely all scientific explanations have to include features like truthfulness, empiricism, objectivity, accuracy and provability. Either way, my study is not restricted only to scientific explanations although this first chapter concentrates almost solely on scientific issues. In any case, what is common to all explanations I am examining here is that they are all answering the question why.

In 1948, Hempel and Oppenheim published an article entitled “Studies in the Logic of Explanation”. In this paper, they presented their view of scientific explanation which they call the ‘Deductive-Nomological (DN) model’. Although there had been similar thoughts expressed before, this was undoubtedly the most detailed presentation on the topic (Woodward 2009 [2003], 1). Therefore it also had such a huge impact on the field of the philosophy of science. Next I will briefly introduce their idea.

Hempel and Oppenheim (1948, 136–137) divide scientific explanations into two different parts: *explanandum* and *explanans*. They note that “[b]y the explanandum, we understand the sentence describing the phenomenon to be explained (not that phenomenon itself); by the explanans, the class of those sentences which are adduced to account for the phenomenon”. Later they add that “[...] the explanans falls into two subclasses; one of these

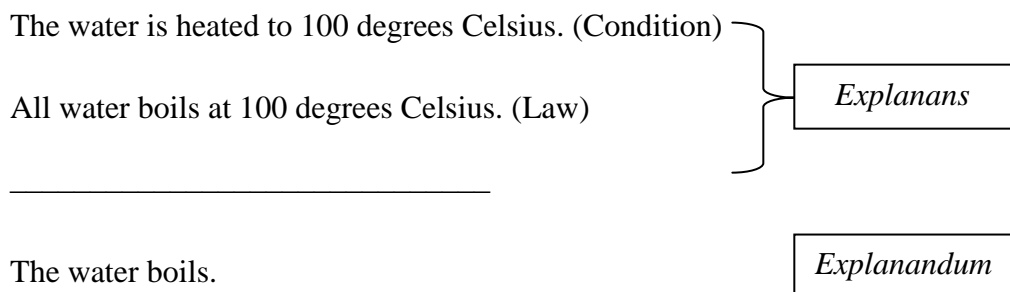
---

<sup>2</sup> Salmon (1984, 10) firstly agrees with van Fraassen on the matter, but later he (1989, 137) seems to have become quite convinced that how-possibly-questions also need special room in the discussion about the role of explanations. For example, we may ask how it was possible for an airplane to crash even though we are certain that it would be impossible to know the reason for sure. By this question, we do not seek one definitive answer but possible explanations by which the phenomenon becomes understandable. In my view, it is still hard to see why these how-possibly-questions require such a unique status. On the contrary, it strikes me as fairly plausible that they can always be reduced to why-questions which are given alternatives as an answer instead of a definitive one.

contains certain sentences  $C_1, C_2, \dots, C_k$  which state specific antecedent conditions; the other is a set of sentences  $L_1, L_2, \dots, L_r$  which represent general laws". To make their view as clear as possible, Hempel and Oppenheim (ibid. 137–138) impose three logical and one empirical condition for an adequate scientific explanation. In short, they are:

- (R1) The *explanandum* must be a logical consequence of the *explanans*.
- (R2) The *explanans* must contain general laws and these laws must be actually required for the derivation of the *explanandum*.
- (R3) The *explanans* must have empirical content; i.e., it must be capable, at least in principle, of test by experiment or observation.
- (R4) The sentences constituting the *explanans* must be true.

What is peculiar in Hempel's and Oppenheim's model is that they require a scientific explanation to have the form of a logical argument. To demonstrate their view I will give one simple example. For the boiling of water, the following explanation can be formulated:



Here we have an explanation for why the water boils, albeit naïve to the extent that it does not really give much information. Nonetheless, it fulfills all the conditions that were defined by Hempel and Oppenheim, so it helps us to understand the model in practice. First of all, the *explanandum* follows logically from the premises. For the second, as well as empirical content there is also a universal law in *explanans*. Finally, all the sentences are true as long as there actually is some water boiling somewhere.

In his later papers, Hempel loosens the conditions for an adequate scientific explanation a bit. He seems to have realized that the condition (R2) is too hard. Because of

the absence of a strict universal law, many potential explanations would lose their status as a scientific explanation. Actually, it is quite rare for us to be able to speak of a real universal law with great confidence. Therefore, in Hempel's most extensive and thoroughgoing writing on the topic "Aspects of Scientific Explanation" (380–393), he introduces two further models for an adequate scientific explanation. What is novel with these new models is that they also approve statistical laws. First of these, the Deductive-Statistical (DS) model is an argument where a statistical law is deduced from *explanans*, which necessarily contains at least one other statistical law. The more important one of these two models is the other one named the 'Inductive-Statistical (IS) model'. Contrary to the DS model, the IS model is meant to give explanations for *particular* phenomena by force of at least one statistical law. This model is called 'inductive' instead of 'deductive' because it is impossible to deduce a particular fact from statistical laws. Consequently, the explanations of the IS model are never absolute, although the statistical laws they contain cry out for high probability. Even though it is sometimes pretty natural to use laws with high probability in explanations instead of ones with a full guarantee, the problem with the IS-model is that those explanations do not cover the chance at the outcome. In other words, even the materialization of the highest probability is in need of an explanation.

As the role of the initiator of philosophical discussion is often unrewarding, Hempel and Oppenheim have received a lot of resistance. However, even if their account does not come across as perfect in the end, the importance of their article should not be underrated. Through the understanding of its faults, we are able to improve it as well as to see what else is needed in order to satisfy the demands of different fields of science. Now, before going any further, I will present two of the famous counterexamples presented against the DN model (it is important to notice that these counterexamples can be directed at the other Hempelian forms of explanation as well):<sup>3</sup>

(1) Bromberger's Flagpole Example:<sup>4</sup>

When the sun is shining on a flagpole, one can explain the length of the formed shadow by referring to the length of the flagpole, to the position of the sun and to the law of rectilinear propagation of light. Now, the problem with the Hempelian account is that, besides forming a DN explanation of

---

<sup>3</sup> To see more of the famous counterexamples: Salmon 1989, 46–50.

<sup>4</sup> Interestingly enough, according to Salmon (1989, 189, n. 12), Sylvain Bromberger never published this precise example which exactly in this form is perhaps the most famous counterexample against the DN model.

the previous kind, it is possible to form an externally precisely as adequate DN explanation which explains the length of the flagpole by referring to the other factors mentioned. However, our intuition should warn us that there is something seriously wrong with explaining some object's size by referring to its shadow. Even though the shadow might be helpful in calculation of the size, it feels odd to call that explaining.<sup>5</sup> As James Woodward (2009 [2003], 11) states, “[e]xamples like this suggest that at least some explanations possess directional or asymmetric features to which the DN model is insensitive”. Another famous example of the same type is the one where a storm is explained with a sharp drop of pressure in a barometer.

## (2) The Hexed Salt Example:

Let us presume that a witch or magician has placed a questionable hex on some amount of salt. Since it is likely that nobody who is interested in the philosophy of science believes that a bit of magic can have any influence on the laws of nature, undoubtedly we are allowed to formulate a universal law according to which all hexed salt will dissolve in water. If we agree that hexes do not make any difference, we should agree that hexed salt as well as ‘normal’ salt will always dissolve when put in water. Now, as we will see, it is possible to compose a perfectly adequate DN explanation with this new universal law we have formulated:

The salt is placed in water. (Condition 1)

The salt is hexed. (Condition 2)

All hexed salt dissolves in water. (Law)

---

The salt is dissolving.

It is obvious that there is something wrong with an explanation if it is explaining a phenomenon with some totally irrelevant factors. Therefore, one main problem with the DN model seems to be its vulnerability to explanatory irrelevancies. Another similar example is the one where there is

---

<sup>5</sup> In his work *The Scientific Image* (132–134), van Fraassen offers an imaginary example where the height of a tower is explained by its shadow, yet the story in question is one highly unusual exception.

a DN argument explaining the failure of a male to get pregnant with his habit of taking birth control pills.

It appears that Hempel's and Oppenheim's model for scientific explanation is not sufficient in itself. Both the previous examples suggest that the fundamental source of problems lies in the connection between the condition(s) and *explanandum*. As those examples demonstrate, the lawfulness in the occurrence of two phenomena does not guarantee that one could be used as an explanation of the other. As Woodward (2009 [2003], 12) notes, "[e]xplaining an outcome isn't just a matter of showing that it is nomically expectable". All in all, it comes across as if there is some important piece missing in the DN model as the laws in explanations cannot be mere laws of regularity. Next I will find out if a right kind of notion of causality could do the trick. Even if I think that modifying the idea of causality is not sufficient to make the DN model a perfect structure for all sciences, in my view it is an indispensable improvement if we are to make the DN model a useful piece of our explanatory machinery. In any case, with or without the DN model, we cannot escape the idea of 'true causality' in scientific explanations.

## 1.2 About Causality

### 1.2.1 Hume's Problem

When it is said that Hempel's account of scientific explanation seems to miss a *right* kind of notion of causality, there is a presupposition that he still holds some kind of notion of it. Naturally, it is true. In "Aspects of Scientific Explanation" (348), he states his view of causality by citing Michael Scriven: "To say that X causes Y is to say that under proper conditions, an X will be followed by a Y". This kind of notion of causality is known as the 'Humean cause'. In order to achieve proper understanding of it, we have to go back to the writings of David Hume.

The common understanding of cause or causality is the sense of something being made to come about by some other thing. However, in his originally 1748 published *An Enquiry of Human Understanding* (VII, I, 50), Hume gives us his famous critique of this idea:

When we look about us towards external objects, and consider the operation of causes, we are never able, in a single instance, to discover any power or necessary connexion; any quality, which binds the effect to the cause, and renders the one an infallible consequence of the other. We only find that the one does actually, in fact, follow the other. The impulse of one billiard-ball is attended with motion in the second. This is the whole that appears to the *outward* senses. The mind feels no sentiment or *inward* impression from this succession of objects: Consequently, there is not, in any single, particular instance of cause and effect, any thing which can suggest the idea of power or necessary connexion.

In this notorious passage, Hume challenges the common sense notion of causality. He notes that, in the end, we are capable of perceiving nothing more than various regularities, and thus we can never grasp any true causal relations in the strong sense of the word. While examining a cause and its effect, we are never able to find the very causality in itself between the two events. Of course we may try to explore the empty space left between them, but sooner or later we end up with the same problem. For example, between a brick being thrown and a window shattering, we can find many different events, like the brick hitting the window. However, there is no escape from the fact that there will always be the same kind of divide between the events. Even if the empty space is now filled with the new connecting event, two new divides are born between the brick hitting the window and the two original events. Therefore, because of our incapability to find the connection between a cause and an effect, David Hume (*EHU*, VII, II, 60) concludes that we should change our idea of causality. He thinks that we should avoid talking about something we cannot perceive, and hence it would be better to modify our notion of causality to cohere with our experience. Under the circumstances, he defines “[...] a cause to be *an object, followed by another, and where all the objects similar to the first are followed by objects similar to the second*”. Furthermore, he adds that “[...] in other words *where, if the first object had not been, the second never had*

*existed*".<sup>6</sup> As we can see, the core of this novel idea of causality called 'Humean' is solely the idea of regularity between two temporally consecutive objects. Briefly, Humean causality is causality without the idea of "true causing" or making something to come about.

In my view, Hume appears to be right while denying that any 'true causality' could be experienced between two events. After all, we are able to witness nothing but the endless sequence of separated events. However, as the Hempelian model for explaining proved, this kind of notion for causality is simply not enough. For a competent explanation, it is not sufficient to rely only on regularity in the appearance of separate events. Furthermore, regardless of the popularity of his idea of causality, the whole process comes across as somewhat suspicious. Is it really appropriate simply to give a new meaning to a concept no matter how artificial it feels? As C. J. Ducasse (1926, 63) stresses, everybody (at least those who have not read Hume) uses the word 'cause' in the sense of making something come about. He notes that "[w]hen any philosophically pure-minded person sees a brick strike a window and the window break, he judges that the impact of the brick was the cause of the breaking, *because* he believes that impact to have been the only change which took place then in the immediate environment of the window". Naturally, the person in question might be proven mistaken, but even then there is normally an existence of some other cause suggested. The issue is that since we have so strong idea of 'true causing' associated with our concept of causality, would it not be more appropriate either to stop using the word 'cause' or merely to try to defend the idea of 'true causing' somehow. As Ludwig Wittgenstein (*PI*, 124.) rightly puts it, "[p]hilosophy may in no way interfere with the actual use of language; it can in the end only describe it". Since I believe that it is impossible to eliminate such a crucial idea, I will choose the latter path.

In order to make better sense of the problem, it is best to turn to an article entitled "Causality and Determination" by G. E. M. Anscombe. In this paper (8–13), she makes two insightful remarks about Hume's account of causality and, in my view, they are both indispensable keys to solving the problem posed two centuries earlier. First, she (*ibid.* 10–11) notes that the idea of causality should not be associated with the idea of necessitation.<sup>7</sup> As we can see in the preceding citation, Hume takes the idea of necessity to play a crucial role in the

---

<sup>6</sup> Even if Hume himself seems to think that he is merely saying the same thing in different words, he happens to give an important further condition for his notion of causality. However, this I will not consider more deeply before Section 1.2.4.

<sup>7</sup> Actually, Anscombe herself makes these points in a different order, but for structural reasons I present them in the opposite order. I believe this arrangement will not harm her ideas.

concept of causality. Interestingly enough, usually the common talk of causes does not include this idea. On the contrary, it is very rare to find necessity in causal statements. Even when we talk about laws, we often refer to *ceteris paribus* – laws which mean that they hold only in certain customary circumstances. This is because of the limits of human knowledge as well as because of the enormous task it would be to enumerate all the single exceptions in every situation. For example, a causal statement according to which a billiard ball will move after being hit does not include the idea of necessity. The reason is that there is also a huge amount of possible conditions which could force the ball to stay put. For instance, the ball might be attached to the table. To mention another example, one is allowed to say with great certainty that smoking causes lung cancer while knowing perfectly well that some people may smoke all their lives without ever getting the horrible disease. In any case, often while defending this thought, I encounter objections of the same kind.<sup>8</sup> It is usually argued that, even though we know that there is no law between smoking and lung cancer *per se*, we have the idea that, in some particular circumstances all things considered, it has to be so that the cancer must follow the smoking. Similarly, even if we know that billiard balls do not always move when hit, while making a statement about some particular instance there is always the idea of necessity included which means that we must think that since everything was the way it was, it must have been impossible for the ball not to move. Anyway, I am not ready to go any further than admitting that this would probably be the case, if we started to put pressure on every single individual making a causal statement. It is likely that everyone with sufficient understanding would before long admit that, all things considered, his causal claim after all contains some idea of necessity. However, to my mind, this is not yet a proper counterexample, as it should be asked how many of the causal claims made in the everyday context already include the very idea of necessity *in the beginning*. Without the pressure put on poor people making innocent causal statements, it is hard to believe their claims in the original form saying anything about necessity. On the contrary, if I say something about a ball hit and moved during a game of billiards, in reality there is nothing in my claim contradicting the possibility of the ball either not moving at all or moving in a different direction. Of course, after deeper contemplation I would be willing to concede that the ball simply had to behave the way it did, all other things being equal, but nonetheless the original claim about the hit and movement did not take any sides on the question. This is what is meant by saying that there is no necessity in ordinary causal statements. In conclusion, I

---

<sup>8</sup> Once more, I am willing to thank Olli Koistinen and Kari Lukka for highly fruitful conversations.



would like to emphasize that perceiving a causal connection does not require a perception of any necessary connection, and it is actually extremely unusual for a causal statement to express necessity.

The other point Anscombe makes is even more crucial to our solving Hume's problem. At the beginning of her writing (1971, 8), she questions the sense in which Hume is allowed to say that we cannot "find" causality between events. If it is justified to say that we cannot perceive causality in the world, it seems as justified to say that neither can we perceive any movement in the world. After all, in the case of movement it can be claimed that all we perceive is no more than an object situated in sequent locations at sequent moments of time. Yet it is something which we take as a perception of movement. Furthermore, what is more important is that Anscombe realizes that causality lies in the causal concepts. In a world in which no causal concepts exist, neither can there be a place found for the idea of causality. As Anscombe (ibid. 9) states, "...the word 'cause' can be *added* to a language in which are already represented many causal concepts". Verbs like 'make', 'break', 'wet', 'push', 'hurt', 'move', 'fill' and so forth are all causal verbs, and their justified use is a sign of causality perceived. As Anscombe (ibid. 10) notes, "[a]s surely as we learned to call people by name or to report from seeing it that the cat was on the table, we also learned to report from having observed it that someone drank up the milk or that the dog made a funny noise or that things were cut or broken by whatever we saw cut or break them".

In my view, Anscombe is absolutely right in emphasizing the facts that the idea of causality lies in the causal concepts and that the justified use of those concepts signifies that there is a causal connection perceived. Nonetheless, I still believe that Hume is also right when saying that there is no causality to be perceived between separate events. During the next sections, I will present my own view about causality and the world itself. Even though I am trying to clarify and continue Anscombe's thoughts, I want to make it clear that my talk about properties should not be associated with her philosophy. Anyway, the next section is one of the most crucial ones in my work, since I am about to present my view of the nature of events and other entities. After all, it is of utmost importance to understand what we are talking about, in the end, when talking about the world, since the act of explaining as an act of answering a why-question is evidently nothing but one instance of discourse. As is said, here the foundation will be laid for the chapters to come.

### 1.2.2 Living in the World of Properties

What I mean to say with this heading is that, to my mind, it is simply impossible for one to grasp anything without doing so through its properties. In its plainest form, this is the line of thought I wish to express and defend in the following paragraphs. In my view, referring to something is necessarily a reference through one or more of its properties, and this implies that we are always stuck with properties as we are not capable of solely passing through them without highlighting them as we refer. Therefore, our talk about the world is never neutral in a sense that it could be independent of properties we choose. This is not to say that we only talk about properties but simply that it is impossible for us to talk about anything without doing so through its properties. Naturally, this is not a statement about the ontology of the things behind properties (if there even are any). I am not willing to make any claims about them, since I am only interested in explaining, that is to say, in the context of speaking and referring. Clearly, most of us share the intuition according to which, in order for a certain property to exist, there must also be a carrier for this property, as it is hard to imagine properties merely floating in the air. However, I am not willing to oppose this intuition as I am even ready to go as far as to call these carriers, say, ‘Xs’ and ‘Ys’, yet I adhere to my claim according to which nothing else can be said about these carriers other than that they, by definition, carry the very properties they are postulated to carry. In any case, it is needless to say that we do not always have to grasp our experiences in any manner. Sometimes we cannot do so even if we wanted to. Everybody knows those weird experiences we cannot describe to anybody. Sometimes we can even see something for which we do not have a word. Either way, this study is about explaining, and that is why I conclude, as I believe Immanuel Kant would, by saying that since we cannot talk about the things-in-themselves, we should not waste our time trying to do so.<sup>9</sup> As we can never find our way past the properties with which all our talk is of necessity intertwined, the only proper thing to do is to accept this insufficiency and concentrate the study on our relation to properties.

---

<sup>9</sup> One must pay close attention to the fact that although I am referring to the division Kant (*CPR*) makes between thing-in-themselves and phenomena appearing to our senses, actually I am myself making a further division between the phenomena and their properties. All in all, I am not talking about how our senses have an effect upon the phenomena we perceive but how we are incapable of talking about these phenomena *per se* without picking up their properties. However, as our distinctions share many similar features, I am inclined to refer to Kant even if I am fully aware of the possible misunderstandings.

Since statements like those above should not be accepted as given, I will try to justify them by giving a plain example. My mother had a dog which was very dear to her, but unfortunately this little fellow passed away few years ago. Now, talking about this animal as ‘a dog’ is nothing but highlighting the property of ‘dogness’ or ‘being a dog’, like calling it ‘an animal’ is highlighting another property. Calling it ‘my mother’s dog’ or ‘a hairy little creature’ is evidently nothing more than bringing out more properties. However, people are used to giving proper names to each other solely to make their dealings easier. As it is known, those names are also given to other creatures and even to artifacts. To return to our example, my mother had named her dog ‘Nelly’. In any case, it is easily forgotten that having a proper name is also a property in a sense that giving a name to a thing is to give a property for a more accurate reference. Names are often highly neutral as the process of deciding names resembles a kind of lottery. Traditionally, they are carried along in the family or, as is more common nowadays, they merely spring from the name giver’s likings. Now, even though they have very little informational value, it does not mean that they have any kind of special status among properties. In our example, my mother’s dog had a property of “being named ‘Nelly’” as it had the property of ‘being a dog’. To be more precise, there was an X which shared the properties of “being named 'Nelly'” and ‘being a dog’. As I stated in the previous paragraph, the ontological nature of X is something about which it is extremely hard to say anything since all our talk ends in its properties.

It strikes me as an extremely easy task to defend this view according to which we cannot talk about anything without referring to one or more of its properties. If somebody wants to argue against it, we may just ask him to mention an object in the world without mentioning at least one of its properties. This seems quite obviously to be impossible. Naturally, he can point at different things with his fingers, but it still remains dubious what he is actually pointing at. Alternatively, his pointing can be seen as bringing out the property of being in a certain spatial location, but to these kinds of properties I will return a little later. Similarly, to use words like ‘it’, ‘this’ or ‘that’ requires that the reference is made clear in some other way, and thus within a pure discussion they do not mean a thing if the target has not been brought out. For instance, while using the word ‘it’ or ‘she’, we must presume that the other person already knows what or whom we are talking about, and this knowledge is likely gained from our earlier use of some less relative properties. Anyway, to proceed, two ways of talking about properties should be distinguished. First, we may refer to a property in a generic sense. Returning to our earlier example, while referring to the property of ‘being a

dog' in a generic sense, we see it as one which is shared by all the dogs in the world. Second, properties can be understood in a particular manner. This means that, when we talk about Nelly's 'dogness', we may see it as a 'dogness' which is unique and which belongs only to this particular dog. These particular properties are often called 'tropes'.<sup>10</sup> Their status is a big issue in modern metaphysics, and naturally there are a lot of questions concerning them, but in some sense I see them as indispensable parts of our understanding of the world. The need for tropes surfaces if we are to think about explaining. To take an example, let us imagine that the constant nagging of a housewife has made her husband lose all hair. Now, in this situation one is undeniably allowed to say that the particular property of 'being bald' has an identifiable cause which is the nagging of that malicious wife. However, it would not seem right to state that the generic property of 'being bald' is caused by this housewife's nagging since nobody in this world is that bad a nagger that she could be blamed for all the baldness in the world. It would also be an exaggeration even to claim that nagging in general is behind all the baldness in the world. This is why it is so important to understand the difference between properties in generic and in particular senses.

Yet it is fairly straightforward to speak about simple entities like dogs and nagers. These kinds of things have their own form and matter which normally remains more or less constant through the time. That means that they are usually something quite concrete. For instance, we may take a dog onto our lap and show it to others saying: "this is the fellow I was just talking about". However, the issue naturally gets much more complicated when we get to entities like pains, thoughts and obligations, since these are examples of things we cannot take onto our lap and show to our friends. Nevertheless, I will not take them under scrutiny now as my main target is another awkward class of entities, namely the category of events. The ontological status of events has concerned philosophers for a long time.<sup>11</sup> The question arises of what kind of entity is, say, 'a falling of a tree' or 'a revolution'. As I mentioned above, there are also some other truly tricky entities like feelings and tasks, but still the case is at least slightly simpler with entities which are static, even to some extent. Although we cannot point at the foregoing entities, we still often think that we have a decent grip on them. Even if the talk is always only about properties, we normally have quite a strong sense that there is some more or less constant X behind those properties to which we are referring. Especially entities like rocks and trees have their clear boundaries in space, and

---

<sup>10</sup> E.g. John Bacon (2008, [1997]) has written an introduction to tropes.

<sup>11</sup> One good introduction to this discussion is Pianesi & Varzi 2000.

that is why it is relatively easy to talk about them. Evidently, that is not the case with events. Localizing and marking out those vague entities is a process which is always more open to various interpretations. Under the circumstances, the latter part of this section I will dedicate to the category of events.

From what I know, in his paper “On the Nature of Certain Philosophical Entities” originally published in 1969, Richard Montague was the first philosopher to introduce the idea of events as properties. While this theory has not won over the majority of philosophers, it sounds sensible after all that has been said in this section. If we are to admit that our talk is always restricted to properties, why should we assume that there could be any difference in the case of events? On the contrary, when talking about various changes in the world it seems equally impossible to go any further than properties. While ‘being a dog’ and ‘being an animal’ are evidently properties, the same idea applies to the things like ‘being an explosion’ and ‘being a church service’. Nevertheless, if we accept the idea of events as properties, the question subsequently arises about whose properties they are then? Anyway, we should not get confused about this matter, since, as has been stressed various times, what is behind properties is not a thing we are able to talk about (if there even is anything there to begin with). For instance, we may say that there is a phenomenon Y which possesses the properties of ‘being an explosion’ and ‘being a terrorist attack’, yet we should not go further than that. In any case, the basic natures of events and other properties are still essentially different. Even if we take granted that all entities have their location in time, the requirement for temporality is something very different where events are concerned. I believe that it is appropriate to say that diffusion in time is a prerequisite for something to be an event. What I mean by this is that instead of merely existing through time, events are spread in time so that their temporal length is a crucial part of their essence. To go even further with the idea of temporal length, Richard Montague states that events should be defined as properties of the moments of time. In his (1974 [1969], 149–150) own words, “... the event of the sun’s rising will be the property of being a moment at which the sun rises, and in general will form a certain class of properties of moments of time”. Contrary to Montague, I will be more cautious as I adhere to the statement about the sanctity of the carriers of properties. After all, it is important to notice that having a temporal location is in itself nothing but one more property (of course, this does not mean to say that properties cannot have properties). Unquestionably, it feels natural to link the property of ‘being an explosion’ to the space and time where the bomb explodes as well as to link the property of ‘dogness’ to a physical

wholeness which barks, but still we should keep in mind that things like the possession of spatial and temporal locations are nothing but properties themselves. For instance, as I am writing this text here in my office, my spatial location is here where I sit whereas the temporal location of my act is the passing moment. Still, in spite of all, I believe that Montague is right to emphasize the role of temporal aspect in events, since they are always distributed in time. Under the circumstances, I would prefer to say that referring to a phenomenon through an event necessitates some (usually fairly vague) temporal location associated with this event. This is to say that particular events always include some temporal property. In contrast to those, events understood in a generic sense are naturally free from any pinpointing. When it comes to the necessity of the spatial locations of particular events, we have a trickier question. The problem is that there are events which are very difficult to localize spatially. Perhaps the best examples are mental events, like ‘remembering one’s youth’. There is a huge temptation to say that they do not have any spatial location at all, but even then those rather naïve propositions which link mental events to their “owner’s” head or somewhere around it have some weird sense in them. Besides these flimsy considerations, I refuse to draw any conclusion on the matter.

Although we need to pay attention to definitions, our everyday experiences and talk around them should not be forgotten. As we use words like ‘tree’ and ‘rock’ to refer to some physical particles, we use words like ‘explosion’ and ‘revolution’ to refer to some things going on inside certain spatiotemporal boundaries. Nonetheless, what is crucial to remember is that getting inside these boundaries is never completely neutral, since the event we postulate always highlights one or more aspects in the phenomenon. Either way, the problem with events appears to be the vagueness of their boundaries. To use a metaphor, events are not in the world in the same manner as products are on the shelves of a supermarket. Contrary to trees and rocks, the spatiotemporal boundaries of events are open to various interpretations. For instance, finding two Vietnam veterans having a conversation about the war and asking them to define the war spatially and temporally will probably bring out fairly diverse views. However, in order to continue the conversation, it is enough to keep some sort of consensus about the limits of the war. Absolute agreement is never necessary if even possible. All in all, talk about events demands much more creativity than talk about more concrete entities. It is more like drawing borderlines on the spatiotemporal landscape instead of merely picking out ready-made objects.

In any case, there is one notorious problem with the kind of view of events which I myself am also defending. Considering events as properties, instead of some more independent entities, quite easily ends up with an ontological view that allows the presence of an immense amount of events. This brings us to the question of the identity of events. Perhaps the most famous example of the problem is that where it is asked whether the event of ‘Brutus killing Caesar’ was the same event as ‘Brutus stabbing Caesar’. Since our common intuition favors the affirmative reply, all the views confined to give the negative reply are often automatically taken to be mistaken. Another similar example is best formulated by citing Donald Davidson. In his famous paper “Actions, Reasons, and Causes” (686) he presents his own view of the question quite explicitly saying:

I flip the switch, turn on the light, and illuminate the room. Unbeknownst to me I also alert a prowler to the fact I am home. Here I do not do four things, but only one, of which four descriptions have been given.

This is also a good example of a question to which certain replies are almost automatically taken to be wrong. Stating that Davidson actually does four things there at the same time is that kind of answer. No matter what, there are still philosophers who dare to argue against the grain. For the most part, the view is profiled to Jaegwon Kim who stubbornly insists that ‘the killing of Caesar’ was a different event to ‘the stabbing of Caesar’ and that all the actions Davidson did are separate ones (1993 [1976]; 42–44, 1973a; 226–227). As for example Jonathan Bennett (1988, 73) notes, Kim’s account of events is so radical that it frightens away most students of philosophy. Without going deeper into Jaegwon Kim’s view, I will try to defend his responses in the light of my way of thinking.<sup>12</sup>

The view of events as properties naturally entails there being a huge amount of events attached to one and the same situation. After all, one thing can always have an endless amount of properties which may all be different from each other. For example, if we say that something is a dog, it means that most likely we believe that there is some X which has the property of ‘being a dog’. This again implies that it also has the property of ‘being an animal’, since all dogs are by definition animals as well. Now, because the property of ‘being

---

<sup>12</sup> One thing that strikes me as problematic in Kim’s view of events is his (1993 [1976], 35) requirement for named substance to which the property of event belongs. I have myself been avoiding engagements of the kind, since naming a substance in cases like storms comes across as exceedingly unnatural. Of course, every storm has its vague borders in the space dimension, but likely it is not the kind of substance we are looking for. This problem may not be decisive for Kim’s view, but it still makes me even more careful of saying anything else about the carriers of properties than that they are undoubtedly carrying the very properties.

a dog' is evidently something else than the property of 'being an animal', X in question has at least two different properties. To my mind, the same goes for events. Saying that 'the killing of Caesar' and 'the stabbing of Caesar' were different events is simply to say that there was some Y which had the properties of 'being a killing of Caesar' and 'being a stabbing of Caesar'. Although they both refer to this same phenomenon, they do it in different manners, and thus are different properties. Now, if we consider the conflicting views of philosophers, I believe that the basic idea is the same at least with most of them. In the end, there is always something going on in the world which we are trying to reach by describing it and by highlighting its features. Problems arise mostly in our use of concepts. Whereas, according to the Davidsonian view, there is always only one event which is merely described in various ways, I am arguing that there is only one phenomenon to which one may refer by bringing out different events it may be said to carry. As in my view these events are properties of the phenomenon, only in the case in which they are saying precisely the same thing can they be called identical. For example, 'killing' and 'tappaminen' are the same thing, as only the language used is different. In these kinds of cases, the thought remains the same, and hence they are just two different ways of mentioning the same property. However, it must be honestly admitted that the question about the killing and the stabbing is far from the worst scenario for my theory that one can imagine. Pushing forward on the chosen path, at some point we will end up with examples like 'the killing of Caesar' and 'the cruel killing of Caesar' and, as has been stressed before, we must keep stating that they are not the same event. Nevertheless, I want to make clear that I can see no decisive difficulty in these kinds of cases. The famous Leibniz's law states that, in order to be identical, entities must share all the same properties, and I cannot see any reason why this law should not be applied to properties as well. Evidently, only the latter event has the property of cruelty, and thus I believe that there cannot be any other option but to conclude that they are not the same event. Once more, they can still be taken to refer to one and the same phenomenon.

As has been said before, my way of talking about events has its disadvantages. Common intuition almost compels us to say that the stabbing and killing of Caesar are one and the same event, and hence my view goes badly against the grain. In any case, we must remember that this kind of thinking ignores the foundation of my view, which states that all we can grasp are properties and that talk about things behind them must contain some kind of misunderstanding. However, in the light of common intuitions, there are also some advantages in my account. At the same time it should feel uncomfortable to claim that the



event of killing and the event of stabbing are one and the same thing. In the previous paragraph, I mentioned well established Leibniz's law. Now, for instance, stabbing includes the idea of a knife whereas killing includes the idea of dying, and thus they do not share all the same properties. That is to say that, according to Leibniz's law, they cannot be identical. In addition, the difference becomes even clearer when we get to explaining since the two events may cry out for different explanations. Evidently, Brutus had different reasons for merely killing Caesar than to do it particularly with a knife. The other advantage in my view is that talk about generic and particular events makes it easy for us to understand the phenomenon of the recurring of an event. The generic properties are ones that can be made particular by attaching some temporal and spatial features to them. The recurring of an event is merely finding the same generic event in changing conditions.

### 1.2.3 A Place for Causality

The preceding section lured us away from the question of causality for a while. Nevertheless, it is fairly evident that understanding the nature of different entities and especially events is indeed crucial for understanding the nature of causality between them. I believe that this fact justifies "the false steps" taken. However, even though I concentrated mostly on the nature of events, I do not want to go as far as to conclude that all causality is between events, as is often presumed (for example, Ducasse 1926, 58). On the contrary, I think that this is mistaken, at least to some extent. For instance, it strikes me as perfectly natural to say that, as in one of the preceding examples, nagging caused the baldness of some poor man. To continue with this example, we may add that the baldness again made the man depressed. Now, instead of events, we are obviously talking about more constant properties. Of course, it is still possible to argue that even these properties can be turned to events like 'the man losing his hair' and 'the man getting depressed', and undeniably there is some truth in that. It even seems more correct to say that, instead of the property of baldness, the effect of nagging was the event of losing hair. All the same, I would still abstain from forming such a drastic statement. One should ask, for example, what kind of event is the event of getting depressed. By reason of its extremely vague form, this kind of event comes across as so artificial that it

strikes me as a much better option to extend the concept of causality to encompass other kinds of properties as well. Besides, it is possible that, after all, the event of losing hair did not cause the depression but the continuous living with the property of baldness. Now it appears to be more truthful to say that the baldness itself, instead of the event of losing hair or the event of living with it, caused the depression. In conclusion, I would say I am much more comfortable associating the idea of causality with all kinds of properties instead of with events solely.

Either way, we should not get ahead of ourselves. At the end of Section 1.2.1, I noted that, in my view, Hume is right with his famous claim according to which we can never get into actual causality in the world. In spite of this, I turned to Anscombe's remarks in order to excuse the talk about 'true causality'. In addition, the preceding section was about our constricted ability to grasp things around us and, notably, to talk about them. In this section, my aim is to put these pieces together and to present my own view about the problematic notion of causality.

Immanuel Kant (*Prolegomena*, §27–§30) tries to solve the problem Hume posed by stating that, although there is no causality in the world-in-itself, we cannot conceive the world without the idea of it. In other words, Kant wants to stress that causality is something through which we contemplate the world and thus is indispensable for making sense of it. Kant's point here is of the utmost importance. If we attempt to consider the world in its most fundamental form, we are easily able to create an image of some extremely small particles or merely some fundamental energy. In these images there is no causality as everything is nothing but some basic movement or pulse going on *ad infinitum*. In order to find causality in the world, we need to postulate some entities between which the causal relation can be understood. As it should be clear, causal connection is a relation between two or more entities, and hence cannot be understood without these graspable parts to which it can be attached. Anyway, we do not need to get so deep into those fundamental elements to appreciate the nature of Hume's problem. If we observe the world without postulating any singular entities, we realize that there is only an ever-continuous stream of change or movement going on. There cannot be things causing others if there are no singular things in the first place. So, in order to find causality, we need to postulate its related parts in the world. Even so, Hume has his point. Even if we postulate two events which we normally consider being related by a causal connection, say 'the brick hitting the window' and 'the window shattering', we still have the problem. It may be asked what that causal connection is

and how we can get hold of it. As Hume said, all there is to perceive are two separate events which are often sequential. So, now it should be asked, if there evidently is no ‘true causality’ in the world-in-itself, is it reasonable to cling to the notion in question so desperately. As perhaps Kant would have answered, the idea of ‘true causality’ is something we cannot get rid of as long as we are interested in making any sense of the world. It is indispensable for our understanding of the world. This all means that we have no other option but to find a proper place for the notion.

At the end of Section 1.2.1, I raised what Anscombe had to say about Hume’s problem. In her second remark, she stressed how causality is included in causal concepts and how it is dependent upon them. In addition, she also noted that Hume seems to make too big a deal out of the difficulties in perceiving causality. I think that Anscombe is right. Now, what we need to do next is to ask what it means to perceive something in the first place. In other words, when are we allowed to say that we perceive something, that there is something in the world? To take a simple example, what it is to say that there are dogs in the world and that we perceive them? I myself would reply that there are dogs if it is reasonable to think that there are some Xs to which we can *correctly* refer by calling them dogs. Therefore, to perceive a dog is to perceive certain properties which we normally associate with the idea of dogness. Nevertheless, as we have noticed, dogs are a far too easy example in their stability and concreteness. What about relations then? How come we are allowed to say that there are, say, lovers in the world (in the Platonic sense)? The lovers- relation is obviously a symmetric one which means that in order to call some couple ‘lovers’ they both need to love each other. Since words are often misused, the best guarantees of love are to be found in certain gestures which all of us know, or at least that is what I hope. Therefore, to say that Romeo and Juliet are lovers is to say that there are certain signs in their behavior towards each other that makes it justified to call them ‘lovers’. As is evident, there is nothing to be sensed directly between Romeo and Juliet, because plainly these kinds of relations are not something to be sensed in that way. Usually, relations are not something one can grasp as easily as dogs or even events, and this is also the case with causal connection.<sup>13</sup> There are simply certain signs which we consider supporting the existence of a causal connection. Perceiving these signs means that it is appropriate to state that there is a causal relation, and this is what, at least in my view, it is to perceive causality. Naturally, the next question should be about these signs. Unfortunately, the amount of such signs is so enormous that the question cannot be comprehensively

---

<sup>13</sup> Of course there are also exceptions. For instance, it is quite easy to get a hold on positional relations.

answered. In addition, it can be said that every situation is so unique that it is not possible to give a precise definition, even with all the time in the world. On balance, we have now nicely reached the point where Anscombe's remark about the role of causal concepts comes into play.

If we take the earlier example of the brick and the window under scrutiny, it will be easily demonstrated what I have in mind. As has already been discovered, there is no causal connection to be directly sensed between the event of 'brick hitting window' and the event of 'window shattering'. Nevertheless, our mind instantly connects those two events with a causal connection because all the evidence speaks on behalf of it. We know the laws of physics and we have seen similar cases before. Moreover, the continuum in time and space between the two events combined with facts about nature of the entities related is good enough evidence in favor of the causal connection between the two events. Of course we may be in error, but in that case, with all the proof in question, we should not be blamed for our mistaken estimation. All in all, in this kind of case it is highly justified to speak in favor of a causal connection, which is to say that we are allowed *to combine the two events with a causal concept*. The same goes the other way round. If it is justified to combine two separate properties with a causal concept, these two properties should be considered as being in a causal relation. Therefore, as it was with dogs, saying that there is a causal connection is to say that it is reasonable to think that there is some Y which can be described with causal concepts. In other words, this Y has a causal property. To get back to our example, we should first of all say that we do not always need to go through the process of combining properties in order to be allowed to speak about causality. The crucial condition is only to be justified to do so. Now, in our case there are two events, namely 'the brick hitting the window' and 'the window shattering'. As we concluded, in this particular case the evidence strongly supports the supposition of the existence of the causal relation. What does this mean? On the whole, it means that there is at least some causal concept which can be used to combine these two events. Since there is sufficient evidence for using the causal verb 'break' in this particular case, we are now allowed to unite these two events into one, namely the event of 'the brick breaking the window into pieces'. As Anscombe said, the causality lies in causal concepts, and the justification for the use of those concepts means that there is causality perceived. Nonetheless, as it is with every perception, even with the strongest evidence there is always a possibility of error.

All this might seem like merely bypassing the problem. In any case, I believe that the kind of problem Hume posed cannot be solved in any other way than by accepting the limits of our knowledge. As Anscombe noted earlier, if we want stricter rules for our talk about perceptions, we will end up in a situation where there is nothing to talk about any more. In the end, causality is no more than the idea of a relation in which some temporally preceding thing is, on some level, responsible for the existence of some following thing. As Kant stated, the idea of causality is one of the fundamental ideas through which we contemplate the world, and hence cannot be properly reduced to some other ideas. In language, there is a massive amount of concepts which include this idea. On page 25, I gave a few examples of them. By growing up with language, one learns to use these words and, as Anscombe (1971, 9) notes, that is a clear sign of one having the concept of cause. For instance, in order to use the verb ‘wet’ properly, we need to understand the causal nature of the event of ‘wetting’. What is it then to perceive causality? In my view, it is merely to perceive events like ‘wetting’. There simply cannot be causality between properties if there are no causal concepts to be used, but if it is justified to combine them with words like ‘wet’, we are allowed to state that there is a causal connection to be found. To conclude, I would say that, even with the limited equipment we have, we still must keep calling people ‘lovers’ and ‘wetters’, unless we are willing to close our mouths for good.

#### 1.2.4 Causality and Context

In the sections above, I have argued that in some sense the world is much like we make it to be. At least the world described by language is. As it turns out, the concepts through which we consider the world are always limited as they are necessarily man-made, and hence are also very limited. In any case, they are the tools we are doomed to operate with. Every concept is merely one side of a thing, and therefore the linguistic equipment can never reach anything beyond those sides. Without paying too much attention to it, we always name things going on around us such as ‘bank robberies’ and ‘killings of a philosopher’, and evidently the names given highlight certain features in the episodes in question. For instance, calling an event ‘a killing of a philosopher’, we bring out information about the victim as well as make

clear that he ends up dead. At the same time, by using different properties we could have highlighted very different features. There could have been different events, and that is why it is appropriate to say that in a way the world is what we make of it. In the previous example, the story did not tell us, for example, anything about the method of the killing or any fact about the killer at all. For instance, ‘the shooting of an unknown man’ would have told us a completely different story. As it is said, the use of concepts can never be totally neutral as it always leans on certain features which we decide, consciously or unconsciously, to pick up.

These teachings are of extreme importance to the understanding of the nature of explaining. Most crucial to digest is that all the obscurities mentioned before also apply to the context of explaining, since explaining as well is nothing but another play with concepts, that is to say, a play with properties. If causality is just another vague property and explaining is mostly about causality (I will return to the question of the relation between the two in the next section), what could be different with explanations? With all the irony this thought includes, it sure is a hard one to go along with for scientists. As I listed earlier on page 17, to call something ‘science’ requires features like objectivity and accuracy among others, but on the fundamental level all these features appear to be a long way out of our reach. Nonetheless, I am not aiming to write a manifesto against the creditability of science, since we all know how well scientific explanations and predictions work at best. My purpose is nothing more than to slightly open the curtain behind the practice to show that, in the end, all science and its explanations necessarily rely on viewpoints and more or less loaded vocabulary. After all, we never explain pure phenomena or things-in-themselves just as we cannot explain anything with those Xs and Ys behind the properties. All the *explananda* as well as the *explanantia* are about properties and there is simply no other option but to learn to live with the fact.

However, now is the time to get back to where we started from, namely to the problems with the Hempelian view of explanation. At the end of Section 1.1, I proposed two counterexamples for the DN model. Although they were both merely occasional cases, they represented two characteristic difficulties for the model presented in the section. The first one, namely ‘the flagpole-example’, represented the problem of direction. As we will see in the next section, Hempel (1965, 352) undermines the role of causality in explanations by arguing that there are also other kinds of laws on which they may lean. Nevertheless, if these laws do not include a clear direction, we will end up with nasty counterexamples. In ‘the flagpole-example’ the question is about “a law of coexistence” instead of a causal law, but we

will not talk about these in more detail before the next section. In any case, these difficulties disappear if we require causality of explanations as it is necessarily a one-way direction. After all, causality is a relation between entities separate in time and naturally time flows in one direction only. ‘The flagpole-example’ is a problem which stresses the need for asymmetry in explanations. An explanation simply cannot work both ways which means that it is not acceptable to place a condition of *explanans* in *explanandum* and *vice versa*. The other similar example mentioned in Section 1.1 was ‘the barometer-example’ which, on the contrary, includes a causal law. However, these kinds of complications are typical for the idea of Humean causality as it notices only regularities in the manifestations of events without paying attention to their actual relationship. As we know, regularity alone does not tell much about the nature of the connection between two entities. In the very example, the same thing which caused the drop in the barometer caused the storm, yet between these two events there is no ‘true causality’ to be found. Anyway, these problems go away if we only take the notion of ‘true causality’ or ‘making to come about’ seriously and require causal laws to tell more than just about a constant succession.

Proceeding to the second problem mentioned earlier on page 20, we will find the idea of ‘true causing’ truly helpful, albeit not capable of solving the issue once and for all. On the other hand, we do not even have to go that far to get rid of the counterexamples in question. As Hume himself already realized saying that “[...] *if the first object had not been, the second never had existed*” (see pages 22–23), causality is usually supported by counterfactuals.<sup>14</sup> In brief, this means that the following event would not have happened without the first one and this fact makes the appearance of the first one *relevant* for the existence of the second. Because the latter counterexamples were precisely problems of irrelevance, stressing the relevance of the causal laws with the requirement for the back-up of counterfactuals makes those counterexamples miss their target. It is obvious that the salt in the first example would have dissolved in water even if it had not been hexed, and that in the second example the man would not have got pregnant even if he had not been taking the birth control pills. In short, the support of counterexamples tells us pretty clearly which factors are relevant enough to be mentioned in explanation. However, even with the incontestable benefits they provide, counterfactuals should not be overestimated. By founding the notion of causality only on the support of counterfactuals, one will end with nothing but new problems. As Jaegwon Kim (1973b) shows, there are a great number of cases which are supported by

---

<sup>14</sup> The most influential work on counterfactuals is without a doubt David Lewis’ famous “Causation”.

counterfactuals, but where there is no causal connection to be found. One of Kim's (ibid. 570) clever examples is the following: "if yesterday had not been Monday, today would not be Tuesday". On the other hand, there are also a lot of imaginable cases of over-determination where one effect has more than one equal but still completely independent cause, and in these cases the use of counterfactuals merely misleads us to believe that they are both irrelevant. Similarly, saying that a burning cigarette was the cause of a fire does not necessarily mean that the house would not have burnt down without it. In the end, it is possible that lightning struck the house afterwards and thus the house would have burnt down in any case. Nevertheless, the thinking through counterfactuals is one highly important piece of equipment in the process of finding and proving the existence of a causal connection.

In my view, the difficulties with relevance cannot be fully solved with any improvements in the notion of causality or in its definitions. As we have seen, thinking through counterfactuals is one useful instrument, yet it is insufficient to do all the work needed. In his book *The Scientific Image* (115), Bas van Fraassen illustrates this idea with an illuminating example. A man called David has an alarm clock which goes off at seven a.m. and immediately after that he wakes up. Now it seems plausible to claim that he would not have woken up if it was not for the alarm. Obviously, one can never know these issues with absolute certainty, but that is always the case with causality as well as with all other things too. Nonetheless, it seems right to assert that the alarm was the cause of the waking. However, even though that seems suitable, we could still find similar support for thousands of other things. For instance, noting that David would not have woken up if he had not first gone to sleep is just as correct, but saying that his going to sleep was the cause of his waking would be weird. On the whole, in these kinds of cases we should lean more on causal language than on the support of counterfactuals. By following the established causal language, we are able to distinguish appropriate explanations from false ones, but even the causal language has its limits. All in all, I would say that, in the art of explaining, it is hard to define any strict rules. Next I will try to explicate this conclusion.

When we are contemplating the world through its countless properties, we see an endless amount of connections between them. Some properties are relevant factors in the process of understanding other properties, and this net of causality continues to the infinite. As Immanuel Kant (*CPS*) argues, causality is not in the world-in-itself but an innate piece in our process of understanding the world. Since we never explain things-in-themselves or even pure phenomena in front of our eyes, everything is about properties and these properties are



necessarily relative. As explaining is usually about finding the relevant factors in the causal net, we should ask what defines the relevance of a property. In my view, a whole lot of wisdom lies in this question, and thus it is a pity that it is not one to which answering could be considered possible. Of course factors like the support of counterfactuals and the appropriate use of causal language are essential elements in causal connections, but they are not the only ones. It is easily forgotten that, for example, cultural factors are also of the utmost importance in explaining. By this I mean that what is considered relevant to an explanation may vary a great deal if it is given to, say, aboriginal people instead of us. Nonetheless, perhaps even more attention should be paid to the roles of the framer of the why-question and the one who answers it. What are their points of interest? If we agree about the idea of explanations being answers to why-questions, it is evident that all the aspects in the questions are important. What property is highlighted? What is it that the questioner is looking for? Talking about, say, the murder of some poor fellow, the explanation varies according to the property highlighted in the question. For instance, the question about the cold-bloodedness of the murder seeks certain specific property in the causal story. The same goes for, say, a question about the location of the unfortunate event. Nevertheless, sometimes explaining requires more interpretation. Asking why the man in question died in the first place one can be answered in several ways which all would serve well in a particular context. For example “because he was shot”, “because he got a hole in his chest wall” and “because he had had an affair with a gangster’s wife” are all good explanations for different contexts. According to movies, the mobster himself would probably answer that the unlucky guy died because he had very little respect for another man’s property. All in all, the level and the nature of information sought should be known before giving an explanation. To give one more example, the nature of the question about the causes of some disease may vary according to the class and the level of expertise of the questioner as well as of the explainer. Sometimes it is appropriate to point out the living habits of the carrier, whereas sometimes the discussion should be turned to more fundamental elements, say, to cell biology. Usually, the nature of explanation sought is determined easily, but in some cases the question needs to be specified. This has already been acknowledged among philosophers as the conversation about the role of contrasts has emerged.<sup>15</sup> The use of contrasts means the specification of questions by highlighting the crucial aspects in them. For example, instead of merely asking why the bicycle was stolen, one may ask why it was stolen instead of a car. All in all, asking

---

<sup>15</sup> For example, Christopher Read Hitchcock (1996) has written a clarifying article about the use of contrasts in the search for explanations.

questions and answering them is only one mode of communication, and therefore the same skills should be applied to them. Clearness in articulation and carefulness in receiving information are good examples of the factors which set the scene for mutual understanding.

Taken into account all of those factors which might influence the validity of an explanation, it seems evident that giving any specific conditions for it is totally impossible. Clearly, there cannot be any general rules for explanation according to which it would be possible to find the relevant properties from the causal net in all situations. The best we can do is merely to list some features of a valid explanation and leave a whole lot of space for improvising. In the end, there is an endless chain of causal connections sprawling in all directions behind every property, and hence naming the relevant factors is one complicated matter. As Bas van Fraassen (1980, 130) emphasizes, there cannot be any questions or answers without a context and, since the context defines the explanations demanded, we cannot give any precise definition for a valid explanation from outside the context. He (*ibid.*) illustrates this idea in a fairly creative manner. As he notes, it is often thought that some Omniscient Being, say, God should be able to give fully complete explanations, since He always knows everything by definition. Therefore, it seems that being trapped in contexts only manifests the limits of human knowledge as it betokens that we are able only to grasp one side or aspect of the wholeness at a time. In any event, as van Fraassen notices, this whole attitude is clearly mistaken. If God did not have any specific interest and if he did not consider the world through certain properties, there would not be any why-questions directed at Him, and hence He would not have any explanation at all in a sense. He would simply have pure knowledge without any form. On the other hand, if He had a viewpoint, His questions and answers would be exactly as dependent on the context as ours. As van Fraassen (*ibid.*) concludes, “[i]n either case, [H]is advantage is that [H]e always has all the information needed to answer any specific explanation request.” However, Van Fraassen adds that “[...] that information is, in and by itself, not an explanation; just as a person cannot be said to be older, or a neighbor, except in relation to others”. In order for something to be an explanation, it needs to be an explanation of something, and that means that an explanation is always tied to some context to begin with.

### 1.2.5 Causality and Explanation

Before proceeding to scientific mechanisms, there is one more matter to settle. One major issue in the philosophy of science has been the question of the possibility of non-causal explanations. So far I have been associating explanations with the notion of causality, but even though that is what is mostly in question, some philosophers have also wanted to make room for different kinds of explanations. Surely, it would simplify the matter beautifully to define the idea of explanation with the requirement of causality, but the price is still not worth it if it lures us away from reality.

The issue in question is interesting, since most philosophies seem to be quite uncertain about it and therefore unwilling to say anything definitive. As James Woodward (2009 [2003], 8) stresses, virtually everyone admits the crucial role of causality in scientific explaining. Nonetheless, there does not seem to be consensus about whether it is necessary for it. At the very least, it comes across as too harsh to deny the possibility of non-causal explanations for good. For instance, in the preceding section, I mentioned that Hempel avoids binding the idea of explaining to the one of causality. In his “Aspects of Scientific Explanation” (352), he declares his own view according to which causal explanations are not the only types of explanations there are. As we know, instead of causality he wants to give the main role in explaining to logic. To present an example of non-causal explanation, Hempel mentions cases which lean on “the laws of coexistence” instead of “the laws of succession”. In one particular example, the time taken for a pendulum to make a full swing is explained by pointing out the length of a pendulum as well as the law that governs the matter. All in all, more common patterns of non-causal explanations are teleological and functional explanations. The first one is often associated with the Aristotelian world view although Aristotle (*Ph*, II, 3) himself acknowledged the existence of causal explanations as well. The idea of teleological explanation is to explain something not by stating antecedent conditions but by pointing out the goals or purposes of a creature or even of an inanimate object. In modern sciences, teleological explanations are used mostly in biology and social sciences. After all, it is fairly natural to explain human action by referring to the agent’s purpose. For the second, functional explanation is a model of explanation in which some activity or even the existence of an entity is explained by its function or role. Perhaps the best known example

can be derived from biology and it is the activity of heart which is, of course, the pumping of blood and which can be explained by heart's function to deliver oxygen throughout the body.

While there evidently are proper scientific explanations which do not refer to any causal connection, I believe it would clearly be wrong to claim that all scientific explanations are causal explanations. However, in my view, the role of causality should not be undermined in consequence. In the end, all the previous explanations are posited on the image of an underlying causal net.<sup>16</sup> For instance, in Hempel's example of pendulum it is obvious that in order to really explain a particular pendulum's movement at a certain speed, an initiatory force understood as a causal factor is needed. Besides, as we saw in the preceding section, without the underlying idea of causality we may end up with highly problematic explanations which lack a sense of direction. Also in teleological and functional explanations, the causal picture cannot be forgotten as these explanations can hardly be held scientific if they are considered as working on their own somehow. If we take the preceding example of heart's function under scrutiny, the point can be easily demonstrated. Undoubtedly, it would be ridiculous to claim that heart is really doing something only because it is its function and that we do not need any causal story to back it up. On the contrary, the whole functional picture leans on the idea of the causal story which causally explains the functionality in question. In the final analysis, the functional picture of the activity of heart is merely the same story only told differently in order to make it easier to follow. In some cases it is more natural to outline the explanation from a different viewpoint, but if we ever want to consider functional and teleological explanations as proper scientific explanations, we also need to carry the causal story with us all the way. Anyway, I will take a closer look at teleological explanations in the second chapter of my work.

On balance, even though a proper explanation can take many forms, I believe that at bottom we are interested in the causal picture and that all the explanations there are need to be fit in this very image. This is why I dare to associate the idea of explanation with the idea of causality. After all, explaining in its most normal sense is specifying antecedent conditions working in accordance with some general laws or rules in making the *explanandum* come about. Consequently, before closing this study of the captivating notion of causality I would like to cite the four general terms for scientific explanation posed by van Fraassen (1980, 124). I believe that in these terms the idea of explaining is portrayed in a fairly felicitous way:

---

<sup>16</sup> These thoughts were originally expressed by Larry Wright (1976).

- (1) Events are enmeshed in a net of causal relations.<sup>17</sup>
- (2) What science describes is that causal net.
- (3) Explanation of why an event happens consists (typically) in an exhibition of salient factors in the part of the causal net formed by lines ‘leading up to’ that event.
- (4) Those salient factors mentioned in an explanation constitute (what are ordinarily called) the *cause(s)* of that event.

## 1.3 The Need for Mechanisms

### 1.3.1 The Problematic Nature of Laws

In Section 1.2 my aim was to stress the importance of the idea of causing in the conventional sense as well as to defend it against Humean mistrust. Thereby, I tried to show that a right kind of notion of causality is indispensable for our concept of explaining. Two famous problems related to the Hempelian account of explanation were presented in Section 1.1 and, as we saw, at least the one with directional aspects was solved with the right kind of notion of causality. The other problem, namely the one with relevance issues, I concluded to be something which we just need to learn to live with, although we can still find some tools for making the practice easier. Anyway, now we have new questions to ask. First and foremost, is this enough? In other words, is the adding of the requirement for ‘true causality’ enough for the Hempelian model of explanation to work properly? Is it possible to rewrite all proper scientific explanations in a form of arguments where the conclusion is deduced from the premises including one or more universal laws? In my view, an affirmative answer to these questions does not sound right. In this section, I will try to explicate why.

To begin with, I am not willing to say that Hempel’s idea is all bad, quite the opposite in fact. If we take, for instance, Newtonian physics under review, any problem will not seem

---

<sup>17</sup> Once more I would like to emphasize that I myself cannot see any reason why other kinds of properties could not be in a causal relation as well.

to appear. In fact, the Hempelian model seems to be perfect for as highly organized a structure as the image of the world is in classical mechanism. When every event follows so unquestionably from the preceding conditions in the guidance of established laws, every explanation can, or even should, be formulated as an argument exactly in the way Hempel thought (at least this is so in simple cases). Anyhow, the problem lies in the fact that science is not only about Newtonian physics. There is even the temptation to say that in most disciplines the case is much more complex. The instances which first come to mind are biology and social sciences. One should wonder whether there really are such laws on which the argument-form explanation can hinge within the given fields. For example, human beings are complicated to the extent that the establishment of exceptionless laws governing their decisions is rather difficult if it is even possible. When the image of the world gets blurrier and more complicated, the possibility of finding universal laws and forming explanatory arguments seems to glide further all the time. Nevertheless, the urge to explain does not vanish by the same token.

Before proceeding any further, I should say something about the laws of nature in general. To define the term 'law of nature' has turned out such a difficult task that it has become one of the central issues in the philosophy of science. After all, there are not many concepts as crucial in the realm of science. As is often the case, a concept which we use frequently and which normally comes across as fairly lucid may turn out to be a mighty philosophical problem when someone asks us to give a precise definition of it. In the end, knowing how to use a concept does not require the ability to specify it. However, my aim in this section is not to try to solve the problem but merely to outline it. Within the bounds of my study, it is sufficient only to realize the intricacy of the issue.<sup>18</sup>

According to the standard view, three kinds of facts can be sorted out (Lange 2008, 203). First of all, there are *logical* or *metaphysical necessities*. These facts are always valid, because it is simply impossible for things to be another way around. The most used example of logical necessity is the fact according to which a triangle has three sides or three angles. The other two types of facts are *contingent*. For the first, there are *nomic necessities* and for the second, there are *accidents*. It is quite plain what accidents are. They are facts not supported by any kind of law like, for example, the fact that I wear a blue shirt today (even if there was a law behind it, it is one about which we do not know a thing). Conversely, nomic

---

<sup>18</sup> One good overview of the discussion is John W. Carroll's "Laws of Nature".

necessities are facts which spring from laws of nature. However, they still differ from logical necessities in a sense that they are not absolute. Even though laws of nature by definition hold in this world where we live, they are not logically true which means that it is possible to imagine a world where different laws prevail. That is to say, nomic necessities do not carry their truth along with them. All in all, laws of nature are generalizations which necessarily hold in our world but where the necessity is world-dependent. In any case, one notorious problem with laws is the very question of how to distinguish nomic necessities from mere accidents. Usually, the difference is quite clear. While accidents are things which solely happen, nomic necessities require some kind of understanding of *why* they are necessarily true. To use common examples, the fact that “there are no gold cubes larger than a cubic mile” holds accidentally, whereas the fact that “there is no uranium-235 larger than a cubic mile” is a nomic necessity. The first one is (at least with extremely high probability) true, but there is no reason for why it could not be so. On the contrary, the latter one holds because it is “written” in the laws governing nuclear chain-reactions. However, this is not always the case, and therefore there has been effort to make more precise definitions. In some situations where we do not have the knowledge required, laws are formed reckoning solely on a fact about a constant manifestation. Naturally our understanding has its limits, and hence on the most basic levels we have data only about what is observed.

Again it was Hume who sowed the seed of skepticism in the idea of induction. Actually, this skeptical thought is a natural conclusion of the ones presented before. In Section 1.2.1, I showed how Hume denied the possibility of the perception of necessary causal connections. Furthermore, as Hume (*EHU*, V, I, 36) rightly believed, these kinds of connections cannot be any better attained by logical reasoning. Therefore, the natural conclusion was to state that necessary causal connections are somewhere totally out of our reach. Now, getting back to induction, I believe it is justified to say that the idea of necessary connection is the core of valid induction. If we are to make precise predictions about future states, the only way to make those in any reasonable manner is to rely on necessary causal connections. Nevertheless, this train of thought turns void if we take the credibility away from them. The question goes as follows: “Even if we know that in every preceding case of A, B has come about afterwards, how can it be guaranteed that B will also follow the next A?” If there is no necessity to be found in singular connections, how are we able to know that these connections will also hold in the future? How are we allowed to claim that we know for sure that the sun will rise in the morning or that the next cup of coffee will also contain

caffeine? Anyhow, as was the case with causality, it seems that we are pushing the boundaries of knowledge too far here, and continuing in this direction will result only in losing the whole notion of knowing. Even Hume himself admits this. Actually, he (ibid.) goes as far as stating that, without relying ourselves on the idea of induction, we will even lose the ability to act for good. After all, induction always springs up from custom and habit and, as Hume emphasizes, “[c]ustom is [...] a great guide of human life”. However, as it was the case with the notion of causality, we should be highly grateful to Hume for showing the thinness of the foundation of our knowledge. Nonetheless, this does not mean that we should attempt to avoid inductive reasoning in our everyday living. Arranging a day for a picnic, we may add the clause stating that we will not go if the sun is not shining, but saying that we will not show up if the sun has not risen at all would be a waste of everybody’s time.

Now, reverting to the language of properties once again, we should at least try to find a place for the notion of law of nature. Naturally, a fact is a true property. Returning to the three types of facts, one should argue that the first ones, namely *logical* or *metaphysical necessities* are properties which can be logically deduced from other properties. There again, we may state that *accidents* are properties whose carrier, say X, could as well not have them, whereas *nomic necessities* are properties where there is no contingency. Now, by proceeding precipitately, one may infer that laws of nature must be relations between universal properties, instead of particular properties, and thereby state that this is the definition of law.<sup>19</sup> This would clearly be a mistake. In my view, it is correct to say that a law of nature must concern universal properties, since postulating a law between particular properties would merely make the whole concept ridiculous, but the problem is that it is perfectly possible for similar relations which are not laws of nature to exist. Actually, it is solely the case of the recurring of the earlier problem. There definitely are cases where all the particular properties of a kind are in relation with some other kind, yet where the relation between them is not of the nomic type. As we know, in order for a relation to be a law, the nomic nature must be included in the nature of universal properties themselves, but since our knowledge about those natures is sometimes defective, in practice we have to content ourselves with laws derived only from particular instances. In any case, as I have noted before, my aim was not to give any precise definition of the concept of law of nature. I believe that, in my study,

---

<sup>19</sup> Likely, the most recognized proponent of the theory which regards laws of nature as relations between universal properties is D. M. Armstrong (1983). However, even though I criticize the idea that this finding would settle the whole issue of the problematic nature of laws, I think that Armstrong is right about the basic matter.



it is sufficient that we understand the concept in question at least on some general level. All in all, one may say that laws of nature are rules without exceptions. They are rules which our world has to obey, and therefore they are the basis of our explanations and predictions. Nonetheless, in the dynamic world of science, these laws cannot be taken to be everlasting, since old ones are proven wrong at the same time as new ones are found. Either way, the problem is that there are many disciplines where there are no laws of the kind to be found at all but where there is still enthusiasm to give explanations and predictions. For example, economics seems to be much about explaining and especially predicting, but still it feels fallacious to talk about universal laws within this field, even if the economic explanations and predictions might work in a truly successful manner. In order to talk about laws of nature we need rules which can be described as strict and infallible to the greatest extent and to which it would be difficult to find counterexamples even in theory. Unfortunately, in the field of economics this does not seem possible.

Even Hempel and Oppenheim themselves realized this particular problem lying behind their view of explanation. In “Studies in the Logic of Explanation”, they raise the question, but nevertheless it does not have the impact to make them lose the faith in their cause. On the contrary, they (1948, 251) state that, even though their study is based on physical sciences, the same general principles can be applied to other fields as well. They admit that, for example, not *all* instances of human as well as other animal behavior can be explained by referring to laws of the same precision and generality as in physics or as in chemistry. In spite of that, they still believe that the general character must be similar in all instances of scientific explanation. Even if explanations in, say, psychology, biology and social sciences lack some power because of the incompleteness of their laws, Hempel and Oppenheim (ibid. 252) appear to believe that, in times to come, these laws can be formulated with such skill that problems mentioned will little by little vanish.<sup>20</sup> All the same, decades have passed since the publication of their paper, and the kind of belief they share has become more and more exceptional if not disappeared for good.<sup>21</sup> After all, within the realm of social

---

<sup>20</sup> After presenting an example of some old crisis in the United States’ cotton industry, Hempel and Oppenheim (1948, 251–252) conclude that “[s]uch laws cannot be formulated at present with satisfactory precision and generality, and therefore, the suggested explanation is surely incomplete, but its intention is unmistakably to account for the phenomenon by integrating it into a general pattern of economic and socio-psychological regularities.”

<sup>21</sup> One I can name is Karl-Dieter Opp. In his paper “Explanations by Mechanisms in the Social Sciences. Problems, Advantages and Alternatives”, he defends the view according to which all the explanations must include at least one law as it is required in the DN model. The main idea appears to be to loosen the conditions for the concept of law in certain disciplines. This approach, however, comes across as highly questionable. Furthermore, all the valid explanations do not meet even the loosened conditions since mechanisms may contain

sciences, for example, it seems to have become increasingly clear that the belief in finding actual social laws is not a very plausible one. Moreover, even in the trusted field of physics new questions have arisen. Modern understanding about the physical world differs greatly from the classical picture, and therefore it sounds reasonable to presume that the explanatory equipment might also need some refreshment. All in all, it appears evident that not all explanations can be arguments of the type of the DN model or even apt to be formulated as such. This, however, is not to say that there are no perfectly well working DN explanations, but to stress that something more is needed in order to cover the whole scientific ground. The next two sections are devoted to mechanistic explanations which certainly have an important role to play in this picture. On the other hand, whether even they are enough for the process is once more a different question.

### 1.3.2 The Concept of Mechanism

I guess that we all have some kind of idea of what mechanisms are. However, as I may be mistaken, we should spend a moment on thinking about catapults or locks. A catapult is a good example of a mechanism in all its simplicity, whereas locks are more complicated but, nonetheless, more common to our everyday experience. So, what features are shared by these two instruments? My answer would be that in both cases there is an event which launches a process where separate parts of the whole react in a way that a new event is produced. If we take, say, locks under scrutiny, the issue becomes more evident. For instance, when I place my key into my office's lock and turn, I hear a click and the door opens. To be honest, I cannot see what goes on inside the lock, yet I can imagine some cogs and wheels through which the causal process proceeds, opening the door in the end. One interesting feature in these mechanical mechanisms is that the parts which constitute the whole are there all the time waiting for somebody to launch the process. Thus, you can make them go through it all over again as many times you want (or until the mechanism breaks down). However, Stuart

---

true randomness. A kind of similar idea can be found in Pierre Demeulenaere's article "Causal Regularities, Action and Explanation". The problem there is that he seems to misread Hempel in order to justify his own claim according to which mechanisms are not that far from the DN model, in the end. To my mind, he is right in that there are regularities needed in mechanisms and that these regularities resemble laws of nature, but still I believe that it would be excessive to relate them altogether.

Glennan (2008, 376) also distinguishes another type of mechanism which is clearly more general within the scientific context. The kind of mechanism is constituted of temporally extended processes where the sequences of activities produce an outcome. The difference is obvious. While “the mechanical mechanisms” are continuously in a standby position to produce the sought-after effect, the latter ones are unique in a sense that they can only happen once. Naturally, these mechanisms also have more than one manifestation, but the structure can never be the same, since the whole is constituted of temporally extended processes instead of constant particles. Glennan’s (ibid.) own example is photosynthesis. There, factors like water, carbon dioxide and energy from light go through certain activities through which oxygen and sugar are produced. As Glennan states, photosynthesis can be conceived of as the activities of a system called chloroplast whose operation is a mechanical process. Anyhow, even though mechanisms in science are usually of the latter type, nothing implies that there could be no mechanisms of the first type in science. On the contrary, one can easily imagine, say, a social mechanism where all the actors stay put merely repeating the same process all over again.

Jaakko Kuorikoski (2009) has brought out a kind of similar sorting as Glennan. He argues that, in the end, the concept of mechanism consists of two separate notions. First, by mechanism one can refer to a strict componential causal system where all the parts have their given function, location and causal role. Second, by mechanism one can as well refer to an abstract form of interaction which is highly flexible and unordered but still has some regular features due to which one is allowed to call it ‘a mechanism’. As Kuorikoski (ibid. 144) notes, even if this distinction may easily be associated with the distinction between natural and social sciences, it is something which we should avoid. Naturally, the mechanisms of the first type are usually found in disciplines like molecular biology and neuroscience, whereas those of the latter type are more common in social sciences, nevertheless a great amount of exceptions can be cited. For example, a mechanism of evolution by natural selection is a typical example of an amorphous mechanism in natural sciences, whereas in, say, economics it is possible to find remarkably strict and organized mechanisms.

However, even with all the diversity among mechanisms, there must also be some factors to be found in all mechanisms encouraging people to use one and the same concept. Evidently, there must be some factors on which philosophers can rely while making attempts to give a precise definition for the concept of mechanism. Before analyzing those factors one by one, I will cite three famous definitions given to the concept of mechanism. Probably the

most used comes from the classic article “Thinking about Mechanism” by Peter Machamer, Lindley Darden and Carl F. Craver (2000, 3):

Mechanisms are entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions.

Another one worth mentioning comes from Glennan (2002, 344):

A mechanism for a behavior is a complex system that produces that behavior by the interaction of a number of parts, where the interactions between parts can be characterized by direct, invariant, change-relating generalizations.

Finally, a definition by William Bechtel and Adele Abrahamsen (2005, 423):

A mechanism is a structure performing a function in virtue of its component parts, component operations, and their organization. The orchestral functioning of the mechanism is responsible for one or more phenomena.

As the definitions above suggest, mechanisms are always complex structures, which means that no singular unit can be a mechanism, unless we divide it into separate smaller parts. Furthermore, even though the parts of mechanisms themselves might be of the utmost importance, the relations between those parts always carry more weight. That is to say that only a bunch of entities without any relational activity can never be a mechanism. Nevertheless, even if mechanisms always consist of parts, they can be considered as units as well. After all, mechanisms are independently working systems which can be classified according to the function they serve. As Glennan (2008, 377) emphasizes, “[m]echanisms always do something, and we identify a mechanism by first identifying the behavior it produces”. The activity is the essence of a mechanism and, for the changes it produces, the inner activity serves as an explanation. Interestingly enough, the very possibility of considering mechanisms from two different aspects, namely from the macro- and micro levels, entails that mechanisms are hierarchical. This means that getting deeper and deeper into a mechanism normally reveals new smaller-scale mechanisms. A mechanism as a unit can always be an active part of another mechanism and so forth. Finally, one crucial thing with mechanisms is the regularity of their activity and production. In order to call a thing ‘a mechanism’ there must be some features which are repeated every time it goes through. I will

submerge myself more profoundly in this issue when we get to the next section concerned with the explanatory side of mechanisms.

All things considered it seems apparent that, when talking about mechanisms in general, one needs to observe great caution. Under the concept of mechanism lies a massive variety of systems which all differ from each other on account of structure, nature and function. Therefore, the definitions must be laid on a highly abstract level in order to cover the whole variety of instances. For example, it sounds reasonable to presume that physicists and sociologists have very diverse views about mechanisms. Nonetheless, the common features binding all mechanisms together are also there to see. All mechanisms are active, more or less regularly working aggregates which are studied through their parts and the relations between those parts. The crux of the whole mechanistic way of explaining is to approach *explanandum* by stripping down the phenomenon and thus to uncover the underlying causal picture. That is to say, the events on a macro level are explained by those on a micro level. Next I will wrap up the whole chapter by illuminating this novel way of explaining.

### 1.3.3 Explaining with Mechanisms

After realizing the insufficiency of the traditional DN model for scientific explanations, people have started to search for new types of patterns. This, of course, concerns mostly philosophical discussion about adequate explanatory approaches. In the realm of practice, mechanisms have been used for ages in the processes of explaining and predicting. During the actual practice, one very seldom wonders about the theoretical aspects of one's doings. On the contrary, the only thing that matters then is to present the issue so that it captures the essence in as accurate, comprehensive and usable a way as possible. Whether the best tool at hand is a deductive argument or a mechanistic sketch depends on the situation, and therefore certain room for improvising must be left. After all, in the history of science, people have given mechanistic explanations even a good deal before the whole concept was invented. Celestial mechanics is a good example since the proper explaining of the movements of celestial objects requires references to the mechanisms behind them. Either way, the mechanistic way of explaining has usually been considered a long-awaited alternative to

those which seem to rely too heavily on laws of nature. As I tried to represent in Section 1.3.1, even though the concept of law of nature is quite problematic in itself, the biggest difficulties lie in the fact that there simply are no laws of nature to be found in every discipline in which, nevertheless, explanations are traditionally sought and given. While scientific mechanisms do not have to rely on laws (although some amount of regularity is usually demanded), the popularity of the mechanistic aspect to explaining has increased rapidly. This can all also be seen on a larger scale than only as a change in explanatory methods. Roughly speaking, while science in the past seemed to be most about finding and defining laws from the surrounding nature, nowadays somewhat less is required of a valid explanation. In order to explain, a scheme does not have to imply anything too general as long as it helps us to understand the phenomenon in question. This is why mechanists often stress that their approach better resembles the actual practice of science. This process of change has other bearings as well. As Glennan (2008, 383) states, “[r]ecent mechanists see the mechanisms movement as part of a larger trend in which philosophers of science have ceased to think of theoretical physics as the paradigm of science”. Anyway, one good question is whether there is even any ideal way of explaining, particularly after what has been said about the obscurities of the matter. In different contexts different kinds of equipment and language are needed, and therefore one must be as open as possible about the issue in question.

As a recapitulation, I will say a few more words about mechanistic explanations. As probably would be the easiest way to explain the activity of a clock, one might take off the back panel and uncover the mechanism ticking inside. The basic idea is the same for scientific mechanisms, merely in a more complicated manner. To explain, say, a war or a neurological process, one must divide the whole into smaller entities and activities, and thereby show how the change on a macro level is composed of changes on a micro level. To understand why a property can bring about another property, one needs to reveal the causal chain lying behind the bigger picture. Evidently, mechanisms do not only answer why-questions as they may also be appropriate answers to the questions of how. By giving descriptions of processes as wholes, they may provide a huge amount of answers to possible more specific why-questions by the same token. To use the terminology of Jon Elster (1989), I would say that mechanisms explain phenomena by opening up black boxes and by uncovering the cogs and wheels or the nuts and bolts of the underlying machinery.

Mario Bunge (1997, 439) suggests only three necessary requirements for a mechanistic hypothesis or any theory to be taken seriously in modern science or technology. Mechanisms need to be *concrete*, *scrutable* and *lawful*. The first two conditions go hand in hand. Mechanisms must be *concrete* in a sense that they cannot be immaterial. To give an example, Bunge (ibid. 418) states that classical psychology can only give answers to the questions ‘what’ and ‘when’, whereas neuropsychology can answer the questions ‘where’ and ‘how’. By offering mechanisms, neuropsychology can offer explanations which are more powerful and somehow more “real” on their higher level concreteness. This, however, might come across as quite strict, since social mechanisms, for instance, quite often rely on more or less abstract entities like desires, beliefs and intentions. Nonetheless, social mechanisms as wholes are something of which one is able to take a firm grasp, and perhaps thus they can be thought to be concrete enough to have credibility also in Bunge’s eyes. Concreteness also guarantees that wholly abstract structures, like mathematical algorithms, cannot form a mechanism by themselves. By *scrutability*, Bunge secures that mechanisms cannot be grounded on anything supernatural. As we are talking about science, one must be capable of proving mechanisms empirically. The most interesting condition is, in any case, the last one, namely *lawfulness*.

As the definitions on page 52 suggested, the idea of regularity or lawfulness, as Bunge puts it, is quite crucial in the search for mechanisms. In order for an explanation to have any kind of explanatory or especially predictive power, it must be repeatable and thus founded on some known general principles. Needless to say, a mechanism which contains, say, a miracle is not something we should take seriously. However, that does not mean that mechanisms must be absolute, although naturally nothing prohibits them from being that. The point here is of the utmost importance. Mechanisms can be, for example, probabilistic which means that they produce certain behavior with certain likelihood. Furthermore, mechanism does not even have to outline any certain probability since the process may go through purely randomly. As I concluded with Anscombe in Section 1.2.1, the idea of causality does not depend on the idea of necessity. Therefore, since mechanistic explanations rely only on the causal picture, something does not have to happen by necessity in order for us to explain it. Yet, by revealing necessary relations, we make explanations stronger. Instead, what matters is to understand the course of mechanism as thoroughly as possible. For instance, even if a mechanism contains awfully many conditions which make it fail more often than succeed, understanding *why it fails* when it does can make the mechanism a useful piece of

explanatory machinery. Even pure randomness can be allowed at some point of a mechanism, assuming that the unity bears it. Therefore, the idea of mechanism should not be confused with the idea of law of nature. In order to be explanatory, a mechanism does not have to be unexceptional or as transparent as a glass of spring water. Within a certain context and certain conditions, some factors are more important than others, and thus we should keep our eyes on what matters. Besides, even when our knowledge is insufficient, what is known may still be helpful.

This all leads us back to the more fundamental issues in explaining discussed before. In Section 1.2.4, I concluded that the problem of relevance is included in explanatory issues by nature. Contextual matters are an intrinsic part of all discussion to which the category of explaining also belongs. All discussion take place within a context, as the idea of discussion above or outside all contexts simply sounds nothing less than absurd. Since mechanisms are explanations, they are also answers to why-questions, and hence the property highlighted in the question settles which mechanisms there need to be brought out. Citing the example given by Glennan (2008, 377), “the mechanism that delivers blood to the brain will include the heart (and its parts) as well as a system of arteries, capillaries, and veins, while the mechanism that produces thumping in our chest will require a different description of the heart and will not include parts of the circulatory system outside the heart”. So, as Petri Ylikoski argues in his paper “Social Mechanisms and Explanatory Relevance”, the mechanistic approach does not get us any further from the basic problem of explaining, namely from the problem of relevance. This, all the same, should not be considered as a fault in the system since the problem belongs to the core of causality and explaining, instead of being particularly a problem for mechanists. As all the other kinds of explanations as well, mechanisms are always bound up with a certain discipline as well as with a certain context, and therefore one should be extremely suspicious of all the interdisciplinary mechanisms which are intended to cover too wide range of phenomena. This is also the reason why Bunge (1997, 439) warns us about all “the grand theories” such as the Romantic philosophy of nature, Marxism and rational choice theory all of which can be seen as too universal answers to too universal questions.

All in all, the strength of the mechanistic approach is the very room that it gives to scientists. With all the flexibility the idea of mechanism contains, it is assured that mechanisms are truly needed in the world of scientific explaining. Of course, the flexibility also involves its disadvantages as the leeway given comes with great liability. For example,



finding the right level of generality and abstraction is always a challenging matter. One must deeply contemplate to how wide a range of phenomena the mechanism in question is applicable. While with more generality the coverage also increases, more specific mechanisms usually share more explanatory power. However, since mechanisms are often quite particular in all the complexity of their structure, the possibility for wide-ranging use can be very little. Therefore, one should remind oneself that, from certain specific mechanisms one should not be too eager to make overly general statements. Anyhow, even if mechanisms are definitely an important improvement to the explanatory equipment we have, in themselves they are not capable of solving any more fundamental problems about explaining. All the skill that it is needed in the process of giving good explanations is nonetheless needed in mechanistic explaining as well. Perhaps even more is needed since more room in explanations entails more possibilities for no-good explanations.

In this chapter, my aim has been to enlighten my view of explaining and, more generally, of talking about the things going on around us. Through introducing the Hempelian model for explaining, I have tried to defend the idea that something more is needed in our scientific explanatory equipment, namely the mechanistic approach. Now I will lay mechanisms aside for a while as the next chapter will be all about the nature of actions. However, I will return to mechanisms during the second part of work. After all the work done with more theoretical issues, there I will take mechanisms under more concrete scrutiny.

## **2. The Multitude of Actions**

### **2.1 What Actions Are?**

The first chapter of my study was mainly about scientific explaining. There my aim was to introduce my esteemed readers to the idea of mechanistic explaining, and this I wanted to do by showing the undeniable need for the idea in question. As I stated in Section 1.3.1, one crucial reason for the need is the lack of laws of nature in, for example, social sciences. Generally, the problem with social sciences is that the social world is complicated to the extent that it does not seem to obey any precise laws. However, in spite of the apparent lack of laws, I believe that it would be closer to truth to say that the laws there may be are too complex to find or define. As the reality consists of various levels, the laws governing one level may not be able to be stated in another, even though the impact of the laws themselves naturally do not vanish when moving from one level to another. Anyway, I do not want rush ahead of things, since we still have much more elementary questions to ask.

On the whole, this second chapter is all about actions. I must apologize for all the inconvenience one may feel about this bouncing between different topics, but as was said earlier, the foundations must be laid before anything else. Thus, now it is the time to contemplate the various aspects of action. I believe that only through a profound inspection may we find an answer, among others, to the question of why there are no laws of nature to be found within the realm of social life. However, before proceeding any further, we should start with the simplest question there can be: what actions are in the first place? As is often the case in philosophy, the simplest questions turn out to be the trickiest ones and, in the end, nothing is harder than defining the most common terms we have used since we learned the language for the first time.

One pretty universal idea of actions is that they are a subclass of events. That is to say, all actions are events but not all events are actions. In my view, there is a remarkable amount of truth in this idea. If there is an action, there is also an event going on. Now, if we are willing to state that actions are events, we need to recall what was said about events

earlier in this work. In Section 1.2.2, I claimed that events are properties of something about which we cannot talk without referring to its properties. This can all also be said about more constant beings, but what is special with events is that they are always extended in time in a very unique manner. Even if, say, dogs and trees are also extended in time in a sense that their existence usually lasts more than a mere blink of an eye, their temporal extension does not have special significance. All constant beings last some particular period of time, but their existence is something which stays more or less the same through that period. If there is a dog, for example, it will be just as much a dog at every singular moment of time through its period of existence. However, when we are talking about events we are talking about something which lies in some period of time in the way that when dividing the period into separate moments we will lose the very event. This is why the temporal features of events are of such huge importance. Events are extended in time so that, instead of merely existing from one moment to another, they exist only in certain clusters of moments. Now, recalling that events are properties, we may conclude that events are properties which always go along with some property of temporal extension.

If we agree that actions as events are properties as well, we still need to figure out what special feature makes an event into an action. The first thing coming to mind would be to say that in actions somebody or something is doing something. Even if this does not come across as much help, there is perhaps the most important feature of action hidden in the paraphrase, namely that in actions there is always a subject. If an event is to be called ‘an action’, there must be *somebody* or *something* doing something. Naturally we associate this subject with human beings at first, because the most full-blooded actions are taken by us, but nonetheless that would be too hasty. After all, undoubtedly non-rational animals are also capable of doing things. Either way, if we take this a little further, we end up realizing that we cannot delimit the term ‘action’ merely by mapping the possible subjects. In the end, we may say that ‘a train is coming’ or ‘a computer is downloading’, and what would that be if not acting. As long as there is a subject to which an action can be attached, we have ourselves an event of acting. Presuming now that we cannot get any further with the subjects of actions, next we should think about ‘doing’. It is clear that attaching verbs to subjects does not necessarily generate actions. If we consider, say, a computer once more, it is reasonable to claim that, while it is downloading, the computer is actually doing something, but while crashing, something unfortunate is merely happening to the poor machine. The distinction seems explicit. In the first case, the action is coming from the subject, whereas in the latter

case the subject is more like an object of something. While crashing, the computer is playing only a passive role in the event. The same division goes for, say, a man who is killing and a man who is dying. Contrary to the first case where the subject is actually making the thing happen, in the latter case he is merely going passively through the process of dying. However, in the case of dying we may still say that a virus is killing the man, and therefore it is an instance of the doing of something, but nevertheless the man himself is not bringing about 'the dying'.

I am positive that my definition of action will arouse condemnation, since associating actions with trains and computers is truly against orthodoxy. In my view, the controversy is a good thing. After all, only through a kind of bafflement are we able to realize the requirement for further classifications for the term 'action'. So, although I acknowledge the confusion, I will stick with the aforementioned idea. Even if the talk about lifeless objects performing actions is somewhat dubious, I believe that Ludwig Wittgenstein (*PI*, 43.) is absolutely right when stating that "the meaning of a word is its use in the language".<sup>22</sup> In the end, in our everyday dealings we postulate actions to things that, in our eyes, do something, even though their doing may not be more than a kind of mechanical processing. For example, there is an essential difference between a train which moves by its own power and a train which moves because there is a bunch of people from the railway pushing it forward. Differences of this kind speak on behalf of the view I have defended. In any case, as I said, it is also important to notice that some further distinctions must be made in order to keep the picture accurate enough. Even if there are similarities among all actions, some crucial differences also exist. If we think about, say, a train moving and a dog eating a bone, we will find out that there is a very conclusive difference between them. Furthermore, if we continue by comparing the latter action with a situation where a representative of the human race tries to decide which wine to drink with a dinner, we will see a further difference. All in all, it seems quite obvious that, in order to appreciate the various aspects of acting, a deeper study must be carried out. Therefore, in the following sections, I will make the further classifications which I believe are indispensable for a sufficiently comprehensive understanding of actions. After that, I will see what kind of implications those classifications bear.

---

<sup>22</sup> On the other hand, he (*ibid.*) adds that, although this definition suits most of the cases, it still does not cover all of them.

## 2.2 Voluntary Actions

### 2.2.1 Aristotle and the Origins of Action

Aristotle is known to have written the first truly profound studies about actions which have survived to the present day. Naturally, their influence has been immense on the whole field of the philosophy of action. Since nothing has really changed on the basic level, I will use Aristotle's account as my stepping stone to the categories of action. Moreover, I believe that there are some highly crucial points in Aristotle's theory which seem to have got lost somewhere over the centuries.

Aristotle (*DA*, 414<sup>a</sup>, 30–32) divides the powers of living things into five different categories which are “the nutritive, the appetitive, the sensory, the locomotive, and the power of thinking”. However, one does not need to share powers from all five categories in order to share powers from some of them. For example, plants have only nutritive powers whereas human beings may be complex enough to possess powers from all five categories. Anyway, the one category we are interested in here is the appetitive powers. Aristotle (*ibid.* 414<sup>b</sup>, 1–6) states that “[i]f any order of living things has the sensory, it must also have the appetitive; for appetite is the genus of which desire, passion, and wish are the species; now all animals have one sense at least, viz. touch, and whatever has a sense has the capacity for pleasure and pain and therefore has pleasant and painful objects present to it, and wherever these are present, there is desire, for desire is appetition of what is pleasant.” Now, before getting into the species of appetite, I want to make clear why we are interested in the appetitive powers in the first place. The point surfaces later when Aristotle (*ibid.* 433<sup>a</sup>, 22) declares that “[t]hat which moves [...] is a single faculty and the faculty of appetite”. By this Aristotle is saying that all actions living things perform happen because of the appetitive powers. Naturally the nutritive powers can also cause movements like growth, but all self-movement springs necessarily from the appetitive powers. When some living thing shares appetitive powers, it means that it is by its senses vulnerable to various sensations, and by nature all sensitive beings seek good sensations and avoid bad ones. Therefore, all the striving forces there are to make living

things move themselves are the quest for good and the avoidance of bad. As Aristotle (*ibid.* 432<sup>b</sup>, 16–17) states, “for no animal moves except by compulsion unless it has an impulse towards or away from an object”. Earlier Aristotle declared that “for appetite is the genus of which desire, passion, and wish are the species”. As we can see, this tripartition is fully equivalent to the one made by Plato (429–347 B.C.) some time before. In *The Republic* (435b–441c), while comparing the structure of human soul with that of a city, Plato divides it into three separate parts which are the appetite, the spirit and the reason. Even though the names used by Aristotle and Plato may vary from text and translation to another, the basic structure is overall the same. In the end, we have two sources of irrational appetites, namely the desiring part from which the sensual desires and aversions spring and the spiritual part which contains feelings. What is important is that these two parts may also be found in non-rational animals, whereas the one part which only human beings may possess is the reason. As we have seen in the citations before, ‘wish’ is very common translation for Aristotle’s term for rational desire (*boulēsis*).

Even if the kind of “faculty-based” account of living things may appear somewhat odd to a contemporary reader, we should not let it confuse us. What is important for our study, first of all, is the idea according to which there are some crucial differences between living things. Creatures capable of true self-movement are those which can be considered capable of performing actions of a special type. However, the line is a difficult one to draw here while there are all kinds of protozoans and alike. In my view, this is not much of a problem, since the same overall picture can be seen in those cases, even if it is on a scale of the utmost simplicity and small size. As we are able to see in Aristotle’s texts, there is a special role reserved for human beings because of their capability for rational thinking, nevertheless they still belong to the group of self-moving animals. Anyway, secondly, Aristotle brings explicitly out what is special in this type of actions, and that is their very origin. Now is the time to take another look at the question of the difference between, say, a moving train and a dog eating a bone. Even if in both cases the subject in question is in some way producing the action or the movement by itself, only in the latter case does it spring from an appetite or desire. In the end, although we often postulate actions to inanimate beings, in the context of living beings, ‘acting’ has a special meaning. In that context, actions spring from the mental states of the subjects. This idea is the body of the famous causal theory of action.

In *Movement of Animals* (701<sup>a</sup>, 32–701<sup>b</sup>, 1) Aristotle pronounces:

I want to drink, says appetite; this is drink, says sense or imagination or thought: straightaway I drink. In this way living creatures are impelled to move and to act, and desire is the last cause of movement, and desire arises through perception or through imagination and thought. And things that desire to act make and act sometimes from appetite or impulse and sometimes from wish.

Due to this kind of statements Aristotle is often considered as the first “causalist” (for example Mele 1997, 4; Schueler 2003, 8). Alfred R. Mele (ibid. 2–3) defines causalism by giving two conditions for it. Firstly, (1) action needs to be caused in a certain way in order to be an action. The right kinds of causes are usually thought to be various psychological or mental items such as beliefs, desires and intentions. This idea is the basic thesis held by all *causal theories of action*. Secondly, (2) actions must also be explained causally, at least partly in terms of items of the kind mentioned above. Now, whether Aristotle really is a causalist is a more complicated matter which I will consider later, but at least we are allowed to say with great confidence that his theory of action is a causal one. Clearly, in various places he notes that all actions are caused by a desire or want of one of the three types he mentions, and that is sufficient for us at this point.

If we now recall what was said in the last section, we are able to see how it opposes the condition (1). I have been talking about trains and computers performing actions but, according to the condition (1), an action requires a causal history from the mental states of its agent. Therefore, since I am willing to maintain that talk about inanimate things performing actions is justified, some changes must be made to the condition (1). I definitely agree that an action springing from the agent’s mental states is somehow special, perhaps more real and more authentic, nevertheless I want to stick with the idea that it is not the only class of actions there is. With this problem, I will consult Aristotle and his remarkable *Nicomachean Ethics* where he makes his most thorough study of the nature of action. At the beginning of Book III, Aristotle distinguishes voluntary action from involuntary. He (1111<sup>a</sup>, 22–29) defines the difference in the following way:

Since that which is done under compulsion or by reason of ignorance is involuntary, the voluntary would seem to be that of which the moving principle is in the agent himself, he being aware of the particular circumstances of the action. Presumably acts done by reason of anger or

appetite are not rightly called involuntary. For in the first place, on that showing none of the other animals will act voluntarily, nor will children; and secondly, is it meant that we do not do voluntarily *any* of the acts that are due to appetite or anger, or that we do the noble acts voluntarily and the base acts involuntarily? Is not this absurd, when one and the same thing is the cause?

A few remarks should be made about the preceding quotation. To begin with, Aristotle seems to distinguish voluntary actions from involuntary ones by highlighting the cause of actions. In the last sentence, he suggests that since certain actions spring from one and the same source, their nature should also be considered similar. Now, one should ask what then the cause of voluntary actions is. While stating that non-rational animals and small children should also be considered able to perform voluntary actions, Aristotle is defending the view according to which reason is not a prerequisite for voluntary action. Furthermore, as he earlier claimed that all self-movement comes from the appetitive powers, it is evident that the cause in question must be a token of one of the three types of desires. Summing up, we may now conclude that, according to Aristotle, voluntary actions are the class of actions which spring from the agent in the way that their causes are the agent's mental states.

Anyway, I am not willing to take Aristotle's account as it is. For example, Aristotle considers actions performed of ignorance as a subclass of involuntary actions, whereas I cannot see any special role needed for consciousness in this division. For Aristotle's part, the idea makes sense as he does not have any special category for intentional action, but we will later see that it is better for the question of consciousness to come and play its part in another context. Nevertheless, in my view Aristotle's concept of voluntary action is perfect for our needs. An action which springs from inside the agent, namely from the agent's mental states, is voluntary while all the other actions are involuntary or non-voluntary. The origin of involuntary action is always outside the agent. This means that, while acting involuntarily, an agent who or which we consider to possess certain mental states acts in a way which is not in accordance with those states. For example, an agent may not control his body in some situations as is often the case with, say, people with Tourette's syndrome. In those situations, the agent acts either contrary to his mental states or simply regardless of them. On the other hand, actions performed by agents who or which are considered not to possess any mental states in the first place should be called non-voluntary. This is usually the case with trains, for



instance. The reason why they are not capable of voluntary action is simply that they do not have any mental activity. As is said, voluntary acting requires desires and beliefs.<sup>23</sup>

### 2.2.2 Explaining Actions with Desires and Beliefs

Even if the idea of some faculty of appetite as a source of all voluntary action may seem somewhat peculiar in the context of modern philosophy, the basic thought has survived incredibly well over the centuries.<sup>24</sup> The reason is simply that, after all, in all actions done knowingly a kind of striving towards them can be considered to exist. Perhaps the issue is better understood if we turn the composition the other way around. Now the question goes the following way: why would anyone or anything possibly do anything without on some level wanting to do it at least a little? The idea seems somewhat absurd, because it plainly sounds reasonable to think that there must be something there in an agent pushing it towards an action. Since voluntary action is always activity, there must be a power coming from inside the agent which may be called the initiator of the whole process. Of course, the whole question intertwines around the concept of wanting or desiring. I believe we can all agree that all conscious action must have its striving force, but what we call that force is another question. Furthermore, it should be asked whether this force has any role in the explanations of voluntary action. Next I will make a brief survey on the matter.

After the World Wars, there emerged a tradition which aimed to belie the causal theory of action by proving that the relation between actions and the factors explaining them is a logical one instead of causal. One of the figures in this movement was A. I. Melden. Now, what is interesting in Melden's view is that he also emphasizes the role of wanting when talking about actions. Actually, to go even further, he (1961, 123) claims that wanting and doing are logically connected in a way that makes it impossible even in theory to understand one without understanding the other. In his view, it would be exactly as absurd to do something without wanting to do it as to want something without wanting to do anything with

---

<sup>23</sup> As a matter of fact, I will argue during the next sections that, after all, actual mental activity is not a strict requirement for considering agents to possess desires and beliefs. However, in most cases they go hand in hand, and thus this is enough said at this part.

<sup>24</sup> As an example of opposed thinking: Steward 2009, 303.

the object of wanting. However, as a surprising twist, Melden turns this idea against the causal view of action. He (ibid. 128) writes:

If the relation were causal, the wanting to do would be, indeed it must be, describable independently of any reference to the doing. But it is logically essential to the wanting that it is the wanting to do something of the required sort with the thing one has. Hence the relation between the wanting to do and the doing cannot be a causal one.

This idea, according to which a causal relation between two things requires that no logical relation exists between them, Melden (ibid. 53) links to David Hume. Whether this is fair is a question which I will not consider here, but what I think is crucial instead is to examine whether it makes sense to take the statement as given and thus found an argument on it. In my view, it does not.

The whole issue about the incompatibility of causal and logical connections is surely a disturbing one and, for the most part, I am not even quite sure about what the problem is in the first place. In the preceding section, I defined voluntary action to be of a kind which springs from the agent's desires. Therefore, with this definition it is quite obvious that the concept of voluntary action is not logically independent of the idea of wanting while the definition requires that it is always the wanting that causes voluntary actions. The whole problem seems to arise from some kind of philosophic gibberish. In the preceding quotation, Melden defines logical relation to be such that one thing cannot be understood or described without reference to another. Now, in my view, it seems rather evident that causal relations are often in an essential role when trying to understand or define a concept. For example, one cannot understand what it is to be drunk without understanding the cause of it, namely drinking. Of course, one may understand the physiological or mental state of drunkenness without knowing anything about drinking, but similarly one may understand the physiological phenomenon which goes on when somebody is moving his hand without considering it as a voluntary action. Hereby, all I can do is wonder, what kind of obstacle we are expected to find in these kinds of logical relationships to causal ones.<sup>25</sup> As has been said, the causal factors are often among the most crucial ones in the process of understanding various things.

---

<sup>25</sup> Donald Davidson (1963, 695–696) also responds to this critique by admitting that some logical connections among causal action explanations may exist, yet denying that it is a problem. All in all, Davidson also seems to be a little confused about the whole issue.

Donald Davidson's "Actions, Reasons, and Causes" is definitely the most significant singular paper written in the field of the philosophy of action. In this paper originally published in 1963, he is aiming to defend the causal theories of action against the movement mentioned before. However, instead of merely responding to the critique, he also gives a precise definition of what any causal theory of action should stand for. In this project, Davidson does such good work that it has been widely seen as a real knockout for alternative theories, and thus his causal theory of action may be considered as the standard view of action. Next I will briefly introduce the ideas presented by Davidson.

As it is said, before "Actions, Reasons, and Causes" there were a lot of texts written about how reasons for actions explain actions without actually causing them.<sup>26</sup> This tradition owns a great deal to Ludwig Wittgenstein who, as it is known, aroused general interest in logical and grammatical relations between concepts. In the context of action explanations, it naturally means that, instead of talking about reasons as causes, philosophers thought that reasons have more of a logical connection to actions they are meant to explain. The overall idea is that, since actions always have various meanings according to the way we interpret them, the reasons explaining them are somehow logically included in those interpretations. Of course, the tradition did not vanish for good after the publication of "Actions, Reasons, and Causes", nonetheless its role decreased substantially. Georg Henrik von Wright's *Explanation and Understanding* is one of the best examples of the whole tradition, even though it was originally published as late as in 1971. Anyway, clearly this "Wittgensteinian" approach has its point. For instance, in Melden's (1961, 86) classic example, a man is driving a car towards an intersection, and hence raising his arm in order to give a turning sign.<sup>27</sup> Now, if we are to explain why the man in question raises his arm, we must clarify that he is actually giving a sign at that moment. As it is clear, this operation does not have anything to do with causality. What we are doing instead of mentioning the causes is to re-describe the action so that the new description includes more information than the original one. Suddenly, we are able to see the social context, and therefore the new description carries much more meaning than the mere physical movement of the body.

---

<sup>26</sup> Davidson (1963, 1, n.1) himself mentions, among other things, Anscombe's *Intention*, Stuart Hampshire's *Thought and Action* and Anthony Kenny's *Action, Emotion and Will*.

<sup>27</sup> Naturally, times have changed since Melden wrote his book, but I believe that we are able to appreciate the idea of his example. If the burden still feels too heavy for a modern day motorist, he should put 'a bicycle' in the place of 'a car'.

Undeniably, it is of the utmost importance to understand the huge power which the freedom of interpretation has while we contemplate the world. Particularly, within the social context, simple actions may serve various ways. Staying faithful to my view of events, I would say that there is a large number of very complicated actions to be found in all phenomena within the social scene, and therefore it is crucial to understand our role in interpreting. However, the next question should be whether that is enough and, moreover, does the search for new meanings have anything to do with explaining in the strict sense. In his “Actions, Reasons, and Causes”, Davidson makes this important question and answers it by the same token. While considering Melden’s case of a signaling motorist, he (1963, 692) notes that there are two ways to understand the situation and with both we end up in trouble. Firstly, we may say that when we understand that there is an action of signaling, we also learn the reasons for the action at the same time, because the reasons are always included in the actions logically. Now, even though we really have the explanation, we may wonder what the actual relation between the reasons and the action is. Moreover, we may ask how one is supposed to know the reason only by seeing the action, since people may do similar things for very different reasons. The other possibility is that the theory requires knowledge of the reasons the agent possesses as well as of the action, after which we would be allowed to conclude that the reasons in question explain the action. However, even with this improvement we end up with the same problem. In the end, one may possess different reasons for performing one and the same action, and hence only by listing the reasons we do not get any particular explanation for the action. All in all, the knowledge of what a person is doing or what kind of reasons he has does not imply the knowledge of why he is doing what he is doing.

As I tried to argue in the first chapter of my work, the idea of true explaining requires genuine causality. In science as well as in other contexts, by talking about ‘explaining’ we should refer to the project of mapping the causal scene, albeit that we undeniably use the term in various ways depending on a situation. Understanding instead is another matter. As I have claimed earlier, explanations should be considered as answers to why-questions, whereas understanding is more a matter of finding answers to what-questions. Although understanding bears more importance than explaining, just as asking ‘why’ requires understanding of ‘what’ and an answer to a why-question may be part of an answer to a what-question, this very study is about explaining, and that is why I am willing to keep up with the causal picture, even in such a complicated context as the social scene. Hereby, I conclude that understanding what

somebody is doing is not the same as knowing why he is doing it and, in order to know the answer to a why-question, we need to know what the cause behind the action is.

According to Davidson (1963, 685), a reason for an action must be a causal force by its nature, and it must be constituted of two separate factors. He states that “[w]henver someone does something for a reason, therefore, he can be characterized as (*a*) having some sort of pro attitude toward actions of a certain kind, and (*b*) believing (or knowing, perceiving, noticing, remembering) that his action is of that kind”. Under the concept of pro attitude, he (ibid. 685–686) includes “desires, wantings, urges, promptings, and a great variety of moral views, aesthetic principles, economic prejudices, social conventions, and public and private goals and values in so far as these can be interpreted as attitudes of an agent directed toward actions of a certain kind”. Later he (ibid. 686) adds that “[g]iving the reason why an agent did something is often a matter of naming the pro attitude (*a*) or the related belief (*b*) or both [...]”. Now, we should pay attention to the adverb ‘often’. If we think about the sign-making driver in the above example, it is easy to demonstrate the matter. Let us imagine a situation where we are trying to explain his conduct and we are able to conceive a large number of different but nevertheless sufficient answers. For instance, we may say that he wants to inform other drivers about his approaching turn. This would be the case of mentioning the pro attitude (*a*). On other hand, we may as well say that he knows that sign-giving is the best way to make others aware of his intention which would be mentioning the belief (*b*). However, besides these options there seem to be countless ways of giving reasons for actions. We may, by the same token, explain that there is an intersection at the end of the road or that his father taught him that one should always obey traffic regulations. The multitude of reasons does not come across as having even the sky for a limit. Anyway, later in the next section I will consider these types of explanations and their relation to mental states, yet one should ask whether all action explanations can truly be reduced to mental states. For example, if we imagine ourselves in the backseat of his car asking the driver himself for his reasons, is it really likely that he would start to talk about his causally effective mental states? To my mind it is not, but for now I will leave this question aside, and shall not return to it before Section 2.3.

The problem about the nature of reasons for action has generated a great deal of literature. One reason for this is that the concept of reason for action seems to be one which can be used and understood in various ways and, therefore, the disputes between accounts spring mostly from different kinds of attempts to compel the whole concept under one

singular notion. In my view, this very problem is also what Davidson facing. However, regardless of the fact that Davidson's idea is guilty of the crime of over-simplification, his view contains a remarkable share of wisdom. Instead of giving a sufficiently comprehensive account of reasons for action, he rather well introduces one notion of the concept, perhaps even the most substantive one. Evidently, this very notion is the most important one at least within the world of scientific explaining. Anyway, I have defined a voluntary action to be of the kind which springs from the agent's appetitive forces, and thus it is natural to imply that, since Davidson considers these appetitive forces to be reasons for actions, his account is no less than perfect for explaining voluntary action. In other words, while mental states are the essence of voluntary action, the explanations of these kinds of actions are naturally those mental states themselves. Since mental states cause voluntary actions and explaining is about naming the causes, the whole picture is etched beautifully. Now, because this Section 2.2 is all about voluntary action, I will devote the rest of it to the notion of reasons for action as an agent's mental states. In Section 2.3, I will concentrate mainly on intentional action as I also want to study another aspect of the concept of reason for action.

### 2.2.3 The Obscure Nature of Mental States

Earlier Donald Davidson argued that, when we are giving a complete explanation for a voluntary action, we need to cite the agent's (*a*) pro attitude and (*b*) belief. Even if that sounds reasonable, the nature of those two elements is far from clear. In the end, although mental states play a rather crucial role in our understanding of the world, their basic nature is somewhat vague. In this section, my aim is to study the question of what we are talking about when we are referring to mental states. I believe that, by answering this question, we may also reach the position from which it is easier to understand why the whole idea of social law is more or less unconvincing. First I will examine the ideas of pro attitude and belief separately, after which I will conclude my view of mental states all together.

If we are willing to claim that all voluntary actions spring from desires, it is obvious that we need an extremely extensive concept of desire for the purpose. If we compare, say, a mouse eating cheese and Knut Hamsun writing *Growth of the Soil*, clearly there are not many

similarities to find. However, Davidson evidently took this challenge seriously when he presented his concept of pro attitude (see page 69). Under that very term he listed various desires from primitive urges to nobler ones. In any case, repeating the list would not serve any purpose here. As the term suggests, it is meant to cover all attitudes aiming towards something. Now, considering this wide-ranging class of attitudes, we should realize that it is not something we are able to get a decent grip on. Little by little it is getting more evident that, instead of being an actual group of states bringing out all kinds of actions, we have a group of causally effective states which are developed only for explaining actions. Even though this may sound peculiar, I will demonstrate the idea by starting from some simple cases.

G. E. M. Anscombe (2000 [1957, §36]) states that “[t]he primitive sign of wanting is *trying to get*: in saying this, we describe the movement of an animal in terms that reach beyond what the animal is now doing”. Now, this thought is of the utmost importance. If we consider, say, non-rational animals, it is incontestable that all we can know about their mental states is through studying their behavior. We cannot ask them what they want (perhaps a dog, for example, may learn some words like ‘bone’, and thereby we are able to ask if it wants a bone, but the answer will likely be nothing more than an action manifesting the want), and that is why we must make our conclusions about their mental states on the basis of how they act. Of course, we may take the issue further from the present moment by learning regularities in their behavior or only by making daring assumptions. We may know our dog so well that, even without seeing it, we may claim that it wants to go out at six o’clock or that the poor creature wants to gnaw bones all the time. In any case, all these desires we have learned through observing our dog’s behavior, and therefore all the facts supporting them are basically leaning on its actions. Now, since the only way of analyzing the pro attitudes of non-rational animals is by studying their behavior, it becomes clear that their mental states have meaning only in the light of their actual or potential actions. Perhaps it is even more accurate to say that their mental states exist only in the light of the behavior they cause. This idea seems to contradict the idea according to which mental states cause behavior, but in my view that is not a worthy problem. On the contrary, it is perfectly normal that causes have their meaning or existence only through the effects they bring about. For example, who would be interested in viruses if they caused no diseases? Anyway, I believe that, after all this speculation, we have a good basis for understanding what D. M. Armstrong (1968, 82) means when he is defining a mental state being “*a state of the person apt for bringing about a*

*certain sort of behavior*". Even if the term itself implies that the basic substance of mental states is their mental nature, in the end the question is mostly about pure dispositions. By studying the behavior of a non-rational animal, we are able to learn the dispositions guiding it, and these dispositions we call 'mental states'. The term 'mental' in 'mental states' is there to distinguish these dispositions guiding voluntary actions from those by which we can understand, say, the movements of a train. Mental states are always inner dispositions, and thus something else than physical facts.

When turning from non-rational to rational animals, the issue gets naturally much more complicated. The world of human beings is so rich that the actions and, by the same token, the reasons behind those actions become so multifaceted that, in order to understand human life, we need to internalize the idea according to which what shows outside is only a little part of what goes on inside. People are capable of expressing themselves not only by acting but by various means of communication, and therefore the dispositions guiding their actions are there to be learned from many different sources. Furthermore, the dispositions of human beings are much more complex than those of creatures interested in not much more than eating and sleeping. Values are one good example of the underlying dispositions which may be considered to exist all the time even if they would never be expressed in any manner. However, in my view, the basic structure is nevertheless the same among rational and non-rational beings. Mental states by way of which the behavior of human beings is explained are in any case nothing but dispositions, which may be postulated not only due to behavior but to many other factors too. As is said, the main difference lies on the level of complexity. For instance, if we know that someone belongs to a certain political party or that someone thinks highly of French literature, we may postulate certain dispositions on them according to which we may explain why they join in marches or like to wear berets. Obviously, Davidson and Armstrong in their previous quotations (see pages 69 and 72) are talking only about human action instead of extending their views to cover the whole animal kingdom. I guess it makes sense since the discussion of reasons for actions within the social world is much more interesting than the study of the behavior of non-rational beings, but I believe that if the factors explaining behavior are the same in both cases, we should not make any definite division between the two. If we are to argue that human actions should be explained with causally effective mental states, why should we make any restrictions for excluding non-rational animals from the domain? Indisputably, we explain all animal action with the very same pattern we use in explaining voluntary actions within the social realm. It is perfectly



sensible to claim, for example, that the dog jumped on the table because it wanted the bone and believed that the bone had been put on the table. As is said, in my view, voluntary actions as well as their explanations obey the same structure among all animals, albeit that the social world is naturally much richer and more complex.

Above, Davidson wrote that a reason-giving pro attitude must be directed towards some type of action and that the action the reason is about to explain must be, on behalf of the agent, considered to be of the type in question. Similarly, Melden thought that we cannot want anything without wanting to do something with the object of desire, and hence wanting is always wanting of some kind of doing. I am not quite certain about this. In my view, it is rather normal that we merely want to own something without any further activity. Nonetheless, in these kinds of cases one may argue that owning or possessing is in itself a type of doing. For example, owning a stamp collection would be just one ultimately passive way of acting. However, even if we accept this highly questionable category of actions, there are still instances where the wanting does not seem to meet the criteria. For example, I believe that it is absolutely appropriate to claim that somebody wants peace without any further activity even if there is nothing one could do with it, not even to possess it. Therefore, I would conclude that the reason-giving pro attitudes must not *necessarily* be aimed at actions as long as the belief which serves as a link between the activity and the target exists. For instance, marching on streets would be explained completely by citing the agent's urge for peace, as long as the person in question also believes that his marching is a means to achieve it. Of course, one may point out that the object of the desire could as well be, say, the act of fighting for peace, yet it would seem a bit like an *ad hoc* –action. In my view, no sufficient reasons for limiting reason-giving pro attitudes only to those aimed towards actions exist.

If we are now to consider Davidson's other condition (*b*), at first sight it comes across as too challenging for non-rational animals.<sup>28</sup> There is a certain sense in believing that we are not too happy to associate with them. Can we really say that the dog in the preceding example knows that the bone is on the table or, to go even further, that its action is the action of reaching for the bone? In my view, the first alternative is valid, but clearly the latter one is a much too complicated belief. While non-rational animals do not have the ability to use language, it is plausible to assume that they cannot possess that kind of information. Later, under the headline of 'intentional action', I will consider a further dimension which is

---

<sup>28</sup> In "Thought and Talk" (162–163) Davidson writes about attributing desires, beliefs and other attitudes to dumb creatures and, as his standards for such are truly high, he is not very enthusiastic about the idea.

brought to the context of action by the ability to use language, but before that I will content myself with more modest ways of knowing. However, I understand perfectly well if Davidson requires higher-level knowledge when talking about reasons for human action, but since I make my division elsewhere, the lower-level knowledge is sufficient for this context. Even if we are not allowed to say that the dog really knows that its action is ‘the action of reaching for the bone’, I believe that it is perfectly appropriate to say that it has the knowledge that its action is a means for satisfying its desire for the bone. Therefore, I would say that the dog wants to eat the bone and has a belief that, by jumping on the table, it will get what it wants, and there we have a perfect explanation for why it is messing the place by jumping around.

One more issue in Davidson’s view to contemplate here is the question about the necessity of the condition (*b*). It is clear that in most cases we need to link the pro attitude to one’s action by showing one’s belief according to which the action is a way to satisfy the desire. However, sometimes the pro attitude seems already to be aiming towards the action itself. For example, when I am making noises only for the fun of it, no further goal to which the belief is needed to form a bridge exists. Of course, the value of this kind of “explanations” is virtually non-existent. That is why Davidson (1963, 688) himself argues that these kinds of cases are better to take as ones where there is no reason at all. This seems merely to be a matter of taste. Even if people often say that some voluntary action was taken “for no reason”, it nevertheless does not mean that there is no explaining desire behind the action. Plainly, it rather means that the action shares no further purpose. However, while the role of appetitive force is so crucial in distinguishing voluntary actions from others, I would still stress the importance of the explanatory desire in actions done for no further purpose. Anyway, there are also other beliefs considered as necessary for voluntary actions. One good example is the idea according to which one cannot perform an action without believing that it is possible. This thought is found as far away as from Aristotle (*MA* 701<sup>a</sup>, 24), yet today it arouses opinions both for and against.<sup>29</sup> Another good example is Anscombe’s (2000 [1957], §36) idea according to which a belief in the existence of the object of the desire or, in some cases, merely the idea of the object is required of the agent. In any case, whether those beliefs are necessary for acting has very little influence on the explanatory side of actions.

---

<sup>29</sup> For example, Irving Thalberg (1972) tries to deny the claim, whereas Rogers Albritton (2003 [1985], 415–416) defends it.

Even if the standard case of explaining a voluntary action would be to mention both a pro attitude and a connecting belief, the reality is often something quite different. Usually in real life situations only to refer to one or other of the two serves perfectly. However, there are also surprisingly many cases in which neither one is mentioned, even though the question is about explaining with mental states. Often voluntary actions are given reasons by mentioning, say, emotional states, purposes or conditions of the past, the present or the future. For example, one may explain a murder by saying that the killer was jealous. Now, by saying this, he does not mention any pro attitude, but nevertheless he is giving us the ground from which the necessary pro attitude may be deduced. In this case, the jealousy can be seen as the motive of the murder. Even if this mental state is not itself a desire, it is the source for certain kinds of desires which are not specified, yet taken for granted. Alternatively, by saying that the killer was aiming to hurt his girlfriend, we are simply using other kind of language. Instead of mentioning a desire, we are mentioning the target of one, and hence the desire is included in the story. In any case, perhaps the largest group of action explanations is the answers which refer to some conditions, events or state of affairs. For example, by saying that the victim had an affair with the killer's girlfriend or that the victim was a sinful man, one may give a decent explanation as well. As was the case with the explanations referring to motives, these explanations also give grounds for certain pro attitudes. Hence, when explaining voluntary action, we may refer to certain conditions as long as they can help us to understand the mental attitudes of the agent. Even if these kinds of explanations do not mention any particular pro attitudes or beliefs behind actions, they may give even more information by pinning down the factors which aroused the mental states needed. As is said, pro attitudes and connecting beliefs are often rather uninformative, although they play the most substantive role in the explanations. In many cases, we learn more if we are willing to go a little further, to the source of those explanatory elements.

Whereas purely physical events are usually explained by pointing out physical events and conditions in the causal map leading to the *explanandum*, in voluntary actions these causal maps must include mental states. This also applies as much to ants building their anthills as to human beings running around in their shopping malls. Anyway, in this section I have been demonstrating how the nature of those entities is something very peculiar. Even though the term itself suggests there being a kind of mental occurrence, in most instances there is only a rather passive disposition at stake. If we consider a drug addict as an example, evidently his desire is very actual and ever-present. However, the situation with, say, a

political activist may be essentially different. People may call somebody a radical left-winger craving for communism, although the man himself spends most of his days on the couch reading *The Phantom* comic books. In the end, the question here is not about what goes on in a person's head, but how he would act in certain situations. As Gilbert Ryle in his *The Concept of Mind* (43) writes, “[t]o possess a dispositional property is not to be in a particular state, or to undergo a particular change; it is to be bound or liable to be in a particular state, or to undergo a particular state, when a particular condition is realized”. Similarly, Armstrong (1968, 86) notes that “one essential thing about dispositions is that we can attribute them to objects even at times when the circumstances in which the object manifests its dispositions do not obtain”.

If we are to consider action-explaining mental states as nothing more than dispositions, we must encounter one important and highly interesting consequence, namely the sudden decrease of the meaning of the very mental nature of those states. In many instances, the idea of mental substance ends up playing a rather small role, while the question is mostly about dispositions helping us to explain and predict behavior. Moreover, even in those cases in which the explaining disposition is one thoroughly mental event, absolute certainty about it is possible only for the agent itself. This is the core of the classic problem of other minds. Although we have to presume that all more or less similar creatures share more or less similar inner worlds with us, we are never able to have any true certainty about the matter. Therefore, all the talk about mental states is based on assumptions which are made according to various external hints about our fellow creatures. Relying on those hints, we postulate mental states to rational and non-rational animals around us without ever reaching the full certainty. Now, taking all this for granted, the idea of postulating those dispositions to agents which most likely are not capable of any mental activity does not come across as such a major error any more. In the pragmatic sense, it is extremely useful, if not even unavoidable, to sometimes see purely physical entities as being capable of acting voluntarily. Lastly, under the theme of voluntary action, I will introduce those “unorthodox agents”.

#### 2.2.4 Unorthodox Agents

I guess that the whole issue of unorthodox agents is best illuminated with one Daniel C. Dennett's famous example. In his *Intentional Stance*, Dennett uses an example of a chess computer in various places to demonstrate his view. Plainly, his idea is that in certain circumstances we may be able to choose from different stances while observing one and the same object. The chess computer is a perfect example for Dennett's purpose as, in all its complexity, it shares various features through which it can be considered. First of all, it is a physical machine which obeys the laws of physics. Second, it is a designed artifact which is created to serve some purpose, and hence to produce certain kinds of outcomes. However, in Dennett's account there is one stance which should interest us the most. What he calls 'intentional stance' is the one which is taken towards a chess computer, while we are playing with it. In the end, the only sensible way to play with the machine is to play with it like one would do with human opponents, namely by analyzing it through its inner dispositions. Of course, one could have a meaningless game with it from a designer's stance as well. As we can imagine, the designer and the machine must have had a huge number of games together during the process of developing the device into its best. In those situations, the designer must preserve his designer's stance towards the computer in order to see which features are in need of improvement. However, if we are to consider actual playing as what the device was created for in the first place, we may conclude that a viewpoint of another kind must be taken.

Even if Dennett is talking about 'intentional stance', in agreement with my use of concepts I would call it 'voluntary stance'. The idea remains the same. If we are to play chess with a computer, we need to consider it as a reasonable opponent willing to beat us. Therefore, we think in a way that requires it to possess mental states, although at the same time we most likely are absolutely definitive that it cannot possibly have any (perhaps those who have seen the movie *2001: A Space Odyssey* could not be quite so certain about the matter). In order to have a meaningful battle, we need to think that the poor scoundrel is after our honored king. These kinds of games are evidently based on various dispositions we postulate to our opponents and, whether we will win or lose, may depend totally on them. For instance, we may believe that the computer is willing to attack from the left hand side or that

it believes that through sacrificing its knight it may get our rook afterwards and, according to these postulated mental states, we predict its moves and plan our own ones. Even if only little children and idiots are allowed to consider trains and cars as having characters and feelings, it is perfectly acceptable for a grown man to curse his inanimate opponent in chess for being devilishly crooked. Nonetheless, as is said, the image of a chess computer as a voluntarily acting agent does not require on any level a denial of the image of it as a man-made machine which is incapable of any mental activity. On the contrary, even though a viewing from two different viewpoints precisely at the same time would be virtually impossible, a sliding from one to another may happen extremely smoothly and even totally unobserved.

I think that it is impossible to deny that we are able to take this ‘intentional stance’ or ‘voluntary stance’ towards artificial objects which do not even have the slightest possibility of possessing any mental states. In the end, the question is only about dispositions through which their behavior is best explained and predicted. After all, with all the instances whose behavior is interpretable through devices of that kind, it should be interpreted through them simply because it is practical. There is a quite obvious sense in the pragmatist idea according to which the best working explanation or prediction should be chosen, even if it might miss certain aspects of truthfulness. Furthermore, the same applies the other way around. When a chess computer crashes, there is no use trying to persuade it to co-operate by, say, buying a new hat for it. In those situations, we need to take the designer’s stance and, at least in most cases, also to call a repairman. Anyway, the possibility of voluntarily acting agents which are not capable of possessing any real mental states has even more interesting consequences than the one we have contemplated, at least in the light of social action. The rest of this section is devoted to the idea of group agents.

Traditionally, the whole area of the philosophy of action has concentrated on actions performed by individuals. This is quite natural while most actions are of the type. However, during the last few decades, the awareness of the fact that the analysis of the behavior of individuals is not sufficient for all purposes has grown. For example, within the realm of ethics it is obvious that most of the larger-scale problems are effects of collective acting. Pollution is a good example of a problem for which almost the whole of mankind is to blame. Evidently, the idea of collective responsibility requires a theory of collective action. In any case, we do not need to go that far in order to demonstrate the matter. For instance, in the context of sports, talk about teams doing things is more than usual. Indisputably, there is nothing wrong in saying that a soccer team plays badly or that a baseball team is willing to

show what it is made of. Since it is undeniable that collectives are treated as acting individuals in many instances, the idea of group agency has long been begging to be studied.

Most likely Raimo Tuomela's *A Theory of Social Action* is the first systematic study of collective action. It was published as late as in 1984. Among other things, in his work Tuomela defends the idea of *explanatory individualism* in the explanations of social action. According to this doctrine "it is conceptually and epistemically possible that all social phenomena and patterns be explained by individualistic explanantia", (Tuomela 1984, 455). This kind of thinking goes under the *singularist* view of collective action.<sup>30</sup> Singularists believe that collective action is always reducible to the actions of individuals. In other words, collective action is nothing but actions of individuals put together in a certain way. Therefore, the only true agents there are must be individual agents. As we can see from the quotation above, this singularist account has some further implications as well. For instance, if we believe that actions should be explained causally with mental pro attitudes and beliefs, in the case of collective action the explanatory equipment must also be reduced to the mental states of individual agents.

At first sight, the singularist account makes perfect sense. What else could it be if a group is acting than the sum of its members performing actions? In the end, there is nothing else happening than a bunch of people performing actions on their part. However, this line of thought can be refuted by various counterexamples which, at least in my view, come across as insurmountable. For instance, for the singularist account it is really hard to distinguish collective actions from seemingly similar actions done without the idea of collectivity. This is best demonstrated by one much discussed example from Margaret Gilbert (1990) where two people are walking in parallel. Now, at least two possible scenarios can be distinguished in this case. In the first, the two have gone for a walk together, but in another they have ended up walking side-by-side solely by accident. Evidently, in the first scenario the question is about collective action, but in the latter there are merely two individual actions happening at the same place and time. It might not be impossible, yet to my mind it seems highly difficult for a singularist to make a distinction between the two scenarios, since the individual actions in both are the same. After all, there is nothing but two people walking in parallel in both cases, but only in the first one does it come down to collective action. Anyway, it is easy to make up counterexamples against the idea according to which collective actions are nothing

---

<sup>30</sup> The term 'singularism' in this sense is supposedly from Margaret Gilbert (1989, 12) originally.

but individual actions put together. For instance, in the case of team sports, it seems that the actions of groups are something more than the actions of the individuals belonging to the group. We may say that soccer team A is attacking fiercely from the left hand side or that it scores, although in both cases there might be only one of the 11 players actually doing something. Similarly, we may say that team A purchased a new player, and here we have a problem even in defining what the team as a whole is. Moreover, one may even go as far as claiming that team A is seeking payback from team B in the cup final as they lost the previous one 60 years ago. This way of talking is fully appropriate, even if none of the players, staff or owners lived at the time of the previous final. In addition, it might actually be that none of them even knows that the team has played in a cup final before.

By emphasizing the existence of collective actions which are something primitive in a sense that they cannot be analyzed only as a sum of singular actions, many philosophers lead themselves to a slightly uncomfortable situation. In groups we now have a new kind of agent which seems to be capable of performing its own actions, even if it does not seem to obey the same principles as the more common agents. By this I mean mainly the fact that the idea of a group mind or group consciousness sounds absolutely ridiculous. As John Searle (1990, 407) puts it, “[a]ll consciousness is in individual minds, in individual brains”. Now, Searle is a good example of a philosopher who ends up with the foregoing problem as he (ibid.) also argues against singularism. By stating that (1) it is possible for genuine group agents to exist and that (2) the idea of group consciousness is totally impossible, but at the same time stressing that (3) actions must be explained causally with genuine mental states, he has the remarkable task of explaining how the group actions are explained while no group desires and group beliefs exist. Searle, however, offers us an explanation. For this special purpose he (ibid. 412–413) introduces a special kind of intention, namely ‘we-intention’, which he uses to solve the problem. These we-intentions are intentions of individuals and, according to Searle, they are the unique piece which is capable of turning a bunch of singular actions into one genuine collective action.

There has been a lot of discussion about we-intentions while their nature is quite far from being straightforward. However, it is useless to open up the problematic nature of the concept here. In the previous section, I presented my view of action explaining mental states, and clearly in my account there is no truly mental *explanandum* needed. If we are allowed to consider chess computers and the robots from *Star Wars* –films as voluntarily acting agents, there will not be any trouble with the actions of groups and collectives. As far as is practical



for explaining and predicting to talk about, say, soccer teams having intentions, wants and beliefs, I myself do not see any restrictions on it. After all, postulating dispositions to groups is often the only reasonable way to comprehend the world around us. In these cases, as in many others, we simply need to ignore the fact that there cannot be any true consciousness behind the agent, even if it acts like it would have.

To my mind, the problem of we-intentions is elsewhere than in finding the right definition for the concept. Even though it is evident that collective action in most cases demands intentions of a special kind from the individuals, one should not go as far as placing necessary requirements on them. For example, one non-singularist who has written about the topic is Gilbert. In her *Living Together: Rationality, Sociality, and Obligation*, she gives many examples of how pro attitudes and beliefs we may postulate to a group might be even totally independent of what pro attitudes and beliefs we consider the members of that group to have. It is quite undemanding to make up one. We may imagine, say, a committee whose members, through voting, commit themselves to hold a belief as a group, even if many of them personally think that it is untrue. However, Gilbert is even ready to go as far as to claim that it is totally possible for none of the group members to possess the pro attitudes and beliefs of the group. Her (1996, 210) inventive example is Hans Christian Andersen's story *The Emperor's New Clothes* where the crowd can be said to believe that the emperor's new clothes are incredibly beautiful, even though all the individuals see that His Majesty is completely naked. I believe that Gilbert is right and that the same goes for intentions as well. For instance, it is common for compromises made in a group to bring about a situation where the shared intention is not equivalent with any singular intentions. In any case, even Gilbert places an inner requirement for group agents or, as she calls them, plural agents. In her *On Social Facts* (164), she defines plural subject the following way:

A participant in a shared action acts in his capacity as the member of a plural subject of the goal of the action. He will count as the member of such a plural subject when, at a minimum, and roughly, he and others have expressed to each other their willingness jointly to accept the goal in question now. They will then count as jointly accepting it, and hence as constituting the plural subject of that goal.

In my view, Gilbert's thought is perfectly reasonable. In order to call somebody a member of a group, it is clear that some kind of verbal or non-verbal mutual agreement must

be reached between the member and the group. However, within the context of everyday actions this requirement seems nevertheless too strict. While postulating collective actions to the world, it is sometimes impossible to know whether there has been any “joint acceptance” reached. In most cases we may know, but in some it is not even very relevant. We may consider, say, an orchestra playing Johannes Brahms’ *Symphony no. 3*. There is no doubt about whether the members of the group acknowledge the participation of the others as well as themselves in the group. It is highly unlikely that, for example, one of the violinists has ended up there by some absurd accident. However, there are also always kinds of borderline cases within which the requirements are exposed to doubt. First of all, one may consider a group of ants or even robots performing a task in collaboration. We may ask what can be this “joint acceptance” which must also be required in these kinds of cases. Similar problems arise among rational human beings as well. If we see, say, a video clip of a riot or war, we must be able to postulate group actions there even if we know for sure that it is simply impossible that “joint acceptance” on any level could tie together all the participants. For instance, in a riot there are typically all kinds of people hanging about on the streets, and likely many of them do not play any part in what is going on. Yet we are justified to count them in as comprehension of the whole would in other cases be impossible. Of course, a requirement of some kind of commitment would be fair, but how on earth could that be possible under circumstances in which a group is more or less randomly gathered? As well as needing to be able to talk about chess computers making voluntary moves, we must also be able to talk about mixed bunches of people behaving according to a collective disposition.<sup>31</sup>

Here I have been examining the concept of voluntary action. I have argued that, with the very adjunct, it is possible to distinguish a special group of action which is also the most important among all actions. For example, in the light of social sciences, the picture of human action needed is mostly from the viewpoint of actions performed because of mental causes. However, even if voluntary actions form the most important group of actions in the objective sense, it is not the only aspect of actions we may highlight. In the next section, I will study

---

<sup>31</sup> Gilbert (1989, 165) also wonders about a similar counterexample. She considers a case where a couple is waltzing together but where another dancer’s mind is completely blank. In these kinds of cases Gilbert is ready to question whether the action is actually happening at all. In her view, we should deny the action in question or at least exclude the participants who do not go through the required mental processes. However, as I have argued before, we cannot really know about the mental events of anybody else besides ourselves. Therefore, the role of those events seems to be a bit exaggerated in Gilbert’s account, even if the requirement makes sense on many levels. In the example I presented above, the situation does not change even if we know about the varying roles of the “rioters”, and thus it seems to be evident that the external factors may matter more than the internal ones under certain circumstances.

another one which is less acknowledged but most likely the dearest one, nevertheless, to all autonomously acting, rational individuals.

## 2.3 Intentional Actions

### 2.3.1 Aristotle and the Teleological Side of Actions

In the preceding sections, I argued that, by adding the property of voluntary, involuntary or non-voluntary to action, we will view it in the light of its mental causes. Voluntary action springs from the agent's mental states whereas the origin of involuntary action is elsewhere in the cases in which the agent is nonetheless considered capable of possessing them. Non-voluntary actions are instead taken by agents which we regard as obeying physical laws only. With this distinction, however, we are not capable of reaching all the crucial aspects in acting that we may be willing to. For instance, there seems to be a very definitive difference between the moves of a chess computer and a more human opponent. Even between non-rational animals and rational human beings, there seems to be a wide gap at least in the case of acting. Of course, for the most part we act like all other animals merely following our instincts and satisfying our primal needs. Simply said, often we are solely doing things without any special kind of thinking. In any case, the difference seems to be that, in spite of all, we are always at least capable of performing incredibly complicated actions as well. Moreover, even in the case of simple actions, we are able to understand our actions more profoundly than any other creatures. The capability for rational thinking and use of language seem to be the keys which raise human actions on their own level. At least for me it appears evident that further distinctions must be made, but first I will return to Aristotle's view for a while.

Earlier, in Section 2.2.1, I speculated briefly whether Aristotle is truly a causalist. As I conceded, his view of action clearly obeys the condition (1) in Alfred R. Mele's definition of causalism (see page 63). Aristotle explicitly stresses in various places that all actions have

their cause in the appetitive part of an animal. However, what should not be forgotten here is that, perhaps quite peculiarly to a modern reader, Aristotle (*Ph*, II, 3) believes that there are actually three types of possible causes besides the effective cause which has been under scrutiny thus far. Of course, the word ‘*aition*’ which Aristotle uses and which is traditionally translated as ‘cause’ has different connotations than our concept of ‘cause’, and therefore the standard translation has evoked a lot of confusion in history. J. L. Acrill (1981, 36) notes that perhaps the term ‘explanatory factor’ would have been less confusing, but it also shares some oddity. Either way, Aristotle’s idea seems to be that, if we are to understand something completely, we must have knowledge of each of the four causes. The four causes are the four essential ingredients of a thing, and hence they all need to be grasped in order to accomplish a comprehensive picture of it.<sup>32</sup>

As is said, ‘effective cause’ resembles our modern notion of cause. The other three of Aristotle’s (*Ph*, II, 3) causes are ‘material’, ‘formal’ and ‘final’ ones. Material cause is the material of which the thing consists. For example, in the case of a bronze statue the material cause would be the bronze. Formal cause is naturally the form of a thing, and in our example it would be the form which we consider as that of a statue. However, the one in which we are interested here and which has evoked the most discussion is the last one. By ‘final cause’ Aristotle means the end or purpose of a thing. In our example that would be the purpose for which the statue is sculpted. For instance, the statue could serve the purpose of depicting Jari Litmanen.

If we are to consider events and actions, it is quite obvious that the most relevant explanatory factors in Aristotle’s sense are the effective cause and final cause. The idea of the goals or purposes of things is rather unique in many instances, but where action is concerned it definitely makes sense. Whereas the effective cause of action in Aristotle’s view is always one of the agent’s pro attitudes, the final cause is the purpose in the light of which the action is taken from the agent’s part. To demonstrate the chains of reasoning which go on in agents’ heads and which lead to decisions, Aristotle uses the terms of ‘practical reasoning’ or ‘practical syllogism’. Those very terms, which G. E. M. Anscombe (2000 [1957], §33) calls “one of Aristotle’s best discoveries”, have confused scholars throughout the centuries. The problem is that, in the end, Aristotle talks about these syllogisms very little, and in addition

---

<sup>32</sup> To be precise, this applies only to instances which share all four causes. Even if Aristotle seems to think that this group covers the majority, he mentions at least one exception. In *Metaphysics* (1044<sup>b</sup>, 10–12), he admits that he is not sure whether an eclipse has a final cause at all.

he gives only a few vague examples of them. In the scope of this study, it is not worthwhile submerging ourselves in that endless study. What is more substantive is merely to appreciate the broad idea here. Besides the fact that we can explain an action by citing its mental causes, there is another way to explain action, and that is to tell *what the agent himself is thinking while making the decision of acting*. Next I will give one example of Aristotle's practical syllogisms from *Movement of Animals* (701<sup>a</sup>, 12–13):

Every man ought to walk. (Premise)

I am a man. (Premise)

---

I'll walk. (Conclusion)

The foregoing example is humorous in all its absurdity, but it nonetheless helps to demonstrate the idea. As I said above, the actual form of syllogisms has aroused a multitude of opinions, but that we may leave to the actual scholars. One thing which should be highlighted before we move on, however, is the nature of the premises in these practical syllogisms. Aristotle (*MA*, 701<sup>a</sup>, 24; *DA*, 434<sup>a</sup>, 16–18) divides the premises into two categories, namely universal ones which are about good, and particular ones which are about possibility. Now, what is highly interesting is how this idea resembles Donald Davidson's view of reasons for action presented in Section 2.2.2. Clearly, universal premises reflect an agent's pro attitudes whereas particular premises link those pro attitudes with certain actions. Most likely the uniformity is not a coincidence, and it is a fascinating fact that Aristotle's view of reasons for action is still highly appreciated in our modern day.<sup>33</sup> Anyway, what is crucial in the light of our study is the role of these syllogisms. It is clear that it is not necessary for an agent to go through syllogisms in his head every time he is performing an action. As Aristotle (*MA*, 701<sup>a</sup>, 25–29) himself states:

---

<sup>33</sup> Actually, Davidson in his "Intending" (85–86) also makes a similar comparison as he too argues that the contents of causally effective mental states can be seen as forming a practical syllogism. However, he (*ibid.* 87) stresses the point that the actual explanations for actions must refer to the mental items, as the syllogisms only rationalize actions.

And as sometimes happens in dialectical questioning, so here the intellect does not stop and consider at all the one proposition, the obvious one; for example if walking is good for man, one does not dwell upon the proposition 'I am a man'. And so what we do without reflection, we do quickly.

So, one does not need to formulate any actual syllogisms in one's head for the sake of being capable of acting, and therefore they seem to be more for explaining actions than to really tell what goes on in an agent's head before the moment of action.<sup>34</sup> The conclusion of practical syllogism is always an action, albeit expressed as a proposition. Logically, an actual action does not, of course, follow from premises expressed in the way we have seen, but the factor which links the premises to actions is Aristotle's idea according to which everybody seeks what he believes to be good. In *Nicomachean Ethics* (1097<sup>a</sup>, 35–1097<sup>b</sup>, 1), he states that happiness (*eudaimonia*) is the final end for all human beings and that it is something one cannot change. Therefore, all things people do are for the sake of their believed goodness and their ability to get us closer to happiness. Now, since human beings always act automatically when they know what would be a good thing to do, Aristotle is allowed to derive an action only from the premises about good and possible.

If we now return to our question of whether Aristotle really is a causalist, we should reconsider how well the condition (2) meets the ideas presented by him. Thus far it has turned out that, in Aristotle's account, actions can be explained both causally and teleologically. While explaining an action the first way, the agent's causally effective mental states must be brought forth whereas, while approaching an action from the second angle, one must reveal what benefit the agent himself sees in his action. Naturally, there arises a question about the relation between the two perspectives. In *Physics* (195<sup>a</sup>, 4–6), Aristotle explicitly notes that one and the same thing may have several causes or explanatory factors. In the end, there seems to be no problem in this account. While referring to mental states explains an action considered from certain angle, revealing how these very same mental states present themselves to the agent himself serves as a valid explanation while using another perspective. Since there is no contradiction between the two perspectives, no problems arise. However, one interesting point here is what Aristotle says elsewhere. In *Parts of Animals* (639<sup>b</sup>, 11–19), he wonders about the ordering of the explanatory factors. He states:

---

<sup>34</sup> Similarly, M. T. Thornton (1982) argues that Aristotle's practical syllogisms should be seen as reasons for actions.

... [T]he causes concerned in natural generation are, as we see, more than one. There is the cause for the sake of which, and the cause whence the beginning of the motion comes. Now we must decide which of these two causes comes first, which second. Plainly, however, that cause is the first which we call for the sake of which. For this is the account of the thing, and the account forms the starting point, alike in the works of art and in works of nature. For the doctor or the builder define health and house, either by intellect or by perception, and then proceed to give the accounts and the causes of each of the things they do and of why they should do it thus.

In this paragraph he quite clearly says that the final cause bears more importance than the efficient one. Of course, it should not be forgotten that Aristotle starts to talk about natural generation, but he still expands his idea to cover human action as well. Even if he considers the subject through the products of human action, it is hard to believe that he would change his opinion in the case of pure acting. However, later he (ibid. 642<sup>a</sup>, 14–17) remembers to once more emphasize that both causes must be taken into account if we are to understand things properly.

In the light of all that has been said thus far, I would not consider it too daring to claim that, in the case of action explanations too, Aristotle values the final cause more than the efficient cause, although he naturally considers both as indispensable for a complete explanation. Crudely, the idea seems to be that, after realizing rationally what should be pursued, the related mental states give the push which drives the agent towards a target considered as good and worthy. Hence, even if the mental causes are an important piece of the puzzle, they nonetheless start to work only after the rational consideration is done (if there has been any). In the cases where there is no rational consideration going on, only a non-cognitive image of the goodness does well. Now, if we need to decide whether Aristotle's view meets the condition (2) of the definition of causalism, it first comes across as credible to stand for the denial. It seems evident that, in Aristotle's thinking, the final cause comes before the efficient one and, if we are to see the picture in the light of modern vocabulary, only the latter is truly a cause in a strict sense of the word. Therefore, actions must not be explained only causally. However, what should come to play its part at this point is Aristotle's idea according to which actions always spring straight from the agent's consideration of what would be the best course of action. In *Nicomachean Ethics* (1145<sup>b</sup>, 24–29), he firstly admits how this famous thought presented originally by Socrates (469–399

B.C.) seems to contradict real-life situations, nevertheless he (ibid. 1147<sup>b</sup>, 13–17) ends up with the position that Socrates, after all, must be right.<sup>35</sup> Later, in Section 2.3.3, I will study the considerations leading to this solution more profoundly, but at this point it is sufficient merely to stress the conclusion. It seems that whatever the agent considers as the most reasonable thing to do materializes as an action. Consequently, it appears more or less accurate to state that, in Aristotle's thinking, the teleological reasons explaining an action always also cause the very action. In other words, if we are to call an agent's grounds for some choice his reasons, we should note that, in Aristotle's account, the relation between these reasons and action is causal. As we will later see, this highly interesting view encounters some serious problems, but for now I will only close this section by saying that, even if it is extremely atypical, Aristotle surely seems to be a causalist as is argued in various instances.

### 2.3.2 Two Aspects of Reason for Action

What Aristotle already understood is something of the utmost importance, and therefore it is a shame how his teachings have been misunderstood or simply forgotten over the ages. Of course, he is not the only one who has emphasized the role of reasons for action as how the agent himself sees the value of his action. For example, Thomas Reid (1710–1796) writes in his *Essays on the Active Powers of the Human Mind* (283–284), originally published in 1788:

[...] I grant that all rational beings are influenced, and ought to be influenced by motives. But the influence of motives is of a very different nature from that of efficient causes. They are neither causes nor agents. They suppose an efficient cause, and can do nothing without it. We cannot, without absurdity, suppose a motive, either to act, or to be acted upon; it is

---

<sup>35</sup> The idea which Aristotle links with Socrates is expressed in various places in Plato's writings. As we know, usually the main character in Plato's dialogs is none other than Socrates who, in the light of his thinking, resembles the historical figure more or less. For example, in *Protagoras* (358d), this Socrates states:

Now, no one goes willingly toward the bad or what he believes to be bad; neither is it in human nature, so it seems, to want to go toward what one believes to be bad instead of to the good. And when he is forced to choose between one of two bad things, no one will choose the greater if he is able to choose the lesser.



equally incapable of action and of passion; because it is not a thing that exists, but a thing that is conceived; it is what the schoolmen called an *ens rationis*. Motives, therefore, may *influence* to action, but they do not act. They may be compared to advice, or exhortation, which leaves a man still at liberty. For in vain is advice given when there is not a power either to do, or to forbear, what it recommends. In like manner, motives suppose liberty in the agent, otherwise they have no influence at all.<sup>36</sup>

Even if there is a huge amount of important observation in the foregoing citation, at this point it is relevant only to emphasize that Reid also realizes that there is an aspect of reason for action which is not about causing but merely about supporting an alternative. Another contemporary philosopher worth mentioning who shares a similar account with Reid is Immanuel Kant (*Groundwork*, 4:448). Recently, the kind of view about reasons for action has been put forward by a continuously growing bunch of philosophers of action. For example, Hugh McCann in his “The Formation of Intention” defends the view according to which reasons for action refer to the contents of mental states instead of the mental states themselves. With this idea in mind, it makes sense to say that two people do something for the same reasons, even though it is impossible for two different people to be in the same mental state. In McCann’s (1998, 151–152) view, the contents of mental states may be the same in separate cases, and this explains the possibility of possessing the same reasons. Afterwards, for instance, G. F. Schueller (2003) has defended a similar position, whereas Jonathan Dancy (2000) has also argued for reasons being propositions but without endorsing the traditional main role of desires and beliefs in action explanations.

What is the most outstanding point in Aristotle’s account is that he realizes that there is more than one way to explain an action (although he emphasizes that all explanatory factors must be acknowledged in order to understand something thoroughly). It undoubtedly seems that instead of having only one aspect, we can use the concept of reason for action in at least two totally different ways. First, we may refer to causally effective mental states when explaining an action. Second, we may talk about the grounds on which a decision has been made. There are also further aspects in the concept of reason for action, but those I will not study in this work. For instance, some philosophers have been willing to distinguish *motivating* reasons from *normative* ones. As an example, Dancy (2000, 2) states that, whereas

---

<sup>36</sup> Even if Reid talks about “motives”, I cannot see any obstacle to taking it as a synonym for ‘reasons’ within this context.

motivating reasons are those in the light of which we act, normative reasons are those which may be considered speaking up for the action due to its moral worth. As Dancy himself writes:

There are these two ways of using the notion of a reason *for* action, which address different questions. There is the question what were the considerations in the light of which, or despite which, he acted as he did. This issue about *his reasons for doing it* is a matter of motivation. There is also the question whether there was good reason to act in that way, as we say, *any reason for doing it* at all, one perhaps that made it sensible in the circumstances, morally required, or in some other way to be recommended, or whether there was more reason not to do it.

There has been a huge amount of discussion about the normative side of reasons for action, but those questions I will not consider here. In the end, instead of being one way to actually explain actions, talk about normative reasons is about evaluating different alternatives for acting from the perspective of ethics.

Even if many philosophers have realized and, hence, also emphasized the role of reasons for an action as only supporting the action instead of causing it, the problem seems to be the same for non-causalists as for causalists. Traditionally in philosophical argumentation, the opposing theory must first be refuted before defending the proposal, and thus the advocates of both sides have been placing numerous arguments against each other. Even if the course of action is quite typical for philosophical discussion, in this particular case it comes across as highly dubious. On what grounds is one allowed to claim that one way of explaining actions or of using the term ‘reason for action’ is wrong and the other way is right?<sup>37</sup> As I argued in Section 2.2, a perfectly valid and, moreover, remarkably popular way of explaining actions is to do it by referring to the causally effective mental states of the agent, or at least to some aspect of those states. However, evidently this way is not the only totally adequate way. One illuminating example can be found, once again, in Daniel C. Dennett’s writings. He (1986 [1969], 158) presents a case where a man starts to psychoanalyze himself when asked why he took a drink. As Dennett states, there is something quite inappropriate answering the foregoing question by saying: “Let’s see. I

---

<sup>37</sup> Naturally, there are also exceptions. For example, Paul Grice makes an incredibly profound plunge into the concept of reason in his *Aspects of Reason* (37–44).

wouldn't be at all surprised if my reason for taking the drink is that I have a death wish and am trying to drown myself'. Most likely, the questioner in the above situation is asking about the grounds for which the man took the drink. He is asking for the benefits which the agent considered there being in the action and due to which he made his decision. This aspect of reasons for action is something very different from the preceding one, yet it would appear fallacious to argue that it is something unreal or meaningless.

Reasons for action as grounds for one are about the subjective understanding of action and its value. The first question should be about what the agent is, in his view, doing, in other words, what his intention is. The second question should be about what kind of value the agent himself sees in his action, in other words, how he would justify the action, most of all, to himself. Even if many philosophers feel quite uncomfortable when too much weight has been put on subjectivity, in this case the subject himself is undoubtedly the only true authority. In other words, if we are willing to know the agent's view about his action and its reasons, there cannot be any other authority but the agent himself. However, obviously there are also some exceptions in this arrangement. First of all, although the agent might be the ultimate authority on his intentional actions, often it is very easy to see what the agent thinks he is doing and for what reasons. This is why we do not need to ask every time what the agent is doing in order to be sure about the matter. For instance, while a person is looking at a book, we are allowed to claim with a great confidence that the person is reading it intentionally. The second point comes very close to this. As we know, people often lie to each other, and even to themselves for that matter, and therefore it is possible that sometimes the best authority is someone other than the agent himself, *even though* there cannot ever be similar certainty in any other instance as there is in the case in which a reliable agent is judging his own actions. For example, in court rooms intentions and reasons for actions may bear remarkable importance when deciding whether to put a man behind bars or even send him to the electric chair, and thus it would not be enough just to ask the accused about them. There would be many criminals running free if we were satisfied only with asking their opinion about their own deeds.

In Section 2.2, I used the terms voluntary, involuntary and non-voluntary as properties with which it is possible to classify actions due to their causal history. Now, I would like to propose another classification with the terms 'intentional', 'unintentional', 'non-intentional' and 'quasi-intentional'. This may be misleading since voluntary actions always include intentional mental states, and therefore they are also intentional in a sense, nevertheless I am

willing to claim that, for truly intentional action, one should require further conditions. Anyway, my suggestion is that an action can be called ‘intentional’ if it matches the agent’s intentions. That is to say that intentional actions are those which the agent himself considers taking. In these cases, we may also suppose that the agent possesses justificatory reasons for his action, because without them it is hard to understand how the agent could consider the specific action as his own. This point is of the utmost importance, and therefore I will contemplate it more profoundly in the next section. The property ‘unintentional’, on the other hand, can be attached to actions in which the agent’s own view of his intentions does not resemble the proposed action. In other words, unintentional actions are actions which the agent performs without noticing. Finally, all the actions performed by agents not capable of use of language should be called ‘non-intentional’. However, even if we are able to handle most of the cases with these three properties, there are also ones which seem to fall under none of these categories.

In one of his papers, Gilbert Harman (1976, 433–434) presents a case of a sniper who has an enemy soldier in his sights. The sniper’s aim is, naturally, to shoot the enemy, but at the same time he knows that firing his gun will alert his enemies to his presence. Now, if he decided to shoot, his killing of the enemy would clearly be a case of an intentional action whereas, say, a killing of a flea in his target’s hair would be a good example of an unintentional action. Anyway, the case of alerting the enemies is much trickier than the foregoing actions. First of all, the sniper knows that, by firing the gun, he will also alert his enemies, and this pleads for alerting the enemies being an intentional action. However, alerting the enemies would not be in any way the purpose of his acting but only an unfortunate side-effect, and by this feature the case differs from the standard cases of intentional action radically. Harman (*ibid.*) himself considers that the alerting of the enemies in the above example is actually an intentional action, but to me this sounds somewhat incorrect. Apparently, there is an essential difference between the acts of shooting the soldier and of alerting the enemies. Alfred R. Mele and Paul K. Moser (1994, 45), on the other hand, use the concept of non-intentional action for the cases of the kind and, in my view, that is a whole lot better option, even though I would prefer to use the terms differently. Instead of ‘non-intentional’ I believe the term ‘quasi-intentional’ would be more accurate. After all, the sniper in question has an intention in his deeds, even if it is a different one. Either way, the idea is the same. Quasi-intentional action is something which the agent acknowledges as being a side-effect of his acting, but which does not drive the agent to act on any level.

Therefore, quasi-intentional actions do not include justificatory reasons. In any case, these quasi-intentions should not be confused with further intentions. For example, if I go to play soccer with my friends, there are many intentional actions to be found in the phenomenon. Among other things, I am exercising as well as I am seeing my friends, yet they are both equally my intentional actions. Even if my main interest in playing soccer were to take care of my physical health, I still have reasons for seeing my friends as well. Consequently, the social interacting is not a mere side-effect of my action. It also has motivational power and, even if it did not have that power as much as the other factor, still it has enough of it to be called a real purpose of my behavior.

Now, it is time to think about the question of what are these intentions and reason for actions as grounds for them. In Section 2.2, I studied the nature of mental states as the causes of voluntary action. I argued that, when explaining an action as voluntary, we must refer to the agent's mental states or at least to the causes or objects of those states. However, it is evident that, where intentional actions are concerned, explanations do not refer to mental states as mental states. By this I mean that explaining an intentional action is not about pointing to the mental states in themselves, even if the explanations may be talk about those states. For instance, when I am answering the question about taking the drink by saying that "I just want to get drunk", I am not explaining the action by referring to the mental state of wanting to get drunk, even though one is allowed to claim that I am in such a state and, moreover, that I am talking about myself wanting to get drunk. In my view, McCann (1998) and Schueler (2003) have a point in calling justificatory reasons "contents of mental states". If we are to analyze the action of taking the drink as a voluntary one, we should explain it by saying that the action was caused by the mental state of wanting to get drunk or, alternatively, we may call the agent an alcoholic or even psychoanalyze him. On the contrary, when considering the same action as intentional, the explanations must say something about how the agent himself experiences that state and how the wanting makes him see some value in the very course of action. I could similarly have answered the question by saying that "I love Martinis" or "I am a much nicer person after a few drinks", and they would have been equally good answers as well. The same goes for explanatory beliefs. If we are to explain from outside why somebody has taken a drink by saying that he believes that drinking is good for health, we are likely suggesting that the belief caused his behavior, whereas the agent himself would explain his own action by providing the content of the belief, namely the proposition 'drinking is good for health'. In the light of all that is said it seems that reasons as ground for

acting are *verbal*. When explaining intentional actions, we form propositions from mental states and, importantly, we do it from the perspective of the agent. Naturally, the same goes for intentions. While an agent is doing something or even only planning to do something, we usually consider him being in a mental state of a kind. Those states are traditionally called volitions or intentional states. Now, if we are to tell an agent's intentions, we must find out how the agent himself sees his states, and therefore we may call intentions the contents of those volitions or intentional states. As Reid noted in a previous quotation (see pages 88–89), since these verbal intentions and reasons for action are “not [...] thing[s] that exist[...], but [...] thing[s] that [are] conceived”, they cannot cause anything. This is the main difference between the two types of reasons. Whereas voluntary actions are caused by mental states, intentional ones are justified with non-causal propositions.

In this section, there is one more issue to be brought up. The whole distinction between voluntary and intentional actions has been introduced by way of taking a glance at Aristotle's thinking. In the end, his distinction between causal and teleological explanations is the foundations on which I have been building my own categories of action. In any case, simply to link voluntariness with causality and intentionality with teleology would be an understandable but, nevertheless, fatal mistake. Even though the target of the agent is more highlighted in the case of intentional actions and the causes are more highlighted in the case of voluntary ones, it would simply be wrong to imply that intentional actions are somehow unconnected with the causal world or that there are no goals or targets for voluntary actions. As has been demonstrated, voluntary actions are also usually taken for a purpose, and thus one rather good way to explain them is to mention the objects of the agent's mental states. For example, there is nothing wrong with saying that a dog does tricks in order to get food. However, what is crucial here is that an action may be voluntary without being intentional. For instance, it is perfectly adequate to say that I am voluntarily making abusive gestures while waving my hands in the street, even though my own intention is merely to have some exercise. In the case of voluntary actions, the crucial point is whether my mental states guide my deeds and whether the things I do spring from within me, whereas in the case of intentional actions the attention is directed towards my own idea of my actions and their reasons. Thus, in the above example, it should be stressed that I am performing the actions of ‘voluntary hand waving’ and ‘intentional hand waving’ as well as those of ‘voluntary insulting’ and ‘unintentional insulting’. In addition, one should be reminded of the fact that, in my account, the four actions are all different as they do not share the same properties (see

Section 1.2.2), and by the same token they cry out for different explanations. Anyway, the point in voluntary actions is not their anti-teleological nature but their emphasis on the external factors with which it is possible to draw a complete causal picture of the action (I will return to the question about the intentions of voluntary actions in Section 3.3.4). Alternatively, in intentional actions, non-causal reasons and intentions are playing the main role. However, this does not imply that intentional actions happen in some world other than the causally behaving one where we live. On the contrary, while reasons are not causal in this picture, some other causal forces are needed. The cause of intentional action is always the agent himself, namely the ‘I’ or the ‘self’ of the agent (later in Section 3.3.5 I will return to this highly interesting question about ‘self’). Whereas in voluntary action the ever-continuous causal chain goes from reasons to actions, in intentional action the chain begins from the agent and continues through the action to its effects. What we should keep in mind, though, is that intentions and reasons in the latter picture are purely verbal, and thus not involved in the causal chain. This does not mean, however, that intentional actions could not share causal power. Next, I will more deeply contemplate this alternative aspect to human action.

### 2.3.3 About Normativity in Intentional Actions

At the end of Section 2.3.1, I mentioned that Aristotle’s view of human action encounters some fundamental problems.<sup>38</sup> The idea according to which human action is always in accordance with the agent’s evaluations does not seem to match our everyday experiences. It is hard to imagine anyone who has not witnessed a situation in which an agent performs an action which he definitely knows is not the best or most reasonable thing to do. As has been said, Aristotle himself also recognizes the pressure from the empirical side, but nonetheless he wants to keep his head. His stubbornness is quite understandable. By linking actions with normative evaluations, we have a harmonious and clear structure by which all actions can be explained and understood. Anyway, a phenomenon where an agent acts against his better

---

<sup>38</sup> Naturally, there are also scholars who are willing to read Aristotle differently (for example, Charles 1984 and Dahl 1984), but to my way of thinking the evidence supporting “the standard view” is much too weighty to be questioned. Even if Aristotle in different places uses varied metaphors to describe the matter, within the chapter actually devoted to the issue (Book VII of *Nicomachean Ethics*) he is quite precise and explicit about it.

judgment is called *akrasia*, and is usually translated as ‘incontinence’ or ‘weakness of will’. Aristotle’s solution to the problem is to claim that, even though people can in a sense act against their better judgment, they can never do it with full awareness. He thinks that, even if an akratic agent possesses the knowledge of the best course of action, sensual and emotional desires are capable of blurring the knowledge enough so that, for the moment of action, the worse course of action comes across as the best option for the agent. Aristotle (*NE*, 1147<sup>b</sup>, 6–8) compares an akratic agent to a man who is drunk or asleep, and states that afterwards the ignorance dissolves the same way and thus the man becomes completely aware of his bad deeds. However, although the situation often goes as Aristotle describes, it is hard to include all the cases of *akrasia* under the same category. Apparently, in some instances we try to convince ourselves that we are doing the right thing and sometimes that happens even without any effort. Suddenly, we forget all our good thoughts as our mind becomes filled with bad ones. Anyhow, it is hard to deny that there are also other kinds of cases. ‘Clear-eyed *akrasia*’ is a phenomenon where an agent clearly realizes all through the decision-making process that the alternative which he is choosing is worse than some other alternative and, as, for example, Donald Davidson (2006 [1982], 143) notes, evidently such cases seem to exist. Certainly, the cases of clear-eyed *akrasia* are rarer than those of average *akrasia*, nevertheless their existence seems undeniable. I truly doubt if there can exist anyone who has not at least once in his life done something which he has clearly and distinctly known to be less good than some other option all through the process. On the strength of the empirical evidence, I am sufficiently confident to claim that actions should not be connected to normative evaluations as strongly as Aristotle does.

Later, during the Middle Ages, the concept of will became an important part of the process of distinguishing human beings from other animals. Aristotle did not have any concept which could be considered as sharing exactly the same role, although in various translations the term has found its place. During the Middle Ages, will was the source of metaphysical freedom for human beings, and by the same token it made it justified to call them to account for their sins as well as for their good deeds. In other words, will was the power to make genuine decisions, and to its credit human beings were not merely at the mercy of their non-rational urges. Even if Saint Augustine has somehow deserved the questionable title of being “the father of the modern concept of will” (for example, Dihle 1982, 123), to my mind it seems a bit of a misunderstanding, at least if we consider ‘modern will’ as an instance independent of the normative evaluations of reason. Indisputably, he



(1991, II) famously tells a story about how he, as a young man, stole some plums, and in that theft he “loved nothing but the theft itself”. However, as should be clear, young Augustine nevertheless sees some value in the very action. As he (*ibid.* II, 16) later amplifies, “my pleasure was not in pears; it was the crime itself”, and thus he evidently admits that he really evaluated the crime, not as good in the ethical sense but as good in its moral unworthiness. Actually, elsewhere he even more explicitly commits himself to the Aristotelian view of the relation between normative evaluations and actions. In *On Free Choice of the Will* (II, IX, 100–102), he states that all human beings seek good and happiness, but the difference is in what one regards as good or as a possible source of happiness. This thought is perfectly equivalent with what Aristotle writes in *Nicomachean Ethics* (1094<sup>a</sup>, 1–3).

The person who wrote the most thorough analysis of human action during the Middle Ages was most likely Saint Thomas Aquinas (1225–1274). As a diligent reader and commentator of Aristotle, Thomas became a loyal follower of his ideas at least in the realm of the philosophy of action. Perhaps the only essential change was the use of the concept of will which Thomas adapted to the construction of the great philosopher of the Classical period. Nevertheless, the connection between normative judgments and actions remained the same while will was nothing but a servant of reason, and all freedom was thereby given to reason (*ST*, Ia 83, 1). In Thomas’ (*ibid.* Ia 82, 4) account, reason makes evaluations about what would be a good thing to do, and the role of will is merely to execute those actions, over and above being the initiating power of the whole process in general. Naturally, there was a huge amount of literature written about human action during the Classical period and the Middle Ages, but if we were to limit our study only to the best-known philosophers, I would claim that René Descartes (1596–1650) at the beginning of the modern period was among the first to detach will at least a tiny bit from the authority of reason and its evaluations. Even if Descartes still emphasizes that all actions are aimed at good, he nonetheless gives more power to will on its own.<sup>39</sup> For example, he (*CSM II*, 40) thinks that, when reason cannot make any decision about the best course of action, will has the power to make the decision between equal alternatives. In addition, in the situations in which there are no good alternatives at all, will may choose whichever option it likes. These kinds of states of indifference Descartes (*ibid.*) calls “the lowest grade of freedom”. Moreover, at least in the interpretation of Gary Watson (2004, 155), will in Descartes’ view is capable of choosing a

---

<sup>39</sup> It is noteworthy that Descartes (*CSM II*, 40) does not consider will to operate only in the practical domain but in the theoretical one as well. In the latter case, its role is to affirm and deny judgments, and its goal is always the truth.

lesser good alternative instead of the better one as long as reason regards both as good. The controversial passage is from *The Passions of the Soul* (177.) and goes as follows: “For if we were wholly certain that what we are doing is bad, we would refrain from doing it, since the will tends only towards objects that have some semblance of goodness”.

All in all, my aim here has been to demonstrate how the connection between normative judgments and choices has been quite strong through the history of the philosophy of action. Although I have only contemplated a couple of philosophers, I dare to argue that the main line of thought has always been the idea according to which actions follow the agent’s view of good or pleasurable. As is said, this kind of thinking has its advantages. First, with the above construction, we would have a more or less comprehensive explanation for all intentional actions. Second, it is hard to imagine any other law governing all intentional actions, and thus intentional actions would become much harder to understand if there was no recurrent pattern. Besides, in most cases the scenario evidently follows the very pattern described, and therefore it seems to be at least not too far from the truth. Keeping these benefits in mind, it is quite understandable that it took so long before people started to argue against the mainstream at full stretch. As late as at the end of the last century there were two papers published both arguing strongly not only against the idea of people always picking the best option but also against the idea of people having to see something good in options they choose. They were Michael Stocker’s “Desiring the Bad: An Essay in Moral Psychology” and J. David Velleman’s “The Guise of the Good”. As they both argue, even though connecting intentional actions with normative evaluations is undoubtedly a tempting idea, it nevertheless makes human beings too rational and, at the same time, too simple creatures, and hence ignores all the other numerous factors which may motivate an agent.

In my opinion, the best counterexample against the traditional account can be found in Velleman’s paper. He (1992, 20–21) presents a case where he himself has fallen into deep bitterness and despair. Even though he in this state of overpowering depression and cannot see sense in any constructive action, he is nevertheless perfectly capable of acting intentionally. As a matter of fact, under the influence of a mood of that kind, a person is likely to express it, perhaps only for the sake of expression itself. As Velleman (ibid. 21) states, those actions do not only spring from despair, but are also done *on the grounds of* despair, and thus the state provides a justifying reason as well as a cause for the nihilistic behavior. Now, at least to me, it seems quite easy to imagine an intentional action which an agent performs without considering the action to possess the merest hint of any kind of

goodness. Velleman emphasizes that it seems even more likely that, if there was something good or valuable in the action, an extremely depressed person would actually avoid doing it. What motivates this kind of person to act is not the value of an action but, on the contrary, the totally negative evaluation of an action. For example, probably by smashing things around him, he would not achieve any pleasure or even relief, since it would be merely a miserly attempt to express the state he is in. Now, of course one could still claim that the depressed agent in the above case sees the expression of his mood as a good thing, but in my eyes it does remarkable harm to the very concept of good. If we are to assert that this person acts in the light of some goodness in the action, the whole concept becomes so wide-ranging that even our darkest urges would become rational powers striving for a better life. Therefore, in view of the convincing example presented, I would rather claim that the long-lasting tradition of binding choices to normative evaluations should reach its end.

Anyway, what can be said about intentional actions at this point? If we are to abandon the necessary connection between normative evaluations and actions, is there any rule or regularity which could help us to understand all intentional actions? In my view, Velleman (1992, 21) not only presents a powerful counterexample to the one of strong traditions in the realm of the philosophy of action, but he also offers us a perfect substitute to fill the space left empty. Now, we should ask what is common in all the various instances of intentional action. We should ask the previous question and consider some examples like the poor Velleman smashing things and the Good Samaritan helping an injured traveler. While we are not allowed to say that they both are trying to do something good, we can say that they are both doing something which *makes sense* to them in the light of their nature and prevailing mood. It would not make sense for an extremely depressed person to help others while he cannot even help himself whereas, for the Good Samaritan, it would not make sense to beat the poor traveler he has found lying on the road. On the contrary, they are both expressing themselves, and hence also doing something to which they can relate. In the case of the Good Samaritan, it is reasonable for him to help a person in need, and thus he can understand himself while doing it. However, sometimes an intentional action is easier to understand if, instead of possessing some worth, the action is totally unworthy, and this is exactly the case with the highly depressed Velleman. What he is doing is totally unworthy, but it is something which the agent can consider as his own action, something coming from his self. After all, if justifying reasons are contents of mental states the agent possesses, it is only natural that he can relate to these mental states. Nevertheless, evidently there are also cases in which an

agent's voluntary actions spring from mental states which he tries to deny or which are somehow hidden in the unconscious parts of his mind. In these kinds of cases, it is harder for the agent to relate to his deeds, yet he must do his best to understand himself or, alternatively, to desperately convince himself of some forced justification for the action. If it turns out impossible, there is no other option but to admit that one cannot identify with the action. This is often the case with, say, people who have drunk too much and thus done things they cannot relate to afterwards. Yet it sometimes is simply more a question of willingness than one of capability.

Intentional action does not have to include any worth, but it must always make sense to the agent on some level. Therefore, the existence of reasons is required in order to distinguish intentional actions from unintentional and quasi-intentional ones. An intentional action is an action which the agent performs knowingly and willingly, and thus it necessarily reflects the agent's nature, mood, feelings, condition and so forth. Only to intentional actions is one able to truly relate. This can be well demonstrated by another brilliant example from Velleman. In *Practical Reflection* (1989, 27–30), he introduces a plain everyday case which is familiar at least to all philosophers who famously are rather absent-minded people. Sometimes we find ourselves in situations in which we have forgotten what it was that we were doing. For instance, we may find ourselves walking down the street without knowing where we are going. As Velleman states, in these kinds of situations we immediately stop walking or whatever we are doing, and we will not continue before figuring out the reason why we are doing what we are doing or even what we are doing in the first place. This is because, in the end, we simply cannot do anything intentionally before we understand the very action. Intentional actions are those to which the agent can relate and which the agent can call his own, and therefore intentional acting logically requires understanding about the very action. Of course, we may invent some new reasons by which we may continue to act the way we have done before, but something must be figured out before acting intentionally. For example, when finding myself walking towards the city center without remembering any reason for it, I can reconcile myself to the situation and continue walking just for the fun of it or, say, for exercise. Nevertheless, I need something which I can consider as a reason for doing it before I can continue. It is plainly impossible solely to lie back and begin to observe one's own actions from outside without really identifying oneself with those actions. Even if we could do that, the actions would be voluntary at best but not intentional. Intentional

actions do not spring only from inside the agent; they spring from the agent's acknowledged self.

#### 2.3.4 The Concept of Will

In the preceding section, I noted that, even though the term 'will' may be found in different contexts with different meanings in the translations of Aristotle's writings, it is overall a quite generally accepted thought that he does not operate with any notion of will which would be equivalent to the one used in the modern philosophy of action (for example, Kenny 1979, vii–viii). Aristotle has a term for a rational desire, namely *boulēsis*, and a term for a rational choice, namely *prohairesis*, but neither of those can qualify as 'will' in the sense that the term has normally been used from the Middle Ages until the modern day. Of course, the whole term is fairly cryptic while philosophers seem to use it whichever way suits their general view of human action the best. Anyway, some generalizations can nevertheless be made in order to outline the term even slightly. Usually, will is considered as a capability only of rational beings. Will is the power to make truly authentic decisions, whereas non-rational beings seem to be more like slaves of their passions. This does not mean that the choices of will must be ultimately rational but only that rational agents are capable of evaluating their options, and therefore also having a genuine impact on what they will do when the time comes. The problem with Aristotle's concepts mentioned earlier is that they are both connected only with rationality in a sense that they exclude all the actions which have their origin, for example, in bodily sources. The problem naturally may not be a real problem if we decide not to operate with the very concept of will, but considering the role it has played through the centuries, it would be really hard to give it up for good.

To my mind, Saint Thomas Aquinas makes an essential observation when he realizes that the term 'voluntary' used by Aristotle is not sufficient to make all the distinctions needed. By the same token, he makes a further distinction between perfect and imperfect voluntariness. He (*ST*, IaIIae 6, 2) states:

[...] it is essential to the voluntary act that its principle be within the agent, together with some knowledge of the end. Now knowledge of the end is twofold; perfect and imperfect. Perfect knowledge of the end consists in not only apprehending the thing which is the end, but also in knowing it under the aspect of end, and the relationship of the means to that end. And such knowledge belongs to none but the rational nature. But imperfect knowledge of the end consists in mere apprehension of the end, without knowing it under the aspect of end, or the relationship of an act to the end. Such knowledge of the end is exercised by irrational animals, through their senses and their natural estimative power.

Consequently perfect knowledge of the end leads to the perfect voluntary; inasmuch as, having apprehended the end, a man can, from deliberating about the end and the means thereto, be moved, or not, to gain that end. But imperfect knowledge of the end leads to the imperfect voluntary; inasmuch as the agent apprehends the end, but does not deliberate, and is moved to the end at once. Wherefore the voluntary in its perfection belongs to none but the rational nature: whereas the imperfect voluntary is within the competency of even irrational animals.

Now, obviously Thomas' concept of perfect voluntariness is fully equivalent to my idea of intentionality, whereas imperfect voluntariness resembles my idea of voluntariness. What Thomas realizes is that, even if rational beings often act like non-rational ones, they nonetheless possess the capability of rational evaluation whether they actually use it or not. Since my act of finding an apple on the ground and eating it quite completely resembles a similar act done by a monkey, the essential difference (presuming that there is one) must lie in the capability of evaluating and understanding the action instead of lying in the actual mechanisms of the actions. Both the monkey and I are capable of possessing mental states, namely a desire and an intention, but only I am capable of interpreting those states and thus to have contents for them. I am capable of understanding that I am eating an apple and that my reason for it is the sweet taste of the fruit in question.

One question to contemplate is whether the whole distinction between rational and non-rational animals is too simplistic or artificial. One may even wonder whether it is so important to make one in the first place. While reading newspapers, one is able to see how

various researchers are trying to prove that non-human animals can also be capable of rather complicated rational processes or even of using language on some level. However, I am not willing to say anything about the question of how close those non-human animals can get to our rational abilities, since my knowledge about the matter is insufficient. On the contrary, what I am willing to claim is that the idea according to which intentional actions require a capability to use language in a fairly advanced manner is the best way to distinguish two fundamentally separate levels of action and hence to find a lucid place for the complex concept of will.<sup>40</sup> Since my intentional eating of an apple requires that I am capable of understanding that ‘I am eating an apple’, we have a natural way to distinguish agents capable of intentional actions from those who only voluntarily follow their instincts. Furthermore, as well as understanding the intention, I am also capable of understanding the reasons for the intention, and only with those very reasons am I able to truly identify with the intention. This is not to say that no non-human animals could be capable of the same, but only to show the level they need to achieve in order to be considered as capable of acting intentionally. The difference between being an authentic source of decisions and being a mere playground for sensual and emotional whims lies in the capability of using symbols, because only with symbols we can understand and evaluate our actions deeply and profoundly enough. The ability to evaluate actions on rational grounds provides sources of motivation which are really hard to associate with any non-rational animals. Whether we use the ability is not a question of whether we have a will or whether we use it, but solely a question of how well we use our wills. All in all, will is not in the actions, but in our capability of understanding our actions and their reasons.

Here I have argued that intentional actions and the concept of will are both founded on the same principles. Next, I will continue with this thought and claim that actually will is nothing but the capability to act intentionally, namely to act in a way that the agent can identify with his actions. Therefore, to use one’s will is simply to act intentionally. In the history of the philosophy of action, the act of will is often associated with some kind of mental event or act which makes the action happen. This line of thought may be captured fairly explicitly if we take a look at John Locke’s writings as an example. He (1975, II, XXI, §30.) states that “[f]or he, that shall turn his thoughts inwards upon what passes in his mind, when he *wills*, shall see, that the *will* or power of *Volition* is conversant about nothing, but our own Actions; terminates there; and reaches no farther; and that *Volition* is nothing, but

---

<sup>40</sup> Similar thoughts have been expressed by Anthony Kenny (1975, 20).

that particular determination of the mind, whereby, barely by a thought, the mind endeavours to give rise, continuation, or stop to any action, which it takes to be in its power". I want to emphasize that there may be nothing wrong with this picture, even though, for example, Gilbert Ryle (1949, III, [2]) has famously placed his heavy critic on the very idea. Clearly, there are mental states which may be linked with actions and even be considered as their causes, yet the problem is that similar states can be found among all creatures. Again, if we are willing to regard the power of will as something special, we need to associate it with something more complicated than mere mental events. In the light of philosophic tradition and its use of the concept of will, it would be too peculiar to consider, say, earthworms and amoebas using their mighty power of will. As is said, there is not necessarily anything wrong with this way of using concepts, but I myself prefer to use them in a different way.

In *Philosophical Investigations* Ludwig Wittgenstein (615.) says:

“Willing, if it is not to be a sort of wishing, must be the action itself. It cannot be allowed to stop anywhere short of the action.” If it is the action, then it is so in the ordinary sense of the word; so it is speaking, writing, walking, lifting a thing, imagining something. But it is also trying, attempting, making an effort,—to speak, to write, to lift a thing, to imagine something etc.

To my mind, Wittgenstein talks about will and willing in a highly sensible manner. Whereas in Locke's thinking will is the mental cause of actions, Wittgenstein emphasizes that it should not be considered as something which happens before actions, but which belongs to actions and to attempts to act themselves. When we are, say, walking, we do not first will after which the action happens as an effect of the willing. On the contrary, our intentional walking is the willing, and thus to say that we are willing our walking is absurd speech if we do not talk about will in a sense of a rational desire or wish. To will is to do something intentionally, and the requirements of intentional action define whether we should talk about willing within the context.

One interesting feature in the concept of willing is the possibility of pure willing. Earlier I defined intentional action as one where the action performed corresponds with the intentions of the agent (see page 92). That is to say that a person does something intentionally, if he shares the intention of doing it. However, sometimes people act with intentions which do not mesh with the actual events. Firstly, one may be mistaken about



some feature in one's action. For example, a person may believe that he is waving to the Pope, although the old man he sees is somebody else. In this case, what he is willing to do is something other than what he is actually doing, and thus his 'waving to an unknown old man' is an unintentional action even if his 'waving' is intentional. Secondly, it is also quite common that our willings remain mere willings. By this I refer to situations in which we try to do something but fail. For instance, a person may try to hit a baseball but miss. In this situation, the intentional action is perhaps best described as 'trying to hit a baseball' as we cannot really say that he performed an intentional action of 'hitting a baseball' while actually missing. However, what we could say is that what the person was doing was that he was intentionally hitting a baseball while trying it, but the action turned into no more than a mere unsuccessful attempt. On balance, what is interesting in this scenario is that, even if actions fail, the attempts do not vanish by the same token. What I am underlining here is that, even if I do not associate willing with any mental event, it is still right to say that a person is willing something while he is merely attempting it. If willing is intentional acting, for the use of one's will it is sufficient only to act whether it comes with success or no. As Rogers Albritton argues in his brilliant paper "Freedom of Will and Freedom of Action", with chains and ropes we can only limit an agent's ability to move instead of tying down his will. For willing even a possibility to try is enough. Anyway, with this statement I end the first part of my study, and thus I will address the question of the freedom of will, among other things, in the latter part.

**PART II:**  
**THE APPLICATION**

### **3. Actions in Scientific Mechanisms**

#### **3.1 Two Types of Mechanisms**

In the first chapter of my work, I took a glance at how the idea of scientific explaining has developed from the strict DN model to more flexible explanatory mechanisms. Of course, a great deal also happened in between the two, but my idea was not to give a complete review of the discussion in any case. On the contrary, I believe that the best way to understand scientific mechanisms is simply to compare them with the DN model, and therefore I proceeded straight from the DN model to explanatory mechanisms. Through the faults of the first, it is easy to understand the virtues of the latter. However, it should not be forgotten that the idea of causal explanation itself has its intrinsic problems, and hence it seems impossible to present an account of explaining which would work perfectly in all instances. Nevertheless, explanatory mechanisms function quite well in the widest range of fields. Naturally, the looseness of the concept puts it at risk of allowing worthless explanations, but the looseness is needed if we ever want to find a model of explaining which could serve all fields of science. For example, in social sciences the requirements of a valid explanation should not be too strict in any sense.

In the second chapter, I left mechanisms aside as I devoted the whole chapter to the concept of action. I argued that the concept as a whole is extremely difficult to define comprehensively enough, since actions can be understood in very different ways. Perhaps the only overarching feature with all the various types of action is the rather vague notion that when acting “something or somebody is doing something”. Consequently, I distinguished three different categories which all contemplate actions from totally different perspectives. All these categories present actions as having a place one way or another in the same causally functioning world, but nonetheless they highlight different features in actions and, in particular, in their explanations. From the perspective of pure physicality, one is naturally interested in the physical properties of action as well as in its physical causes. Voluntary actions, on the other hand, bring up mental factors. However, the class of intentional actions

is the most peculiar one. Whereas in the first two cases we explain actions causally, while talking about intentional actions we are interested in reasons and intentions which are not in a causal relation. Intentional actions, nonetheless, belong to the causal world, but the factors by which they are explained, namely reasons as grounds, do not explain them causally. All in all, what is crucial here is that in one and the same phenomenon one may find various actions. Moreover, by adding properties like ‘voluntary’ or ‘intentional’ to action, we bring out different aspects, and thus end up with specific actions with a certain kind of causal nature. To be precise, after the stance is chosen, we are not talking about the same action any more as the properties attached to it are radically multiplied or changed.

Now, in this third chapter it is time to apply the lessons learned in the two previous ones. From the beginning, my interest has been in the pictures which mechanistic explaining draws of acting individuals, and finally we have the necessary equipment to take the main issue under scrutiny. The question is about how human action should be understood in different kinds of scientific mechanisms and, additionally, outside of them. In other words, what roles can agents play in causal mechanisms and what kind of features should be associated with these roles? For example, how should we understand the self or the “I” of an individual being working as a cog or wheel in a causal mechanism? In this chapter, I will argue that the images of the acting individuals in causal mechanisms are very different from those of intentionally acting agents, and therefore they should be sharply distinguished. I will once more compare the different perspectives on action with each other, and show that it would be better to use different concepts while changing from one to another. Even if this changing of concepts cannot be carried out in practice, their altering meanings must be acknowledged.

However, before proceeding with our investigations, one more distinction must be made. If we are to consider the role of human actions in the mechanisms of various sciences, I believe that those mechanisms should be divided into two separate groups on the grounds of how they treat actions. First, human beings and their actions can be seen as mere parts of the physical world so that the explanations do not rely on any special features which depart from the order of the surrounding physical world. This means that there are no non-physical entities involved due to which actions could be distinguished from other physical events, although there is one extremely interesting exception, namely mental actions themselves. A good example of a mental action is mental arithmetic. Mental actions are actions as well, yet it is really hard to study and explain them without taking the mental aspect into account.

After all, the mental occurrence is the essence of such an action. However, since it is generally believed that mental activity has some relation to neural activity, the physical mechanisms in question try to explain mental activity by studying physical events in brains. Therefore, through the perspective of pure physicality, mental activity becomes something which can be reduced to physical phenomena at least on some level, and hence it loses its special status. All in all, if one is to contemplate actions only in the light of the physical story, evidently the study must follow the same rules shared with all the other physical studies, namely the rules of physical causality. By this I mean that there should not be any non-physical factors involved in these stories. Second, there is another group of mechanisms, and in these mechanisms actions share a rather special role. What I mean by ‘special’ is that, in these mechanisms, actions depend on certain features which do not belong to the physical realm and which thus make actions stand out from the even mass of pure physical events. Naturally, they are the inner dispositions of agents which lift them above the mere physical level. On this new level, there are also new concepts, like ‘will’ and ‘pro attitude’, needed to cover the various features of this more complex sense of agency. Whether the actions are to be found in *explanans*, *explanandum* or both should not matter in the issue. On balance, this second group of mechanisms contains those which emphasize the specialty of human actions in the world of events, and which thus also base their explanations on such entities which may not have any physical counterpart or at least that counterpart is not the object of interest.

I have a strong urge to refer to this distinction as one between the mechanisms of natural and social sciences, even though I acknowledge that it does not fall under this categorization without problems. By natural sciences I mean the branches of sciences aimed at studying the natural world, whereas by social sciences I refer to those concerned with society and humanity overall (often ‘social sciences’ are used as an umbrella term for the fields outside the natural sciences). Clearly, for example, the mechanisms of biology as a whole cannot be placed in either of the sides in my categorization. Biology is, as a matter of course, a good example of a natural science as its aim is to explain and understand the phenomena of nature. However, even if there are branches like molecular biology which divide the human body into pieces and thus study it as no more than as another part of nature, traditional biology offers a special role to the actions of human beings as well as to those of other creatures. These actions play an important part within nature, and in traditional biology they are usually understood and explained by similar factors as in social sciences. However, the problem does not lie solely in my distinction. The field of evolutionary psychology is one

good example of a branch which is really hard to place under either natural or social sciences. On the contrary, it seems evident that it shares elements of both as it tries to explain the actions of human beings within the social context with the terms of evolutionary biology. In any case, even with all the exceptions, I am ready to take a risk and talk about the distinction between natural and social sciences when I want to highlight the difference between the abovementioned two groups of mechanisms. Since all categorizing is more or less artificial, I am confident that it cannot do much harm to use these somewhat misleading terms. However, this decision has one more consequence worth mentioning. If we are to place the mechanisms of traditional biology under the mechanisms of social sciences, we must also accept the fact that they bring mechanisms involving non-rational animals along with them. Of course, it sounds quite contradictory to allow such a mess as letting dogs and cats enter the social context, but there is certain sense in it. After all, as I tried to argue in Chapter 2.2, the same pattern is used in explaining voluntary human actions and the actions of non-rational beings, although the first ones can naturally be of a higher complexity. So, without further ado, I will start by contemplating the mechanisms of natural sciences under this classification.

## 3.2 The Mechanisms of Natural Sciences

### 3.2.1 The Idea of Mechanical Philosophy

When the term ‘mechanism’ is used to refer to explanatory mechanisms in science, it is often a source of confusion in philosophical discussion. This is because the concept also contains another sense which is even more widely acknowledged among the philosophers of science. The other sense stems from the writings of the philosophers of the early modern era who shared a more or less uniform view of the basic structure of the world. Even though these mechanisms or ‘mechanical philosophies’ may come across as somewhat radical to a modern reader, they have nevertheless had a crucial influence on the modern world view at least in Western countries. As Aristotle during the classical antiquity laid the foundations of science for the whole medieval period, during the early modern era the tradition was replaced, and by

the same token the new foundation was laid for the forthcoming centuries. Therefore, an understanding of these mechanical philosophies is important for grasping a wider picture of modern scientific thinking.

I will devote this section to the historical overview of the original sense of the term ‘mechanism’ in the philosophy of science. I hope that it will help us to clarify the picture by dissolving the confusion hanging in the air around the concept of mechanism. Of course, there lies a source of further misunderstanding as talk of mechanical philosophies may blur the picture even more if it is on any level associated with the mechanisms of the modern day sciences. Therefore, it is impossible to emphasize too greatly the fact that *these two ideas should not be confused*, although they naturally share some similar features. Mechanical philosophy is most of all a world view, whereas mechanisms in science are no more than explanatory devices. Moreover, as I noted on page 56, the adherents of mechanistic explaining usually shun all kinds of excessively wide-ranging interdisciplinary explanations. After all, a mechanism in this sense is always an answer to a question, and thus it sounds far-fetched that there could be one answer with which all questions could be answered. Mechanical philosophies, on the other hand, are less answers to questions as they are rules according to which those questions should be answered. As they tell us something about the basic structure of the world, by the same token they tell us something about the nature of adequate answers. This is a distinction which should not be forgotten while the similarities are taken under scrutiny. Anyway, reducing confusion and increasing the level of education are not the only reasons why we should take a look at the traditional meaning of ‘mechanism’ in philosophy. As they present the world in its strictest possible form, through them the conflicts between different kinds of explanations are also more easily demonstrated. Under the circumstances, before anything else, I will take a brief glance at the birth of the mechanical world view and mechanical philosophies.

The idea of mechanical philosophy can be traced back to the scientific revolution which took place in Europe during the 16th and 17th centuries. With his book *On the Revolutions of the Heavenly Spheres*, Nicolaus Copernicus (1473–1543) is usually considered as the true initiator of the revolution. There he presents his heliocentric world view which means that he places the Sun at the center of the universe instead of the Earth. Similar thoughts had already been expressed in the classical antiquity, but Copernicus played the role of bringing them out again as he defied the official opinion of the Christian Church. However, although Copernicus gave the initiating push to the process, the revolution itself

did not start properly before the time of Johannes Kepler (1571–1630) and Galileo Galilei (1564–1642). They were both loyal followers of Copernicus’ thinking, and thus aimed to substantiate his thoughts. Galileo especially achieved essential results in the light of the mechanical world view. Since the classical antiquity, Western scientists had been following Aristotle’s thinking, but the heavy critique from Galileo’s side changed the scientific world view for good. For example, Aristotle’s (*Ph*, III, 1.–3.) account of motion is based on a thought according to which motion must involve the very essence of the body becoming enhanced or fulfilled. As Richard S. Westfall (2007 [1971], 19) notes: “Local motion—what alone the word ‘motion’ means to us—had been to Aristotle only one example of a much broader conception meant to embrace all change”. Westfall demonstrates this by giving examples; for Aristotle, the education of a youth and the growth of a plant are instances of motion just like the vertical fall of a heavy body is. However, the crucial move from Galileo’s part was to separate motion from the nature of bodies. As Westfall (*ibid.*) puts it, “[m]otion is merely a state in which a body finds itself; and as Galileo repeated over and over, a body is indifferent to its state of motion and rest”. If we take a falling stone as an example, the difference is easily demonstrated. For Aristotle the stone falls to the ground because it is in its essence to be close to the ground, whereas for Galileo it falls not because of its nature but because of the external forces that have an impact on it. In Galileo’s picture, the motion does not depend on the essence of the stone, since all motion, as motion, is identical. This means that all motion should be explained with the same factors. This new way of thinking was fundamental to the new Copernican world view as it helped to explain the possibility of being in movement without even noticing it. Because we are indifferent to motion, it is possible that we live on a planet which keeps moving constantly at immense speed.

Even only in the question of motion the difference becomes quite definite. Whereas in Aristotle’s account all non-violent motion springs from the essence of a particle, Galileo pointed out that all particles are at the mercy of the surrounding world of causes and laws, so their essences do not play any role in the case of motion. Suddenly, all motion followed the same rules so, at least in this aspect, the essence or nature of things was separated from their physical presence. Moreover, in Galileo’s view, motion in itself does not need any specific cause but only the changes in it. As A. Rupert Hall (1954, 87) puts it, “[i]n Aristotelean physics [...] rest was normal and motion a state requiring special explanation; in Galileo’s physics where space was a vacuous geometrical framework [...] only changes in the state of motion or rest required explanation”. Now, even this rather shallow understanding of the



foundation laid by Galileo helps us significantly to see the circumstances in which René Descartes wrote his philosophical studies. For mechanical philosophy, Descartes is definitely the most important figure as his dualism was the required metaphysical foundation for the new scientific understanding. The reformations of this rate have their impact on all understanding of the world, and thus, by the same token, the metaphysical side of science needs to be updated. For this purpose, Cartesian dualism was simply perfect.

As I believe is widely known, Descartes (*CSM I*, 63.–65.) considered the world composed of two types of substances which are distinct on all levels and in all possible aspects. These substances are the mental and the physical, namely the mind and the body. Whereas the essence of the mental substance or spirit is thinking, the physical or material substance is best characterized by the attribute of extension. Because the two substances must be completely separate, they cannot share a single common property. This means that all the features traditionally associated with matter cannot be found in spirit, and *vice versa*. The mental substance cannot have any form or location, for instance, and therefore it is hard or even impossible to examine it empirically. On the other hand, the realm of matter naturally shares all these features which thus make it a perfect research subject for the natural sciences. With this division, Descartes defined the role and target of sciences by taking away the spirit and leaving all the rest under inspection. For Descartes, spirit is an active and enigmatic force, whereas the physical realm is merely passive and inanimate machinery governed by certain rules and laws. The task of a scientist is to reveal those laws, and hence to make sense of the manifestations of this enormous body.

Even though only a few of the contemporary scientists were ready to follow Descartes all the way through, there were many who approved the main line of his thinking, namely dualism. As before, most philosophers had considered nature as consisting of a boundless amount of organic occurrences, now the world was seen as composed of inanimate mechanisms which are not only easier to study but also easier to control. In Descartes' account, the world manifesting itself to our senses is just an enormous mass of matter governed by causal laws. Nonetheless, in this picture there is at least one important exception which is the human being who is also able to think and to have a soul.<sup>41</sup> The most interesting curiosity in this special creature is its twofold nature. Even if it is capable of having mental abilities and properties, it still has its place in the physical world as well. This twofold role of

---

<sup>41</sup> Descartes (for example, *CSMK III*, 361–362) talks about angels and God as well, but I will not take them under scrutiny in the scope of this study.

human beings is of the utmost interest. Whereas the soul or mind of human being is naturally mental, the body in itself is a sheer mechanism being a part of the world of matter and laws of nature. This combination has highly interesting consequences which I will consider later in my work. Anyway, one peculiar piece of demarcation from Descartes' side comes within the rest of the animal kingdom. As he strongly adhered to dualism, he excluded all non-rational animals from beings possessing mental abilities and properties. In Descartes' (*CSMK III*, 365–366) thinking among mortal only human beings can have a soul and this fact leaves the rest of fauna only with a bare body free of all thoughts. However, he (*ibid.* 366) did leave some special features to dumb creatures as he stated that “I do not deny life to animals, since I regard it as consisting simply in the heat of the heart; and I do not even deny sensation, in so far as it depends on a bodily organ”.

We may now leave aside the soul and its mysterious nature for a while, since our main interest in this section are natural sciences. So, in accordance with Descartes' dualism, the substance of extension, namely the mechanical body of the world around us, became the subject for natural sciences. Consequently, the philosophers and scientists did their best to unravel the laws governing the physical world as well as the basic nature of this world in itself. Naturally, opinions agreed and differed in various aspects. According to Descartes, the universe is a plenum what means that no void spaces exist among matter. He thought that matter extends covering the entire universe as it is infinitely divisible. On the other hand, another contemporary “mechanist” Pierre Gassendi (1592–1655) presented an alternative view of the physical world. Gassendi was an atomist meaning that he believed that the universe consists of an infinitive number of extremely small units which cannot be divided. This kind of view implies that, behind or around the units, there is a void in which those pieces of matter have their locations. (Westfall 2007 [1971, 39]) Therefore, contrary to Descartes, Gassendi did not think that the matter comes down solely to extension as it can be characterized also by solidity, hardness, resistance and impenetrability (Boas 1952, 430). The atomist world view was not a completely new invention as there had already been famous atomists in the classical antiquity. As the Aristotelian account of physics suffered from inflation, Democritus (460–370 B.C.) and Epicurus (341–270 B.C.), for instance, attracted fresh attention. However, even with all the disagreements, the two mechanists were nonetheless of the same mind about many essential issues. All in all, they thought that the physical side of the universe consists of qualitatively neutral matter whose motion produces all the phenomena there are to be found within the very realm.

In any case, the concept of mechanical philosophy is usually most strongly associated with Robert Boyle (1627–1691). Even though he himself tried to make very plain that he did not belong to either of the abovementioned camps, it is nevertheless rather evident that he built his own “corpuscular” world view on the foundation laid by Descartes and Gassendi. He also thought that the whole physical world consists of matter and motion, and that all the perceived phenomena can be reduced to these two basic elements (Boas 1952, 267–269). However, contrary to Descartes, he believed in small, solid and indivisible particles which are the fundamental building blocks of matter forming clusters and further clusters *ad infinitum*. Boyle thought that the particles have their own shape and size, but unlike Gassendi, Boyle denied that they could have any special “inner nature”. In Boyle’s account, the motion of particles does not spring from them or their nature as the whole construction is ordered and governed by God. In this respect, Boyle paralleled Descartes. To the highly religious Boyle, it was also important to prove that, contrary to general opinion, an atomistic or mechanical view does not need to be inconsistent with Christian outlook. Boyle often described the world as machinery built and governed by God (Cook 2001). However, in the end, Boyle was not so keen on creating a theoretical world view as he was on merely studying empirical science. For him, to consider the universe as a mechanical structure was more like a methodological fact based on experimental data than a pure theoretical hypothesis (Boas 1952, 487–488).

Keeping all that has been said in mind, it seems rather difficult to give a general description of the idea of mechanical philosophy. Evidently, the mechanists of the early modern era had quite varied opinions of what kind of mechanism the physical world is. Other mechanists worth mentioning are Thomas Hobbes (1588–1679), Christiaan Huygens (1629–1695) and John Locke (1632–1704), at least on some level.<sup>42</sup> Mechanical philosophies have mostly been tightly related to contemporary discussions and achievements in science, and hence an account which could be considered as the archetype of the whole idea does not exist. Nevertheless, there are some features which can be quite generally associated with the very idea. Typically mechanical philosophies are materialistic, atomistic and, most importantly, deterministic theories of the basic structure of the physical world. In the next section, I will continue with mechanical philosophies but focus mainly on the last aspect.

---

<sup>42</sup> Lisa Downing (1998) has written an article about Locke’s relation to mechanical philosophy.

### 3.2.2 Determinism *versus* Indeterminism

Now, as both senses of ‘mechanism’ have been introduced, the difference between them should be quite clear. For example, Garland E. Allen (2005, 263) studies the two in the context of biology and ends up describing them as follows. About mechanical philosophy, he states that it “[...] is the view that likens organic (or other complex) entities to the interaction of material components in a machine”. On the other hand, he describes an explanatory mechanism as “[...] the step-by-step description/explanation of how the components in a system interact to yield a process or outcome: for example, how an enzyme molecule interacts with its substrate molecule to yield a product, or how neurotransmitter molecules interact with membrane-bound receptors to produce an action potential”. Perhaps the difference does not need to be further stressed. Regardless of the context or field, mechanical philosophy is always primarily an ontological statement, and therefore presents rules for explaining rather than giving any actual explanations. According to mechanical philosophy, all or some specific phenomena can be reduced to the sole movement of matter, and hence any explanatory mechanisms relying on mechanical philosophy must be given according to the very idea. Obviously, explanatory mechanisms do not need to rely on mechanical philosophy or, equally, on any other specific ontological view. As was noted in the first chapter, explanations are merely answers to why-questions, and thus their nature depends on the given context. Of course, explanatory mechanisms always commit themselves to a particular idea of causal ontology, but these commitments may vary a great deal.

One thing worth noticing is that the scale of mechanical philosophy does not have to cover the whole universe. For example, in the case of human behavior one does not need to make statements about the basic structure of the physical world as a whole. On the contrary, in contemporary discussion it is more common to talk about local mechanisms than to insist that the whole world behaves deterministically under some strict laws of nature or diktats of God. Evidently, the best example of a local mechanism would be the hypothesis of neurophysiological mechanism. The term refers to the hypothetical situation where a neurophysiological theory, which can completely explain and predict all the movements of human bodies except those stemming from outside forces, is established. This idea assumes that human body is a comprehensive causal system which follows some strict laws of nature.

Therefore, it would be possible to reveal these laws and to uncover the underlying mechanism by which we would be able to see distinct neurophysiological causes for all movements springing from inside the agent. Naturally, these theories have not yet been found, and it is questionable whether the whole idea is plausible, but nevertheless the hypotheses of the abovementioned kind have served as fruitful stimulants for theoretical speculations. The most famous example of this is perhaps Norman Malcolm's paper "The Conceivability of Mechanism" in which he contemplates how the revealing of such a theory would affect our lives.

Anyway, the most enthralling feature in all mechanical philosophies seems to be determinism which means that within mechanisms everything happens of necessity. In other words, within deterministic models all events are consequences of some other events in the past, and the causal connection between them is governed by some strict laws of nature. Thus, determinism denies all genuine alternatives, as, in the end, there is always only one possible outcome for all particular events. Naturally, contemplating the idea of determinism brings up the question of the power of making genuine choices. If everything is settled from the beginning, how could anyone make any real decisions ever? This very question has been one of the most inspiring problems in philosophy through the centuries. For example, David Hume (*EHU*, VIII, I, 73) calls the juxtaposition of liberty and necessity "the most contentious question of metaphysics, the most contentious science". Since the problem comes very close to the life of an ordinary man, not only professional philosophers have spent their days wondering about the notorious tension. Leo Tolstoy is one good example of this as he mulled over the very question in various instances. In *War and Peace* (1952, 688), he writes that "[i]f there be a single law governing the actions of men, free will cannot exist, for then man's will is subject to that law". Later he (*ibid.*) continues by stating that "[t]he problem is that regarding man as a subject of observation from whatever point of view—theological, historical, ethical, or philosophic—we find a general law of necessity to which he (like all that exists) is subject. But regarding him from within ourselves as what we are conscious of, we feel ourselves to be free". In this quotation from *War and Peace* lies a great deal of wisdom, even though there is something which Tolstoy himself seems to miss.

As is said, whether the world has been considered a predestined story written by God or mere automatically working machinery, the difficulty of finding a place for free will has been a significant problem for philosophers as well as for others. Naturally, several cliques have developed. *Compatibilists* believe that determinism and free will can, in fact, fit in the

same picture without being mutually exclusive. This could be made possible by redefining the concept of free will in a manner which does not require a possibility of changing the course of the predetermined physical world. *Incompatibilists*, on the other hand, deny such a possibility. They think that the true concept of freedom cannot be defined without the real possibility of choosing between alternatives. This means that they insist that, in order for people to be free, the world simply cannot be predetermined. Obviously, there are two different schools within this group. First, *libertarians* believe that human beings must possess free will, and hence determinism cannot be true. Second, *hard determinism* is the view according to which the world is deterministic, and this fact makes it impossible to find a place for any notion of genuine free will. The debate has reached enormous proportions. Besides the fact that the question of free will is of the utmost importance for all of us personally as we are, undoubtedly, constantly choosing and acting individuals, the whole issue also has many further implications which cannot be ignored. For example, the idea of moral responsibility is usually attached only to rational, freely choosing beings. Therefore, if the possibility of free decisions vanishes, all talk containing hints of moral praise and blame may lose its meaning by the same token.

Considering all the fuss about the possible threats of determinism it would sound rather harsh to challenge the importance of the whole issue. My aim is, nonetheless, to do so. In the beginning it must be stressed that, all in all, it is quite difficult to say anything about whether determinism is a plausible view of the world we live in. With the results of contemporary research in the field of physics, the universe is beginning to come across as more complicated than people could ever imagine. Perhaps it, for the moment, seems unlikely that some kind of determinism could prevail, at least on the most fundamental levels of the universe, nonetheless I would feel more comfortable merely avoiding having to choose side in this question. After all, I have good reason to do so as I cannot fully understand the importance of the debate between determinism and indeterminism to the question of free will. Actually, to my mind, Tolstoy already highlighted the most crucial point of the issue in the above citation. He pointed out that the whole problem culminates in the tension between two different viewpoints. From the objective viewpoint, one can reduce the acts of men to causal chains stemming from earlier history, whereas from the subjective viewpoint an agent himself is the actual source of his actions. From the latter perspective, the agent not serves as a route for a chain of events as he is the one who initiates a new one. The same distinction is brought out by Thomas Nagel in his *The View from Nowhere* but with one important novelty.

As he (1986, 110) states, this very problem does not, in the end, require the idea of necessity in order to achieve its form in a full scale, although it naturally dramatizes the matter a great deal.

In my view, the greatest achievement in Nagel's book is that he finally sets the phrasing of the question in the right direction. As is known, human beings are creatures of a special kind in their ability to reflect. This means that they are capable of not only living and acting in the world but also of stepping out of themselves and thus viewing themselves doing what they do as parts of the whole. Nagel's book concentrates on the two very different standpoints from which we can study the world and, most of all, ourselves. While acting we see ourselves from the subjective viewpoint and there we naturally play an important role in making things happen. From the objective viewpoint, on the other hand, we see the world from outside. There our important role starts to diminish as we see how all authentic decisions are mere pieces of the never-ending flow of the stream of events. Now, instead of fussing about the true nature of the physical world, Nagel states that solely within this very division lies the source of numerous philosophical problems of which the question of free will is a good example. I think he is right. Nagel (1986, 110) writes as follows:

Something peculiar happens when we view action from an objective or external standpoint. Some of its most important features seem to vanish under the objective gaze. Actions seem no longer assignable to individual agents as sources, but become instead components of the flux of events in the world of which the agent is a part. The easiest way to produce this effect is to think of the possibility that all actions are causally determined, but it is not the only way. The essential source of the problem is a view of persons and their actions as part of the order of nature, causally determined or not. That conception, if pressed, leads to the feeling that we are not agents at all, that we are helpless and not responsible for what we do. Against this judgment the inner view of the agent rebels. The question is whether it can stand up to the debilitating effects of a naturalistic view.

In the end, whether events do or do not follow some strict laws of nature does not appear so relevant any more. Even in an indeterministic world, events must have their causes as it is extremely difficult or even impossible to imagine something happening without a cause. Now, whether an event is the necessary effect of its cause or not, does not change the fact that

it is nevertheless its effect. This means that the cause has made the effect happen and as the cause must in itself have a cause, the stream can be traced back to infinity. If we are to consider the flux of physical events, it seems quite difficult to find any place for will, let alone its freedom. All in all, even to substantiate that the world is indeterministic is not enough to solve the problem of free will.

Although Malcolm (1968, 1.–2.) also emphasizes the role of necessitating laws in his idea of neurophysiological mechanism, at the same time he directs the debate towards the question of the possibility of more than one independent explanation for one and the same phenomenon. In my view, this is a more crucial question to the solving of the problem of free will. If we follow Malcolm and imagine there being this kind of all explaining neurophysiological theory, it will necessarily challenge all the other possible explanations for actions. As Malcolm does, we should ask what role could all the explanatory mental states, for example, play if there was already one perfectly complete explanation for the action under scrutiny. Malcolm (*ibid.* 22.–27.) himself derives rather interesting conclusions from the matter. He goes as far as claiming that the vindication of neurophysiological mechanism would have such powerful implications that even our language would lose its meaning. Therefore, he deduces that the vindication of mechanism would in itself be a sort of paradox, since the vindication would be rendered meaningless by the same token. However, those are somewhat peculiar conclusions. Anyway, in all likelihood, the problem has become sufficiently well demonstrated already; while giving purely physical explanations for actions, it is really hard to find any role for alternative explanations which would include more intrinsic elements of acting than mere physical happenings. As Nagel (1986, 111) puts it:

There seems no room for agency in a world of neural impulses, chemical reactions, and bone and muscle movements. Even if we add sensations, perceptions, and feelings we don't get action, or doing—there is only what happens.

Perhaps it is unnecessary to once more state that this goes for all explanatory mechanisms which explain human actions by mere physical entities and processes. The fact of whether the explanatory mechanism itself is a strict one obeying some explicit laws of nature, or more like a loose sketch including indeterministic relations, does not change the drift. In the end, within both alternatives one can provide an explanation sufficient to challenge all other possible ones.



Now, I believe that it is the time to leave aside the question of determinism, since determinism is not a requirement for modern explanatory mechanisms on any level. Besides, the problem manifests itself well enough even without the idea of determinism: within a physical explanatory mechanism it is extremely difficult to find any place for concepts usually used in describing and explaining human actions. All that has been said in this section leads us to the question which I think is most important for the topic of my study. Plainly, is it possible that there could be more than one complete causal explanation for one and the same phenomenon? Inspired by Malcolm's paper, Jaegwon Kim (1989) takes this wider implication under scrutiny, although he also naturally associates the problem strongly with the tension between physical and mental explanations for human action. In his paper "Mechanism, Purpose, and Explanatory Exclusion", he demands explanatory realism as he is ready to defend Malcolm's idea at least to some extent. He (ibid. 95) formulates the following epistemic principle for two independent explanations aimed to explain one and the same phenomenon: "*No one may accept both explanations unless one has an appropriate account of how they are related to each other*". Of course, this makes sense. There is something fundamentally odd in a situation in which one event could have many perfectly sufficient explanations which would not have any relation to one another. However, Kim is not as ready as Malcolm (1968, 8.) to deny the possibility of some kind of relation between two or more explanations in the case of human action. On the contrary, he (1989, 269–271) sifts through in detail all the possible relations which two explanations might share.<sup>43</sup> In any case, my aim here is not to proceed any further with this discussion before some more primary issues are considered. First, I would once more like to ask what the actual *explanandum* of those explanatory mechanisms relying only on physical entities is. Second, I would like to raise a question about the nature of acting and agency as well as their relation to the physical realm. In the next section, I will consider these questions in the light of the inspections carried out in the first part of my study as I believe that the problems posed may be solved only with careful contemplation of the nature of causality and action.

---

<sup>43</sup> Evidently, only a few possibilities of those which Kim goes through are worthy alternatives to describe the rivalry between mental and physical explanations for human action.

### 3.2.3 Actions and Agents in the Physical Domain

In Section 3.2.1, I introduced the foundation for modern science laid by Descartes at the beginning of the modern era. Although the strict dualism served perfectly as a theoretical background for modern science, there were still questions not properly solved. The main challenge has been to explain what kind of relation there could be between the two types of substances which could not share any of the same properties. Nevertheless, the idea of enclosed physical substance with its own laws is still quite popular. After all, within the field of physics there is constant endeavor to grasp some ultimately small pieces of matter which could be regarded as inorganic building blocks of the whole physical realm. Now, since it comes across as more or less sensible to consider the physical world as a closed system of matter, it is also plausible to presume that all that happens in it has a physical story behind it. In the first chapter of my study, I argued that there is no causality in the world-in-itself as it is only a necessary device for our understanding of the world. However, even without talk of causality, it seems reasonable to regard the physical world as consisting of an everlasting stream of ‘something going on’. After all, one of Galileo’s (see page 112) revolutionary thoughts was to claim that constant motion is the most natural state in the physical world. With all that has been said, I think it is quite plausible to presume that all events in the physical world have their causes within the same realm. Besides, I could not even imagine what it would be for a physical event not to have a physical cause. In other words, even though my understanding of physics is not advanced enough to prove my claim on any level, the other way around would simply be something which I could not properly understand. Therefore, I do not consider it essential whether any good physical causal explanations have already been given to the actions of human beings. On the contrary, what I believe is more important in the light of my study is that it is highly natural to assume that they are there to be found. Since the most challenging problem here has been to understand the relation between mental and physical substances, I am greatly tempted to endorse the idea according to which the physical substance is likely to manage perfectly by itself.

Now, in light of all that has been said in this section, I cannot see any problem in postulating imaginary physical explanations for the things that I do. Since it is impossible to carry out any empirical experiments on the issue at least for the moment, the second best way

to analyze the consequences of an explanation of this kind is most likely to imagine one and then to contemplate the issue. Let us take an example under scrutiny, say, my calling an unknown man on the street 'a scoundrel' without any apparent reason. Presuming now that there is a physical chain of events leading to this incident, we may call it 'causal explanation P'. Whether P includes facts about neurophysiology or something else is not relevant to our analysis. What matters instead is that it is an explanation which refers only to pure physical entities and events. In the preceding section, I superficially demonstrated the ongoing discussion of the possibility of two or more independent explanations for one and the same event. In any case, to my mind the possibility turns to necessity if we let the scene proceed a little further. Let us continue by imagining that my innocent victim turns to me surprised and insulted, demanding an apology and, most of all, an explanation. Now, if I try to satisfy this man's thirst for knowledge by citing explanation P, obviously that will not do the trick as it is more likely that he believes I am capping my unfair insult by making fun of him. Clearly, the content of P does not matter here. Whether I am talking about brain chemistry or atoms does not change the situation, since the explanation the angry man is crying out for is something fundamentally different than a causal explanation about physical states. In my view, it seems to be beyond dispute that physical explanatory mechanisms cannot fulfill the need for other kinds of explanations. This leaves us with a few options. First of all, we must decide whether we take the alternative explanations sharing the same *explanandum*. Only in the case of choosing the affirmative answer, need we make further investigations about whether there must be some kind of relation between the alternative explanations. To my mind, it is not necessary as I will defend the negative answer.

As I noted above, one possible response to the problem of two or more independent explanations for human action is to deny that they are actually explaining *the same thing*. This is evidently the most straightforward solution to the problem, and naturally it has been presented in various forms in many instances. For example, this kind of reaction to the problem was popular among the representatives of the "Wittgensteinian" approach (see Section 2.2.2). A. I. Melden (1961, 118) distills the line of thought aptly when saying:

Where we are concerned with causal explanations, with events of which the happenings in question are effects in accordance with some law of causality, to that extent we are not concerned with human actions at all but, at best, with bodily movements or happenings; and where we are concerned with explanations of human action, there causal factors and causal laws in

the sense in which, for example, these terms are employed in biological sciences are wholly irrelevant to the understanding we seek. The reason is simply, namely, the radically different logical characteristics of the two bodies of discourse we employ in these distinct cases—the different concepts which are applicable to these different orders of inquiry.<sup>44</sup>

The whole idea emerges quite lucidly from the above quotation. Melden is willing to distinguish the language of actions from the language of bodily movements or happenings, and thus to completely separate two “bodies of discourse” or “language games”. So, in our preceding example, explanation P would merely explain why my body has moved whereas by saying, for example, “I know that you are the one who dishonored my sister”, I would explain why I did what I did. Another “Wittgensteinian”, G. H. von Wright (1971, 9.) treats the question in a similar way. He distinguishes basic actions from actions with some further meaning, and hence considers there being two different *explananda*. In this case, there would be a basic action of uttering some words and an intentional action of offending a fellow citizen.

Norman Malcolm (1968, 6.–7.) takes these kinds of arguments seriously, even if he is not willing to see their whole worth. Instead, he poses the following question: “It is true that the two kinds of explanation employ different concepts and, in a sense, explain different things: but are they really independent of one another?” He presents a case where a man climbs a ladder in order to retrieve his hat from a roof and where the very action can also be given a sufficient neurological explanation. Justly Malcolm points out that, although there definitely are two separate explanations referring to totally different kinds of entities and emphasizing different aspects in their *explananda*, nevertheless they both serve as an explanation for the fact that man has ascended onto the roof. Now, if the presumed neurophysiological explanation is by itself sufficient to explain the fact, what possible role could intentions, beliefs and desires play? This is undeniably a very good question. After all, the man is on the roof and it seems possible to explain it in many different ways. If one of the ways is complete enough as it is, the role of all the other explanations is rightly questioned.

---

<sup>44</sup> It must be remembered that Melden denies that action explanations are causal, and therefore he associates causality only with all the improper explanations for action (see page 66). However, in the second chapter of my work, I tried to argue for the opposite thought. I believe that reasons for action can be understood as causal as well as non-causal depending on the perspective we choose.

I believe that now is the time to return to the example we contemplated before. As we have seen, there are at least two *explananda* which may be distinguished. First of all, there is the physical movement of making sound, and then there is the action of uttering an insult. Now, Malcolm would admit that they are two different things on some level, but he would nevertheless deny that they could be totally separated. He would likely think that, in the end, they are solely alternative descriptions for the same phenomenon, and thus an explanation for one is an explanation for another as well. Either way, this is something with which I will disagree, even though the basic idea is quite popular. In fact, to my mind, the key to the solution lies right here. In the light of what I have said in the first chapter of my study, it should be easy to clarify my thoughts. In Section 1.2, I presented my view about causality and, before anything, about how we experience the world and talk about it. I argued that all that we can ever grasp are properties and that we can refer to the phenomena behind the properties only by highlighting those very properties. Plainly, that is to say that the phenomena-in-themselves are always a way beyond our discourse. In that same section, I also included events and actions under the category of properties. It can be said that behind events there is most likely a phenomenon-in-itself going on, nonetheless we can talk about it only through its properties. Therefore, the phenomenon-in-itself can never be the *explanandum* of our explanations as we can only explain the properties it possesses. Similarly, causality does not exist anywhere behind the properties as causality is in itself a relational property which cannot be attached to the phenomena-in-themselves but only to their properties. Now, if we keep all this in mind it is easy to see that the problem lies only in our confusion as it is hard to accept our inability to grasp the world as it is behind all the aspects from which we must choose.

So, once more, let us get back to our example. I believe that we can all agree that behind both events there is one single phenomenon to which these events can be attached. However, what I wish to emphasize is that this phenomenon is something which we cannot talk about in itself, and thus it is something which cannot be explained. There is nothing to be said about this phenomenon apart from talk about its properties. Therefore, the properties must be those we explain. Now, the two properties under consideration, namely ‘a purely physical action of making noise’ and ‘an action of insulting a fellow citizen’ do refer to this one and the same phenomenon, but *they are not the same property*. Thus, it is justified to claim that they are not the same event and that they may require dissimilar explanations. In the end, it is quite surprising that in the case of events, the idea of diverse explanations for the

properties of one and the same phenomenon appears so extraordinary, although it is almost a self-evident fact in the case of some less debatable properties. For example, I doubt whether there is anyone who would disagree about the fact that in, say, a book, the properties of, say, blackness and thickness demand different explanations in most instances. Although it is one and the same book which is black and thick, the two properties are nonetheless not the same, and hence should be explained differently. Now, if we are to consider the insulting scene as a whole, it is evident that the two events mentioned are not the only ones it may be said to possess. For instance, further events like ‘a man behaving rudely’ and ‘a man standing up for his sister’s honor’ can be found. When it comes to explaining, many actions are so close to one another that they can be explained together. For example, it is likely that the actions of ‘insulting a man’ and “calling a man ‘a scoundrel’” could be explained with the same *explanans*. However, clearly this is not the case with the physical event of ‘making noise’ and the more meaningful action of ‘insulting a man’. They are both events of totally different kinds, and hence they require different kinds of explanations. The first one could be explained with physical facts whereas the latter one demands reasons referring to mental or propositional entities. All in all, to me it seems evident that there could be no physical mechanism serving as an explanation for the action of ‘insulting a man’. Undoubtedly, there is a tendency to place the physical domain above all else, and thus to consider the explanations of the kind to be capable of explaining everything else by the same token. Of course, this is somewhat understandable as the whole world is usually regarded as structured on the foundation of the physical matter. Following this line of thought, one could conclude that when we reveal the laws governing matter, we reveal everything else at the same time. However, even if this kind of thinking is understandable, it is rather naïve all the same. As there are an endless amount of further levels and meanings to be found for physical occurrences, there are naturally various perspectives which we may take towards them. These perspectives cry out for different concepts and causal structures, and therefore the explanations of one may not mean anything in another. As was stressed in Section 1.2.4, an explanation is always an answer to a question, and thus different questions demand different kinds of explanations. Furthermore, when it comes to the ranking of perspectives, in the end, the question is far from simple. After all, as we all are individuals living in a more or less civilized society with our fellow citizens, it is a quite weird conclusion to draw that the most important aspect is that of physical facts. It should be asked what the statements and explanations of physics could possibly mean to “the man on the street”, especially, if he is

called 'a scoundrel' all of a sudden. After all, perspectives of actions as voluntary and intentional are those with which we deal mostly in our everyday life.

There is one more issue we should comment on in this section. As the headline suggests, the question about a place for actions and agents in the physical realm is still left to contemplate. In the second chapter of my study, I introduced different ways of understanding actions. In Section 2.1, I noted that there is nothing wrong in speaking about actions in the case of subjects which are mere physical objects. Of course, the issue is clearer when talking about, say, robots to which we may postulate mental states, but neither did I see anything wrong in agents to which it is hard to postulate any such states. In my view, it is acceptable to say that a train or an internal organ is doing something. After all, there is an evident disanalogy between the cases of a healthy heart pumping blood and of an artificial operation where the heart is made to pump blood because of an emergency. However, in some cases, the line is vaguer. Let us take the following example. When throwing something up in the air, we do not regard that the thing is doing anything but merely going through the event of being thrown. Either way, if we put, say, a boomerang in the place of a thing being thrown, the case changes a bit. Now, while the boomerang is flying in the original direction, it does not do anything, but the situation alters when the boomerang starts to make its turn. This is the case at least when we throw it for the first times. In the beginning, the turn may surprise us, and thus we may consider it as something which is not in our power but which is made by this mysterious item. In any case, after practice the situations changes. We learn how the boomerang behaves in the air, and similarly we get the feeling that we control the whole process. With this example, I have been trying to demonstrate how the actions as well as all the other properties are not something existing regardless of us, as they are more like the products of our experiencing the world. Simply said, when we feel like something is doing something by itself, we may say it is acting, but when we see it as solely going through a process, the action turns to a mere event. Therefore, the justification of action talk depends only on our understanding of the world. By way of another example, we may imagine a situation in which I see an airplane making a figure eight in the air. For me it is perfectly natural to state that the subject of action is the airplane, whereas people better informed in the line of business see the plane only as an instrument of the pilot. Similarly, people more aware of the laws of physics may be able to see the turn of the boomerang from the beginning as a happening governed by those laws, whereas for a surprised layman it seems that the boomerang is steering itself. All in all, for justified action talk it is impossible to give any

general rules. In the end, it seems only to follow our interpretation of the world and, where we see ‘somebody or something doing something’, we are allowed to say that there is an action going on. There are no restrictions considering the subject. Anyway, I am totally aware that this kind of view of actions diverges greatly from more orthodox ones. One good example of alternative thinking is the quotation from Thomas Nagel on page 120. There he concludes that when there are only physical or mental happenings at stake, there seems to be no place for real actions and agency. However, even if it seems that nobody is really doing anything when there are only physical or mental happenings to be found, at least to me it comes across as totally impossible to exclude these from the category of actions. Simply, it would contradict the common discourse too heavily to claim that, say, irrational animals are not actually capable of doing anything, since their actions are mere consequences of mental happenings in their head. Preferably, I choose to argue that actions are there where we consider them being.

Now, presuming that we accept the line of thought I have been demonstrating, I believe that we are justified to talk about actions even in contexts which are all about pure physicality. As is written even in my somewhat amusing definition according to which in action ‘somebody or something does something’, actions always require a subject which is the performer of the action. In the case of actions understood only in physical terms, it is natural to take the whole physical body as the subject of the action. When, for example, we take the stance of physics towards my actions, it is natural to see my body as the subject of the action. However, even if we may refer to these physical bodies using names, the name nonetheless does not include the self as it does when it is used in the case of voluntary or intentional actions. Even though there are views which associated the self of a person with his body or some part of it, they undoubtedly belong to the minority.<sup>45</sup> The same goes for many other concepts familiar within the action talk. For instance, the concept of will does not seem to find any place in the physical domain. Simply, what on earth could be the physical counterpart to will? Where could one find will in, say, the body of a man, let alone in that of, say, a moving train? Of course, one may claim that what the subject does in the action is the manifestation of its will, but to my mind this kind of talk goes too far. Even if we admit that there are acting subjects in the physical realm, we do not need to accept that there are agents

---

<sup>45</sup> To give an example, Lynne Rudder Baker (2000) defends a constitution view of self. Her (ibid. 91) account stands for the idea according to which a person is constituted by its body, but denies that they are identical, as bodies can clearly exist without a personhood. On the other hand, for instance, Roderick M. Chisholm (1989, 125) presents an imaginative view according to which a personhood dwells in Leibnizian monads, namely in little physical particles which govern the rest of the body.



with wills making decisions. As has been said, if we really contemplate the action only within the physical story, there is no counterpart which could be linked with the concept of will without ending up with a fully absurd picture of agency.

In the light of all that has been said in this section, to me at least it seems obvious that, even with all the information the mechanisms of natural sciences may provide, they cannot say anything about agency in its full sense. As there are no wills to be found in the physical domain, it is undeniable that only by referring to physical facts can we conclude neither positive nor negative facts about them or their freedom. Whether the world is, in the end, deterministic or indeterministic does not have any influence on the issue. If there are no wills in the whole context, nothing we say within it can provide any information about them. In my view, this does not mean, however, that we cannot talk about actions in the very context. Anyway, I believe that now is the time to move forward and take the mechanisms of the other kind under scrutiny.

### 3.3 The Mechanisms of Social Sciences

#### 3.3.1 The Role of Social Sciences

Even to a layman the general purpose of natural sciences comes across as fairly plain although the actual practice may go over his head. The target of natural sciences seems to be, of course, the world about us as natural scientists try to reveal the rules according to which it behaves. With all the information gathered, it should be possible to explain singular and more general occurrences as well as to predict those yet to come. Naturally, there can be many gaps in this brief definition, but all in all at least the basic idea cannot be very far from truth and, moreover, it is hard to believe that there could be much dispute about it. However, surprising complications arise even in the very basic matters when we move from natural to social sciences. Although the history of social sciences as such is truly shorter than that of natural sciences, there has been a remarkably large amount of controversy even about the very role of social sciences itself during this relatively short period. Here, before proceeding to further

questions such as about the explanatory mechanisms in social sciences, I will try to illuminate those more fundamental issues.

I suppose that there must always be two extremes in every dispute, and so is the case in that which we are contemplating. I will call them *sociological positivism* and *interpretive sociology*, and next I will introduce them one at a time. I believe that very seldom, if ever, can the truth be found on an extreme side of any argument, and so it seems to be in this dispute as well. On the contrary, philosophers and researchers are usually not so lost in their thinking that one could find no sense in their thoughts. Moving to one extreme naturally means the denial of the opposite, and therefore I have learned to be cautious about getting too close to far edges. In this section, I will argue that, in this case too, the most sensible ideas are found somewhere in the middle, and hence I will try to dispel the juxtaposition on our way to compromise.

Naturally, sociological positivism belongs to the continuum of the tradition of logical positivism or logical empiricism which, according to Richard Creath (2011, 1), is more like a movement than a set of doctrines.<sup>46</sup> The tradition is usually associated with the Vienna Circle and the Berlin Society for Empirical Philosophy which were both influential in Europe in the 1920s and 1930s. After the unfortunate incidents in the region during the following decades, the movement found its new headquarters on the other side of the Atlantic Ocean. Anyway, even though the terms include extremely wide-ranging views, I am sure that it is not too daring to make some simplifications about the most general ideas. First, empiricism in its strictest form is the feature which comes first to mind when talking about logical positivism (interestingly enough, Creath [ibid.] denies that even empiricism is an idea which all logical positivists could be considered as sharing). Undoubtedly, at least most logical positivists subscribed to Ludwig Wittgenstein's strict verificationism (ibid. 14). This line of thought suggests that all meaningful claims must be derived from empirical observations excluding, of course, analytical truths. Thus, it renders traditional metaphysics, for example, nonsense. In all its strictness, this kind of thinking is hard to fit into the everyday context within which people talk rather incautiously but, at least for the purposes of sciences, it offers an attractive tool. Second, there is an idea of unity of sciences which, in the light of our study, is the most important feature in logical positivism. According to Creath (2011, 21), Otto Neurath (1882–

---

<sup>46</sup> Although the two are usually considered synonyms, their identity may still arouse minor doubt. However, even if some little differences in nuances can be found, I believe that there are no serious risks in associating the two the way it is normally done. As Creath (2011, 2) himself states, “[t]he most that one can say is that if a distinction is to be drawn, logical empiricism is the wider term”.

1945), more than anyone, was the one who championed the thought. His motivation stemmed from two main worries. Firstly, he was concerned that, if social sciences did not commit themselves to similar methodological rules as natural ones, they would be in danger of heading for results which relied on untenable metaphysical factors. Secondly, Neurath thought that, only through a proper unification of the methodology and vocabulary of the various fields of science, could a true collaboration between them be made possible. In the end, the motivation of Neurath and his fellow positivists is rather easy to understand. After all, even the term ‘science’ cries out for some basic principles for defining its boundaries. The line between science and non-science should be as clear as possible in order to make the distinctions in practice. One good example of positivist thinking is the DN model (see Section 1.1) which Carl G. Hempel saw as an ideal explanatory device for all sciences.

In the light of what has been said about logical positivists, it is relatively easy to figure out the ambitions of the representatives of positivism in sociology. For them, social studies should not diverge from natural sciences, at least in the methodological aspect. The functionalist paradigm to social sciences is rooted in the very tradition of sociological positivism. Gibson Burrell and Gareth Morgan (1979, 26) describe it the following way: “The functionalist approach to social science tends to assume that the social world is composed of relatively concrete empirical artefacts and relationships which can be identified, studied and measured through approaches derived from the natural sciences”. They (ibid.) add that “[t]he use of mechanical and biological analogies as a means of modeling and understanding the social world is particularly favoured in many functionalist theories”. As one of our main interests in this study is the idea of social mechanism, the latter quotation must have piqued our interest. The most used example of functionalist thinking in social sciences is Emile Durkheim (1858–1917) as he stands boldly behind the positivist picture in its strictest form. For him, the object of social studies does not diverge one bit from that of natural sciences in the sense that they both study their targets objectively from outside and only by means of what is observed from that viewpoint. As Anthony Giddens (1993 [1976], 138) puts it, “[a]ccording to Durkheim, the object of sociology is to construct theories about human conduct inductively on the basis of prior observations about that conduct: these observations, which are made about externally ‘visible’ characteristics of behaviour, are necessarily ‘pre-theoretical’, since it is out of them that theories are born”. Obviously, according to this kind of thinking, the inner lives of the agents as well as the cultural meanings do not carry much weight. As a natural scientist observes his target from outside, founding his theories only on

what he sees, Durkheim is willing to step out of the social realm and take a similar position as a neutral bystander aiming to find causal connections and regularities in the social world.

Even if the unity of science makes sense on one level, after further consideration the issue begins to appear far more complicated than that. After all, the social world differs greatly from the natural one in many aspects. The antithesis for sociological positivism is interpretative sociology which wishes to emphasize those differences and thus to separate social studies from natural ones once and for all. When it comes to the differences mentioned, a few of them should be contemplated next. If we consider nature around us, the things going on there cannot usually be interpreted in too many ways. For example, a falling tree is merely a falling tree, and therefore it is rather hard to find any further meanings in the process. Similarly, the explanation for the falling should be quite easy to find and validate. However, even if one could choose a totally different aspect to the thing, it is still a person outside the event naming it. For instance, we may consider the falling of a tree as a kind of self-expression of God who naturally has His own reasons for the act, but still it would be we who determined what was going on in the picture. Naturally, this is so because asking a tree or God about their deeds is not an option for a healthy young man. Anyway, the issue is quite the opposite in social sciences. As one simple hand gesture may have various different meanings depending on the context and intentions of the agent, it is usually of the utmost difficulty for an alien observer to get a decent grasp of the scene. In the end, there is a dramatic difference between, say, a fanatic National Socialist and a comedian on a stage making exactly the same hand gesture. Therefore, in order to really understand what is going on, a social scientist must figuratively find his way inside the context and agent's head since, to outsiders, the gestures look the same. Furthermore, as Giddens states in his *New Rules of Sociological Method* (168), “[s]ociology is not concerned with a ‘pre-given’ universe of objects, but with one which is constituted or produced by the active doings of subjects”. Later he (ibid.) continues that “[t]he production and reproduction of society thus has to be treated as skilled performance on the part of its members, not as merely a mechanical series of processes”. At this point, we arrive at one of the most fascinating features in social sciences, namely that which Giddens (ibid. 170) calls *double hermeneutic*. As a social scientist as well must, at least on some level, belong to the social world in order to understand it, the information does not go one-way only. After all, since social scientists themselves as well as their studies are integral parts of the social world, it often is that they have influence on the behavior of their targets.

In Section 2.2.2, I wrote some words about the influence that Wittgenstein had upon the philosophy of action. Plainly, the idea of A. I. Melden, Georg Henrik von Wright and their companions was to stress the understanding of meanings in actions in the place of causal connections of any kind. They saw actions and their reasons as belonging to the same contextual web of logical relations and, as they believed that there cannot be causality between logically related entities, they went as far as denying the possibility of causality in action explanations. Instead of explaining actions causally, the true aim of action explanations should be the understanding of their meanings. Perhaps the most important representative of interpretative sociology was Peter Winch (1926–1997) as he defended it in its strictest sense. However, as Giddens (1993 [1976], 56) points out, the significance of Winch's work did not derive from its originality as similar thoughts had been expressed before by his many fellow "Wittgensteinians", but more likely from the fact that he was the first one who brought those ideas from the philosophy of action to social sciences in the fullest sense. To Winch's (1958) thinking, the whole social world is a contextual network where all the meanings are conceptually related. Whereas Durkheim, for example, was willing to divide the whole into parts which could be investigated separately and through the causal relations they share, in Winch's view the logically connected bundle of meanings must not be torn apart. He went as far as claiming that causal connections have importance only in natural sciences while in social sciences they should not be the object of interest. The concept of 'social rule' plays an important part in Winch's account as he believes that, instead of causes, the rules according to which people act are what we must figure out in our process of understanding the social world. Interpretative sociology clearly relies on hermeneutic tradition in its holistic thinking (Burrell & Morgan 1979, 235–238). Winch and others emphasize the undeniable relation between the whole and its parts as they embrace the impossibility of understanding one without another. Therefore, they believe that the whole process of understanding the social realm is more about skillful interpretation than straightforward learning through observation.

Obviously, an interpretative attitude towards social studies has its point. Sociological positivism does not appear sufficiently workable as it seems incapable of grasping all the aspects of the extremely rich social world. Merely by observing from outside, it may be impossible to fully understand what is going on in some unfamiliar social processes. This can be easily demonstrated with a plain example. Let us imagine an aboriginal tribe whose customs and behavior we are trying to understand. We observe that, every time the chief of

the tribe beats a drum, his subjects begin to dance wildly wherever they are at the moment. This weird procedure happens many times during a day. Clearly, even after only a short period of observation we are able to unravel the cause and effect in the scene. We are able to predict that, when the chief beats his drum the next time, there will be a fast-paced whirl of motion. However, even if we learn to explain the dancing by mentioning its cause, namely the beating of the drum, this kind of explanation is as empty as can be. If we compare this kind of knowledge with the true understanding of the meaning and history of the ritual, obviously it is not of much worth. In this scenario, knowledge of the mere causal connection does not have much purpose. As is stressed, within the social world even the most basic acts can be surprisingly rich in their meanings as the context may serve as a colorful playground for highly complicated self-expression. Thus, there may be a huge amount of action-properties to be found in seemingly simple body movement. However, the next question should be about whether these facts undermine the importance of causal connection as was suggested by Winch. Let us further imagine that, after infiltrating the tribe, we find out that the beating of the drum is a sign of a possible threat which the chief intuitively senses. Within the tribe there is a strong belief that dancing is the best way to keep trouble away from the village. Now, only after understanding the ritual properly through its meaning are we in a position to say that we truly have substantive information about the process. Nevertheless, does it mean that the causal connection between the drum and dancing is of no importance at all? I would not go that far either.

In the preceding example, the causal relation between the drum beating and dancing was clearly not very valuable information given that there was no information about the meaning of the whole procedure. Nonetheless, the fact does not make the causal connection between other properties of the same phenomenon dispensable. For example, the causal connection between the collective fear and dancing is a highly interesting and significant relation. Furthermore, would it even be possible to understand the whole picture without the idea of causality playing the major role? As stressed in Section 1.2, causality is one of the most basic structures according to which we assess our surroundings. Undoubtedly, the fear in this example is a truly informative causal cause of the act of dancing. All in all, even if causality is not all that we need, it is nevertheless something which we do need.

What I have wanted to emphasize here is that, even though the mere re-describing of phenomena may serve as the main course of action in many instances of sociological research, the fact does not render the idea of causality and causal explanation totally

meaningless in social sciences. On the contrary, as the idea of causal relation plays a dominant role in our understanding of the world, it does so in the social realm as well. I believe that Marja-Liisa Kakkuri-Knuuttila (2006) is right when arguing that the shunning of causality in social sciences has its origins at the pure misconception about the nature of the very thing. For example, the thought according to which there must be lawfulness where there is a causal connection has made many social scientists avoid talk of causality. After all, it is a well-known fact that it is extremely hard to define any law-like relations in social sciences. However, as I argued in Section 1.2.1, the idea of lawfulness should be separated from talk of causality, since the idea of causality in its most ordinary form does not embody the very idea. For instance, Winch (1958, 124–125) has an example of a sergeant yelling “Eyes right!” and his men doing as his says. Now, as Winch states, there evidently is a conceptual connection between the acts of commanding and obeying as it would be impossible to understand one without another. For Winch and his fellow “Wittgensteinians”, this is enough to deny the causal connection as they stubbornly held the rather peculiar opinion that there cannot be a causal relation between entities with a conceptual nexus (see page 66). Moreover, it is obviously impossible to formulate any law between the two as there have fortunately always been people who do not obey the commands of those above. In any case, despite these factors, to my mind it is incredibly odd to seriously claim that the command is not the cause of the actions of the sergeant’s men or that the possible causal relation does not carry any weight. Only after the type of clarification of the notion of causality done in Section 2.1 are we able to clear up these misapprehensions and find an appropriate place for causality also within the social world. As Kakkuri-Knuuttila (2006) states, both causal explaining and interpretive understanding are needed in a social study, and therefore they should not be taken as options of which only one must be chosen.

### 3.3.2 About Social Mechanisms

Now, if we proceed to talk about social mechanisms, it should be quite evident that their use is based on the positivist tradition in social sciences. As social mechanisms are necessarily causal chains, they provide explanations of a similar type as those of natural sciences.

Therefore, they can be considered an overarching device by which various sciences can find a common language. However, as I tried to argue in the previous section, there is some definite uniqueness in social sciences which requires keen attention. Firstly, the social world is so rich in its overlapping meanings that one must be careful about which features are related with causal relations. Above I presented an example of the aboriginal tribe to illuminate the fact that it is remarkably important to pay attention to which properties of the same phenomenon we combine with causal relations. In order for an explanatory mechanism to be genuinely informative, it must pick out as descriptive properties as possible. Secondly, an acting social individual is as complicated an entity as can be in the light of the various ways we can understand it. As we all belong to the social realm ourselves, we must be truly cautious about not confusing the way we see ourselves with the way we must see the individuals acting as parts of mechanisms. In this section, I take a look at the use of social mechanisms in general.

Naturally, the actual use of mechanisms has been more popular in some fields of social sciences than others. As Peter Hedström and Richard Swedberg (1998, 3–4) note, mechanisms have traditionally been quite frequently used in economics and psychology, whereas in sociology and, most of all, in history their use has been rarer. For example, in cognitive psychology the processes studied are often seen as causal mechanisms. Similarly, the whole discipline of economics is founded on one “master mechanism”, namely on the market which can be traced back as far as to the 18th century. However, although history and sociology have been more about understanding and describing institutions, little by little mechanisms have found their place in these fields too. Mostly, in sociology there has been rapidly growing enthusiasm for mechanisms in recent decades. The term *analytical sociology* has been established to cover the tendency to find explanatory causal mechanisms in the social world. As Hedström and Petri Ylikoski (2010, 58) state, “[t]he identity of this movement is not based on a common object of study, a shared historical tradition in sociological theory, or the use of a specific research method or type of empirical data, but on some general ideas about what good social science is all about”. By these ideas, they mean “the importance of the intellectual virtues of precision and clarity”. Consequently, the movement of analytical sociology is willing to bring true explanations to the social realm in the place of mere descriptions, and by this they try to close the gap between natural and social sciences.

Even if Hedström and Swedberg (*ibid.* 5) trace the concept of social mechanism back to the beginning of the 20th century, its explicit use commenced not before the middle of it.



During that time Robert K. Merton, who is often regarded as the father of the whole trend, started to use the term frequently in his writings. He associated social mechanisms with his well-known idea of middle-range theorizing in social sciences. He (1949, 5) describes theories of the middle range in the following way: “[T]heories intermediate to the minor working hypothesis evolved in abundance during the day-by-day routines of research, and the all-inclusive speculations comprising a master conceptual scheme from which it is hoped to derive a very large number of empirically observed uniformities of social behavior”. Simply said, Merton wisely believes that social theories should avoid both extremes in the scope of their study. As all too wide-ranging theories built in order to explain a huge mass of social phenomena are dubious, it is also rather hard to see any point in theories which can be applied only to some very special circumstances. Merton saw social mechanisms representing middle-range thinking perfectly. After all, they are generalizations of some repeatable procedures, but they aim to explain phenomena only of one type instead of being some sorts of all-encompassing explanatory schemes. Merton’s own account of *self-fulfilling prophecy* represents the idea of middle-range theory quite well as it is also one of the prototype examples of social mechanisms in general.

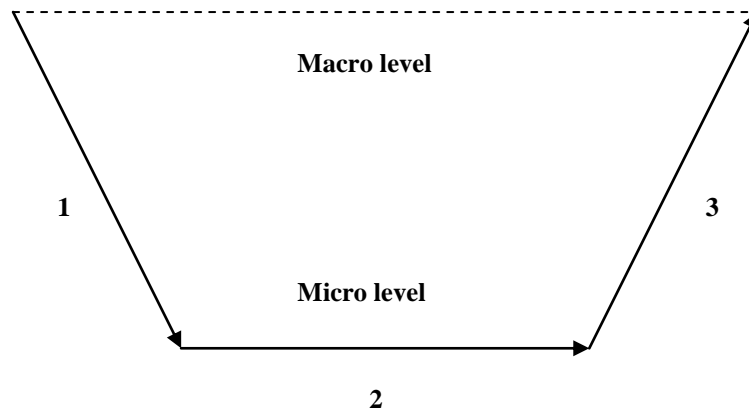
In his *Social Theory and Social Structure* (1936), Merton quotes W. I. Thomas as he is willing to name one of the basic theorems of social sciences: “If men define situations as real, they are real in their consequences”. This ‘Thomas Theorem’ is the core of self-fulfilling prophecies. What the theorem means is that a false belief may have just as powerful consequences as a true one, since what counts in the light of the effects of beliefs is not the truth value but the possession of the belief overall. However, the mere existence of false beliefs does not make a self-fulfilling prophecy but the change in their truth value. As Merton (ibid. 180) himself says: “The self-fulfilling prophecy is, in the beginning, a *false* definition of the situation evoking a new behavior which makes the originally false conception come *true*”. In other words, in self-fulfilling prophecies people with false beliefs act in a way that makes the beliefs become true. In these kinds of mechanisms, the special nature of social sciences becomes highlighted beautifully. As Merton (ibid.) points out, the same does not go for natural sciences. For instance, it is impossible for whatever predictions we make about the return of Halley’s Comet to have any influence on its orbit. Instead, an economic prediction, for example, may affect greatly what happens in the world of economy. From this, we have a short path to Merton’s famous bank-example whereby the idea of self-fulfilling prophecy becomes well demonstrated.

To introduce a practical instance of self-fulfilling prophecy, Merton (1949, 180–181) presents a detailed example of the bankruptcy of one most likely imaginary American bank called ‘The Last National Bank’. Before “Black Wednesday”, the bank in question is simply flourishing. Nevertheless, during this ominous day, some depositors begin to gather around the bank to withdraw their money as they have heard a rumor that the bank is on its way to bankruptcy. Now, the facts that the rumor is actually incorrect and that the bank is doing perfectly well do not change the situation, if the depositors cannot be budged with them. Unfortunately for the bank, besides taking their money away from the bank, at the same time the depositors are either knowingly or unknowingly galvanizing their fellow depositors into the same behavior. After all, every cent taken from a bank is one small step towards its doom. As more and more people begin to believe the originally false rumor, little by little it starts to come true and, before long, the whole situation ends in actual bankruptcy. One of the most interesting features in self-fulfilling prophecies is their tendency to reward those who possess false information. In Merton’s bank-example, the withdrawing depositors will be highly satisfied with their actions as they were able to foresee what would happen and thereby save their money, but the problem is that they probably will not see their own role in the process of fulfilling the prophecy. Anyway, self-fulfilling prophecy is an extremely useful mechanism as it can be found in many instances within the social world. For example, the placebo effect is one widely known mechanism of the same type. It refers to situations in which medication works only because the patient believes in it.<sup>47</sup> The mirror image of self-fulfilling prophecy is *suicidal prophecy* which refers to the reverse situation, namely to one where originally truthful predictions turn out false. A good example is one about elections presented by Eerik Lagerspetz (2006, 45). Sometimes the predictions of the dominance of a candidate may actually promote the reverse result. These kinds of predictions may cause, say, over-confidence and passivity in the winning team as well as extra effort in the other camp, and hence change the course of elections.

Hedström and Swedberg (1998, 21–23) divide social mechanisms into three categories. This pattern is often illustrated with the diagram ‘Coleman’s boat’. With the model in question, also known as the ‘macro-micro-macro model’, James S. Coleman (1986) tries to conceptualize collective social action. “The boat” can be drawn as follows:

---

<sup>47</sup> Michael Biggs (2009) has written a profound review of self-fulfilling prophecies and how they appear in various instances.



As we can see, mechanisms never work only on the macro level. On the contrary, the concept of mechanism invites us to move from the macro level to the micro level in order to find explanations (see Sections 1.3.2 and 1.3.3). What the numbers in the diagram represent are the three types of mechanisms that exist. Before I sift through them one by one, I want to stress that the types of mechanisms often work together in forming one grand mechanism. This means to say that, although the three types can be distinguished, in many cases they serve as mere parts of mechanisms which requires others of different levels. After all, mechanisms are always hierarchical, and thus inside one mechanism are more and more mechanisms of lower levels. Coleman’s boat itself illustrates a situation where the phenomena of macro level influence agents at the micro level in a way that makes changes which, again, have effect at the macro level.

Now, the three types of mechanisms are (1) *situational*, (2) *action-formation* and (3) *transformational mechanisms*. As shown in the figure, (1) situational or macro-micro mechanisms are those where a social event or situation affects individual agents in some specific way. Good examples of these are Harrison White’s (1970) “vacancy chains”. The term refers to situations in which, say, a new opening in the labour market brings about a chain of events. For instance, somebody may leave his job for the new one and thus make an opening for some other person who may do the same in order to get the job left by the first one. This chain can go on and on until all the positions are filled. (2) Action-formation or micro-micro mechanisms are directed to explain how certain combination of desires, beliefs and opportunities are likely to produce certain kind of behavior. Naturally, a great deal of psychological or social-psychological mechanisms work only on a micro level. George Ainslie’s (1992) theory of hyperbolic discounting is a perfect example of action-formation mechanism. Hyperbolic discounting refers to situations in which an agent very irrationally

prefers an inferior good over a superior one as it is obtainable in a much shorter period of time. All sorts of addicts typically place short relief above a more stable good if the first is near at hand. (3) Transformational mechanisms cover the transition from micro to macro level. In these mechanisms, a number of individuals act in a certain manner which has intended or unintended collective outcomes. For this, we already have a good example in the abovementioned bank case of self-fulfilling prophecy where the individuals acts of individual people are capable of driving a bank to bankruptcy.

If we are to consider the structure of social mechanisms, it is apparent that, like all mechanisms, they consist of separate but causally related parts. However, while the parts of mechanisms in natural sciences are usually mere physical or organic entities or events, things are more complicated in social ones. Arthur L. Stinchcombe (1991, 372) distinguishes four main types of causal units which may operate as parts in the mechanisms of social sciences. They are “(1) individuals, (2) social actors that can be treated as individuals (such as firms), (3) situations, and (4) patterns of information”. Naturally, by (1) individuals, Stinchcombe means people who, in most cases, play the main role in a mechanism. Interestingly enough, he (*ibid.*) notes that (2) social actors who have a stable preference function (for example, “maximize discounted present value of the firm”) should also be treated as people for the purposes of analysis. This is highly fascinating in the light of what was said in Section 2.2.4 about unorthodox agents, but we will get to this point later. By (3) situation, Stinchcombe (*ibid.* 372–373) refers to “a time and place in which there is continuing communication or interaction, such that the actions and communications of one person are facts to which the others respond, and such that some ‘objective’ features of the situation come to be defined in a common way by the people in the situation”. As an example, he presents a case where a criminal is sent to prison. Now, in this example it is possible to separate various causally effective situations such as his crime becoming known to the police, the indictment and the trial. What is crucial about situations is that the people inside the boundaries act in a certain special manner which differs from their normal behavior. Furthermore, if we are to study the statistics of certain types of crime, most likely we will again find all kinds of regularities in, say, the arrest warrants, arrests and sentences. According to Stinchcombe (*ibid.* 373), this is so because certain crimes types, for example, bring about certain kinds of (4) information which again have many consequences. For instance, it depends on the information how likely the criminal is to get caught and what kind of sentence he will probably face. Since in the parts of mechanisms the most important feature is their causal nature, the same goes for

information within mechanisms. Therefore, not only its content matters, but also its reliability, source and nature overall. For example, evidence in rapes is usually quite similar in many aspects, and hence with it certain facts about the arrests and sentences can be explained quite well.

I guess it is understandable that social mechanisms have usually been associated with methodological individualism on some level (for example, Hedström & Swedberg 1998, 11–13). This is the view according to which all the facts explaining phenomena on a macro level are facts about individuals on a micro level. However, this kind of thinking has recently been challenged by Hedström and Ylikoski (2010, 59–60), for instance, as they note that it might be too strong for the purposes of the mechanistic perspective. Instead, they suggest some form of structural individualism according to which “all social facts, their structure and change, are in principle explicable in terms of individuals, their properties, actions, and relations to one another”. The main difference between methodological and structural individualisms seems to be that the latter emphasizes the importance of relations and relational structures in explanations as well, like “facts about topologies of social networks; about distributions of beliefs, resources, or opportunities; and about institutional or informal rules and norms” (ibid.). The issue, however, is far from clear as Pierre Demeulanaere (2011b, 11) claims that methodological individualism has been misunderstood and that, in fact, structural individualism should be taken as inherent in it. Anyway, even with all the possible variations of social mechanisms I dare to claim that the actions of individual agents are always present at the center of social mechanisms. In the next section, I will contemplate how these agents and their actions are depicted in social mechanisms.

### 3.3.3 Actions in Social Mechanisms

Undeniably, there is always a human action at the core of social mechanism as the whole field of social sciences is all about the actions of social beings. If nobody ever did anything, there would not be a single social mechanism, and moreover, there would not be anything for social scientists to study. Clearly, this is all rather axiomatic, but it brings out the importance of actions in social mechanisms as well as in social studies overall. Even if Hedström and

Ylikoski (2010, 60) claim that the idea of mechanism does not in itself tell us how to conceptualize human action, with its explanatory purpose and the idea of causality playing the main role, there are some broad outlines we are able to draw. Nevertheless, it is true that excessively restricted views of human action should be avoided at least on a general level. For instance, sometimes *rational choice theory* is associated with social mechanisms (for example, Abbott 2007). It is not, in any case, very wise as all of us know that often human actions are systematically guided by quite irrational incentives, like emotional disorders. However, it should not be forgotten that also oversimplified assumptions may stand up for themselves in some particular mechanisms in the pragmatist sense. For instance, within the mechanisms of economics, acting units are often seen as agents aiming only to maximize their own profit. After all, in the scope of the economic world, the above presupposition is useful, and thus may be defended. All in all, even though we know that a rule cannot govern all social agents, in some instances it can form a foundation for a workable mechanism. As Stinchcombe (1991, 396) points out, sometimes even the harshest generalizations and oversimplifications are needed in forming working micro level mechanisms. After all, mechanisms are allowed to include exceptions and, besides, it is always possible to specify one more precisely if necessary. What is crucial is that the generalizations and oversimplifications are understood to hold only in a particular mechanism in question, and therefore should be applied further only with considerable caution.

In the second chapter of my study, I introduced different ways of picturing human action. In my view, actions and events are only properties, and yet by postulating further properties to them, we may get different perspectives of them. For instance, in one hand movement we may highlight very different features, and thus the movement in question can also be placed in different causal chains. I have argued that it may be possible for an action to be seen at least from three different perspectives, namely as a mere physical movement, as a voluntary action springing from the agent's mental states, and as an intentional action performed as a truly free and authentic manifestation of will. Now, if we take social mechanisms into consideration in the light of this distinction, it should be evident which perspective must be used in this particular discourse. Since we are not interested in mere physical movements in social sciences, it is clear that we need to include some further elements in the picture. Similarly, as mechanisms are most of all causal explanations, we need an account which explains as much as possible instead of relying on such cryptic concepts as metaphysical self or free will. Besides, as the freedom of will requires initiation

of a completely new causal chain (see page 95 and later Section 3.3.5), apparently intentionally acting beings are not very good candidates to operate as parts of mechanisms. Furthermore, as groups and institutions may also serve as agents in social mechanisms, we need very loose causal factors to attach to these entities. It is perfectly fine to explain, say, the activity of a bank through mental states understood as mere dispositions, whereas it would be odd to talk about its self or free will. Therefore, I dare to argue that actions in social mechanisms must always be seen as voluntary. While an agent stands and acts in the middle of a mechanism, we need to understand his actions as voluntary which means that we must picture them through the agent's mental states. Of course, mere action alone is seldom enough for a proper explanatory mechanism, so the causal history behind the mental states causing the action as well as the consequences of the action must usually be discovered in order to achieve full understanding. To mention one example of an account of actions among analytical sociologists, there is a widely accepted model for action explanations, namely 'the belief-desire-opportunity model'. Jon Elster (1989, 20), for instance, argues that actions should be viewed in the light of the agent's desires and beliefs about his opportunities. In other words, in order to understand why somebody does something, we need to know what he wants to do as well as what he thinks he is able to do. Obviously, this is just one modification of Donald Davidson's causal theory of action (see Section 2.2.2). In the end, it is natural for social mechanisms, perhaps without exception, to be founded on the Davidsonian view of action. After all, as social mechanisms are inherently causal apparatuses, their idea of action must also be inherently causal.

Now, if we return to the above-mentioned bank example of self-fulfilling prophecy (see page 138), we are able to see how the causal picture is drawn. First of all, the most important factor in all self-fulfilling prophecies is naturally the originally false belief which is about to come true. In this case, it is also the main causal factor explaining why people are running into the bank to withdraw their money. Of course, it cannot be the only factor, as there must be others in order to complete the picture. For example, the belief in itself will not be sufficient to explain the action, if there is no presupposition of a desire driving the people. Hence, the mechanism in question implicitly includes the idea of people as beings who want to retain their wealth. The idea is so self-evident that it needs not to be mentioned, but still it must be there implicitly. As has been stressed, the generality of the kinds of presuppositions do not pose a problem, since one important feature in mechanisms is their flexibility. Even if there were people without the "in-built" desire, a sufficiently large amount of people

possessing it would still be enough for the mechanism to work. For example, there might be a person with savings who hears the rumor but does not act, because money has meant nothing to him since the love of his life ran away with his best friend. Nonetheless, even if he does not care about money, there certainly are others who care enough to make the mechanism work, assuming that the rumor spreads properly. Opportunities are also central to the understanding of actions as the beliefs and desires cannot do much, if it is not possible for the person to act according to them. If we consider mental states as mere dispositions, they are something which comes into play only when a suitable opportunity arises. As the saying goes, ‘opportunity makes the thief’. On the whole, in order to understand a mechanism properly, we must understand the whole chain as well as the implicit factors not mentioned probably because of their obviousness.

As the actions of individual agents are at the core of social mechanisms and as they are explained through dispositional mental states, evidently those mental states play a quite important role in the mechanisms of social sciences. In Section 1.3.1, I considered the possibility of social laws, and concluded that the whole idea seems very unconvincing, even on a theoretical level. With what has been said about the obscure nature of mental states and their important part in the causal chains of the social realm, it is even easier to see the problem. If mental states are required to explain actions and if they are mere dispositions, no wonder that the first social law is yet to be formulated. In the end, it is rather hard to imagine any disposition that all people must necessarily possess. Besides, the sole possession is not sufficient as there may always be some stronger dispositions taking the reins when the time to act comes. For example, it is sometimes suggested that all people as well as all other beings have a strong will to survive. This thought can be traced back to Baruch Spinoza (1632–1677) (*E*, III, p6) and his idea of *conatus*, the innate inclination to continue existing. However, even if there may exist such power, there are still people who end their lives by their own hand, and so there is no strict law to be formulated even about the effects of the striving *conatus*.

Earlier, in Section 2.2.3, I noticed that, although mental states are needed to explain voluntary actions, sometimes it may be more natural and more illuminating to talk about their causes or objects instead. The sole mentioning of a desire or belief is seldom very worthwhile, if its existence is all that is given. On the contrary, it is more illuminating to refer to, say, a childhood trauma which of course implies that the trauma causes the action through some mental state. Often among explanatory mental states intentions are mentioned and, in



the end, they are truly of major importance in social mechanisms. After all, it is evidently crucial to know at what the agent is aiming and whether the results are something wanted or even expected. This fact casts doubt on my claims that voluntary actions should be distinguished from intentional ones and only the first ones belong to social mechanisms. Clearly, the question demands more clarification, so I shall devote the next section to the question of intentionality in social mechanisms.

### 3.3.4 The Two Senses of Intentionality

So, in order to be truly explanatory, a mechanism needs to be understood thoroughly. Therefore, as there are individual actions at the center of social mechanisms, these actions must be understood as well as possible. If we are to find regularities in actions and their explanations, we need to know what causes them, at what they are aimed, and what their actual consequences are. As I tried to prove earlier with my example of the aboriginal tribe (see pages 133–134), within the social realm for a mechanism to be truly explanatory, it must catch the meanings in actions instead of mere movements. Of course, there does not have to be a further meaning in every action. For instance, a person may just stroll around on a sunny day, and that is pretty much all that there is. However, social beings are capable of much more complex actions as well. Good examples of these are the so-called ‘communicative acts’ of which Anthony Giddens (1993 [1976], 94) gives the following definition: “A communicative act is one in which an actor’s purpose, or one of an actor’s purposes, is linked to the achievement of passing on information to others”. If we now hark back to A. I. Melden’s example of the signaling motorist (see page 67), it is fairly easy to see a communicative action in practice. Evidently, the driver is trying to inform his fellow motorists that he is about to make a turn in the indicated direction. As Melden himself points out, even the mere grasping of this meaning is worthwhile within the social studies. However, even if we imagine that this motorist is driving in the village of the abovementioned aboriginal tribe where the signal is not known and understood, it is still important for a social scientist to understand the purpose of the gesture, albeit that it might not have the intended

meaning any more in the very context. All in all, the importance of intentions should not be underestimated.

The most interesting social mechanisms are those formed due to the ignorance or indifference of people. This, of course, presupposes that there are also mechanisms intentionally created by individuals working as cogs and wheels in the very mechanism. The problem with this latter group of mechanisms is often that, besides being less interesting, they do not serve too well as explanations which should be the main idea of all explanatory mechanisms. Seumas Miller (2011 [2007], 17) calls these ‘joint institutional mechanisms’ and cites elections as an example. In elections, people with the right to vote usually know how the system works and what must be done in order to be a part of the mechanism. They will give their vote to the candidate they prefer and, when the time comes, the candidate or candidates with the most votes are elected. Naturally, this meets the definition of a mechanism as there is a macro-level effect brought about by micro-level activity. However, as scientific mechanisms are above all explanations for an event or fact, this voting mechanism serves as an explanation only for those who do not know the mechanism, and in a democratic society it is hard to find many people who do not know about the idea of voting. What is interesting instead and what is crying out for explanation is the results of elections, among other things. Now, if we consider, once again, the bank example of self-fulfilling prophecy, it becomes clear that what makes it fascinating is the fact that, as a side-effect of people’s actions, the whole bank goes bankrupt. There most likely are people who withdraw their money knowing that it may contribute to the failure of the bank, but it is hard to believe that the idea of bankruptcy is the driving force for any of these depositors. Quite the opposite, the poor bank goes bankrupt as a mere side-effect of the acts of individual people only trying to save their capital.

Merton already paid close attention to the issue when he wrote an article entitled “The Unanticipated Consequences of Purposive Social Action”. First, he (1936, 895) distinguishes purposive from non-purposive action. In his view, purposive action is distinct from mere behavior, and thus involves motives and a choice between alternatives. Purposive actions are usually what are at stake in social mechanisms, but it is also possible for non-purposive actions to play a part in one. Either way, Merton (ibid. 896) emphasizes that the purpose of purposive action does not have to be clear-cut and explicit in order to serve as a purpose, as there are a lot of habitual actions which have become rather nebulous and hazy over time. Anyway, naturally only purposive actions can have anticipated and unanticipated

consequences, since non-purposive ones do not include any expectations. In other words, even if non-purposive actions may obviously have consequences, they cannot be called ‘anticipated’ or ‘unanticipated’ as the agents have not paid enough attention to their doings and to the possible outcomes. Now, as purposive action by the definition has necessarily one or more purposes, this or these purposes are naturally the anticipated consequences. Moreover, as might be the case, a purposive action may have other anticipated consequences which the agent acknowledges but which are not wanted in any way. As would be a case with a depositor who knew that the bank would go bankrupt partly because of his withdrawal but who cared more about his own money than anything else, the anticipated outcome of driving the bank towards the bankruptcy is a mere acknowledged side-effect of his action. All the other consequences of purposive action besides those mentioned are unanticipated consequences.

Merton (1936. 898–902) also distinguishes different kinds of reasons for actions having unanticipated consequences.<sup>48</sup> The most important and similarly the most common reason is, of course, ignorance. In the end, the world is simply much too complicated and multifaceted for anyone to be able to know a mere fraction of the consequences his actions might have. On the contrary, usually we have to act in the light of very limited knowledge as choices have to be made quickly and without too much contemplation, not that even all the time in the world would guarantee complete awareness of possible outcomes. However, as Merton points out, there is also an economic problem in the issue. After all, time and energy are limited resources, so it must be figured out how much of those resources it is sensible to spend contemplating the possible outcomes of one’s action. Naturally, the importance of the decision in question, among other factors, should define the resources to be used. The second general cause of unanticipated consequences is error which may lie in all phases of action, namely “[...] in our appraisal of the present situation, in our inference from this to the future objective situation, in our selection of a course of action, or finally in the execution of the action chosen” (ibid. 901). For example, sometimes people rely too heavily on the assumption that things will go as they have gone before. Also, a kind of blindness for certain features in action is often the cause of error. This may vary from a simple neglect to a

---

<sup>48</sup> Interestingly enough, Peter Hedström and Lars Udehn (2009, 35–37) claim that Merton mentions four different reasons instead of three. Moreover, they combine ignorance and error to form one single reason. The other three are complexity, myopia and wishful thinking. I will not take sides in the dispute about which distinction outlines the issue better, but since we are talking about Merton’s paper here, I think we should respect his specific formulation. For instance, he (1936, 901) explicitly counts wishful thinking (or “wish-fulfilment”, as he calls it) in the category of error.

pathological denial. Finally, Merton (*ibid.*) mentions myopia as one of the most general factors causing unanticipated consequences. It refers to situations where the agent pays too much attention to the foreseen immediate consequences of his action at the expense of the consideration of further or other consequences. Actually, earlier mentioned George Ainslie's account of hyperbolic discounting is an instance of somewhat similar behavior (see pages 139–140), although in it the question is not about the missing knowledge of the consequences but about not caring about them enough.

Evidently, understanding a social mechanism properly requires understanding of the motives and intentions of the agents operating in the middle of it. Therefore, the importance of intentions must be acknowledged. While seeing intentions as mental states, they must also be taken as dispositions, namely ones directed towards the future. They are caused by desires and beliefs, and again they bring about actions. Thus, they should be taken as connecting desires and beliefs to actions in causal chains leading to voluntary actions. For instance, my hunger has given rise to the intention of going to have dinner as soon as I get this paragraph done. Therefore, when the paragraph is finished, it is justified to claim that my intention as well as my hunger is the causal force behind my going to eat. However, sometimes intentions are for that moment only, as an action may as well immediately follow an intention. For instance, my hunger may become too strong for me to wait until I have finished writing this paragraph. In that case, it could give rise to a new intention of going to eat immediately, and so my action would follow. In addition, there are situations where there is simply no time to waste. Good examples of these are the sudden moves in, say, dangerous situations in traffic. For instance, when I have to make a swerve in order to avoid colliding with a deer running across the road, the intention to make the swerve must arise in a blink of an eye, if we are to take it as bringing about the very action. However, as it really cannot require much time for a plain mental state to arise, all sudden moves can also be considered as being caused by intentions. After all, as intentions are also dispositions causing voluntary actions, the missing disposition would leave the picture incomplete. Similarly, a desire or belief can arise in a second. In the end, mental states do not need to be identified or recognized at all by the agent as voluntary acting does not require that kind of reflection. Anyway, in cases of this kind, the identification of mental states usually happens afterwards which, on the other hand, enables the postulation of sudden intentional actions.

Now, the main question to be asked is how it can be justified to claim that only voluntary actions, rather than intentional ones, should be associated with social mechanisms.

In my view, the confusion lies in the terms used. Clearly, intentionality is also crucial in voluntary action, and therefore it appears somewhat peculiar to build a contrast specifically for what I call ‘intentional action’. However, in my defense, I have only chosen already-established terms in order to avoid giving rise to more confusion by inventing and defining new ones. Furthermore, even if these terms are used in various ways in everyday living, here I am operating with them only to name some necessary categories of action. This categorizing is done only for the sake of philosophical discourse, and thus should not be seen as a redefinition of already-established terms. Even though I have been talking about adding the properties of ‘voluntary’ and ‘intentional’ to actions, as we will later see, this procedure actually happens without noticing and, above all, without using the very terms. Moreover, what attracts me about the term of ‘intentional action’ is how it underlines the most important factor in the group. This factor is, of course, subjectivity. Most importantly, to distinguish the intentional aspect of actions is not to claim that there cannot be intentions in other aspects but to stress the subjective view in the aspect. Hence, to consider an action as intentional in a sense is to emphasize the view of the agent himself about his own action, whereas to call the same basic action ‘voluntary’ is to shift the target of interest from the subjective point of view to more objective features in action. Therefore, in voluntary actions we have actions in which we are mostly interested in the objective idea of action guided by an agent’s mental states. Among these mental states there are probably states of intention, but nonetheless our interest is directed towards the existence and nature of the very states instead of the question about how the agent himself understands them. In many instances, the difference may be hard to see, but in some it can be found with ease. Furthermore, although the emphasis does not seem so crucial in many cases if we are only to contemplate intentions and intentionality, it nonetheless defines the whole picture of action and agents by the same token. This we will see in the forthcoming sections. Now, however, we should consider a few examples in order to clarify things a bit.

One good way to pick out the difference between the viewpoints is to take a glance, once more, at non-human agents. There is absolutely nothing wrong in saying that a dog or robot does something on purpose.<sup>49</sup> For example, there is an evident difference between a dog barking purposively and non-purposively. In the first case, the dog may want to inform an intruder that he is not welcome. The latter, on the other hand, may happen due to some great

---

<sup>49</sup> In everyday context, it is also right to say that a dog or robot does something intentionally, but within my study I avoid it as it would likely induce even more confusion.

excitement which makes the dog bark even without it noticing the whole thing. Hereby, it is important to distinguish barking with an intention from barking done without one, although the dog does not possess the intention as such in neither case. What I mean by possessing is that the dog does not have any clue that it is barking for the moment. On the contrary, it is more appropriate to say that the dog acts with an intentional state of mind instead of claiming that the dog has the intention of 'informing the intruder'. The difference is of huge importance. The dog cannot use language, and therefore it cannot be said to possess any intentions as propositions, nonetheless it may be in an intentional mental state which causes its actions. There is no subjective perspective for us to investigate as the dog is incapable of reflecting its own inner world. Instead, what is of interest to us is the very mental state with its causes and effects. Sometimes the same difference is hard to see in human action, as human beings often explain their own actions so that the objective and subjective aspects parallel. However, the difference nonetheless exists as, in other cases, intentions refer to mental states and in others to propositions. Next, I will try to illuminate this difference once more.

Let us imagine a case in which a man is planning to go to a theater to see *Macbeth* at nine o'clock. Now, it is definitely justified to say that he has the intention of going to see the play at the time. By this we mean that he is in a mental state of going to see the play and, with this dispositional feature, we are able to predict and explain his actions as the time runs towards the evening. However, let us further imagine that this man is my friend and that I am not aware of his plans. Now, meeting the man on the street during the lunch hour, I may enquire after his plans for the night as I may be looking for company for a game of table tennis. As the man tells me that he is going to see the play, he is telling me his intention, but the question follows whether the man is also talking about his mental state. It seems very unlikely. Instead, the man is interpreting his intentional mental state by which I mean that he answers me by revealing the proposition in the light of which he will behave when the time is right. Now, the intention here is not a mental state but a plan. All in all, even if the difference is there to see, it comes across as pretty artificial as in both situations the matter is more or less about the same thing, namely about the intention to see the play. In any case, the difference should not by any means be downplayed. As I tried to express earlier in the second chapter of my work, the whole picture of agency and acting changes according to our aspect even if the difference between the two views to intentions may not be very striking. Anyway, the difference may also become more strongly highlighted with a different kind of example.

In the abovementioned case, as in most instances, the person was perfectly aware of his plans. Somehow it would seem reasonable to claim that the idea of awareness is included in the idea of intention, but nonetheless this thought may be challenged. For example, Anthony Kenny (1975, 99–101) uses the concept of ‘unconscious intention’ which at first may sound confusing but which starts to make more sense after some thought. In his examples of unconscious intentions, Kenny refers to some famous cases from Sigmund Freud’s psychoanalytical examinations. In Kenny’s (ibid. 101) view, it is possible that a well-trained psychoanalyst can reveal intentions which the agent cannot see or which he keeps denying. However, although I appreciate the idea of unconscious intention, to my mind the examples of Freud’s cases are too perplexing to make the idea more comprehensible. On the contrary, I think we will do better with some less foreign ones. Earlier, on pages 90–91, I used Daniel C. Dennett’s example of a man taking a drink to illuminate the fact that sometimes subjective and objective reasons for actions may vary a great deal. Now, however, we should make the drinker as self-deceptive as possible in order to bring out the difference between the types of intention. Let us imagine that a man with a drinking problem is waiting for the bar down the street to open up. Even though he knows that he needs to be careful with his drinking, he does not admit the problem in all its seriousness. Hence, he believes that it is all right to go and have a drink or two to ease the pain, after which he can come back home as a true winner. However, everybody who knows this man and who has seen how his nights at the bar end is able to predict with full confidence that he will stay at the bar all night and get completely drunk. After all, the same thing has happened many times before. Now, although the concept of unconscious intention may appear somewhat self-contradictory in some aspects, in this example it nevertheless feels right to say that the man is going to get blind drunk. Naturally, there is no doubt about his intention of, say, going to the bar, but the case of the intention of drinking all night is more complicated. For instance, we cannot say that the man is acting in the light of his intention of getting drunk as he does not even consider having such an intention. Nevertheless, in some sense it is justified to state that the man is in the mental state of going to get drunk as soon as the bar opens its doors. There certainly is a disposition of aiming to get drunk according to which his behavior may be predicted and explained, thus it seems right to say that the man unconsciously has the very intention of getting drunk. All in all, the intention exists as a mental state but not as a proposition standing as a basis for authentic decisions.

In the preceding section, I argued that, instead of intentional, actions in social mechanisms should be considered as voluntary. This, however, engenders confusion since evidently intentions play a crucial part in various mechanisms. In this section, I have been trying to clarify the ambiguity around the two senses of intentionality. In my view, we may understand intentions as mental states as well as subjective insights into actions. What is at stake in social mechanisms are obviously the first ones. If we are to give a proper mechanistic explanation, incontestably the objective view of actions carries more weight than the subjective one which is always more prone to error. As has been emphasized, sometimes people are unwilling to accept their actions as well as their causes. Sometimes, on the other hand, they may not even be able to gain access to them, and therefore they simply have to justify their actions as plausibly as possible. For example, the effects of some childhood trauma may be hard to see from the agent's perspective. However, I do not claim that intentional actions should not be in the interest of science as, for instance, in some fields of psychology, the insights of agents themselves are at the center. Only while aiming to provide objective causal explanations, should we limit our line of study to actions as voluntary. Anyway, the difference is not only about choosing sides between objective and subjective (or perhaps between subjective and even more subjective) perspectives as it is much more conclusive. After all, the question is not only about how to understand intentions and reasons for actions but how the whole idea of action and agency is pictured. In the next section, I will compare these two aspects of actions, and hence try to clarify their individual nature. I will also sift through the basic concepts related to action and figure out their possible roles within both aspects.

### 3.3.5 Selves and Wills

In Section 3.2.3, I contemplated the image of agents and actions within the physical domain. I concluded that, even though it is quite natural to associate the physical body of the agent with the subject of the action, nonetheless this kind of idea does not go too well with our idea of personhood. Thus, although it is perfectly valid to talk about physical bodies performing actions, the discourse does not include any idea of self, will or freedom among others. For



instance, we may say that a heart is pumping blood, and at the same time take for granted that there is nothing more to it. The heart in question is a mere physical object without any features suggesting any kind of personhood. However, while talking about voluntary or intentional actions, the concepts like self and will are necessarily brought out. Earlier, on page 120, I cited Thomas Nagel's words which questioned the existence of agency in the world seen from objective perspective. In his view, the world from this aspect is all about causally related events, so within that world nobody ever really does anything. Moreover, Nagel denies that even bringing mental happenings to the picture could help to get rid of the problem. However, I have argued against this thought as I claimed that our everyday discourse allows talk of, say, non-rational animals performing actions, even though in this picture actions lean completely on supposed causally effective mental states. Now, the question arises about the nature of these agents performing voluntary actions. For instance, we may ask how the agents in social mechanisms should be understood. To answer this question, it is best to take a pretty long leap back to the Age of Enlightenment.

To my knowledge, one of the first Western philosophers to truly criticize the idea of substantial self was David Hume. Before that, for example, René Descartes (*CSM II*, 17) had presented his famous *cogito ergo sum* -argument which he used to prove the existence of self by associating it with thinking or the mental substance. Moreover, the idea of substantial self is the very foundation on which Descartes built his whole account of the world. Similarly, there was usually a more or less analogous idea of selfhood in the theories of most philosophers. However, as was said, Hume in all his integrity was ready to reject this notion among others. In *A Treatise of Human Nature* (1.4.6.3) he states as follows:

For my part, when I enter most intimately into what I call *myself*, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never can catch *myself* at any time without a perception, and never can observe any thing but the perception.

Later he (ibid. 1.4.6.4) goes as far as concluding that “[...] I may venture to affirm of the rest of the mankind, that they are nothing but a bundle or collection of different perceptions, which succeed each other with an inconceivable rapidity, and are in perpetual flux and movement”. What Hume does in these passages is that he denies that there is some kind of self or soul which exists as the same through time. On the contrary, he (ibid. 1.4.6.11, 1.4.6.15) compares the identity of a person with that of the ship of Theseus in the famous

paradox, saying that both consist of parts which may change over time so that, in the end, all of the parts could have been replaced by others. In conclusion, he (ibid. 1.4.6.15) calls this kind of identity “fictitious”. One way to take these words is to take them as a denial of personal identity, but the other way is to consider them as the redefinition of the notion of self. This latter way outlines the famous bundle theory of personal identity. Perhaps the best-known contemporary advocate of the very theory Derek Parfit (1987, 20) contrasts it with the so-called ‘Ego Theory’ of which Descartes’ theory is a good example, and defines it in the following way:

According to the Bundle Theory, we can’t explain either the unity of consciousness at any time, or the unity of a whole life, by referring to a person. Instead we must claim that there are long series of different mental states and events – thoughts, sensations, and the like – each series what we call one life. Each series is unified by various kinds of causal relation, such as the relations that hold between experiences and later memories of them. Each series is thus like a bundle tied with string.

In the above passages, Hume denied the idea of an immutable and constant self which could remain the same throughout its life cycle. On the contrary, he claimed that the term ‘personal identity’ refers only to a bunch of mental states which are under incessant change. Now, although Hume talks about personhood as a bundle of perceptions, similarly we may take these mental states as desires and beliefs. Obviously, this kind of idea of personhood or agency is perfect for the aspect of voluntary actions. If we are to take under scrutiny, say, the self of a man acting as part of a social mechanism, we realize that there is nothing more to this personhood than the mental states. Everything this person does is explained and understood through them, hence there is nothing left to consider if those mental states were erased. Therefore, the agent acting voluntarily is nothing more than a bundle of dispositions through which all his choices can be understood. Of course, another way to consider this agent would be to see him as the conjunctive factor of all the states. In other words, within this view the person would be the carrier of the states. However, as all his actions would nonetheless be understood through the mental states, the carrier could not play any role other than that of just carrying the very states, and therefore he would be neutral to the extent that it would become appropriate to even question the purpose of this kind of entity without any further function whatsoever. All in all, even if the bundle theory may be interpreted as a

theory for finding a place for personal identity in views in which it first seems almost impossible, the image it is providing is not too complimentary.

Now, while viewing actions as voluntary, we are evidently binding ourselves to the bundle theory of agency. This is necessary because, with all the other options, we end up with the problem Nagel presented above. In other words, either we accept the idea of self as a combination of mental states or we will lose the idea of self in this context for good which would be problematic since, as we remember, every action requires a subject. After all, besides mental states there is nothing else we could ever grasp. So, what is there to say about concepts like will and freedom within the context? In Section 2.3.4, I argued that the concept of will should be associated only with intentional actions, but undeniably there are also other ways of using the very concept. For example, Thomas Hobbes thinks that will is nothing but the strongest desire which is that which becomes an action (Chappell 1999, xiii). Clearly, I cannot dismiss this way of using the concept as improper since the term has shared various meanings through the ages, but I would emphasize that following this line we will necessarily end up with, for example, non-rational animals possessing the power of will which, on the other hand, contradicts its original meaning as a rational power before anything. By the same token, Hobbes (1962 [1914], I, XIV, p. 66) is willing to see the freedom of this will solely in the negative sense, which means that will is free as long as nothing prevents it from manifesting. Of course, this is one adequate idea of freedom, yet it does not seem to be enough. As Hobbes' sparring partner John Bramhall (1594–1663) (1999, §6) in their famous discussion about the freedom of will suggests, besides being able to fulfill one's desires, there is also a higher idea of free will, namely the power to make deliberate choices between alternatives.<sup>50</sup> Here Bramhall demands of freedom the positive aspect which stands for the idea of a subject having open alternatives in front of him. For instance, a prisoner naturally is free in one sense after being released from prison, but if he is merely left on the street without any money and options to earn some in an honest way, he still lacks freedom in another sense. However, it should not be left unnoticed that, besides stressing the importance of existing alternatives, Bramhall highlights the very power to choose between them. Hence, his idea of the freedom of will is not the mere idea of freedom in its positive sense, but something much more substantive, namely the power to make a difference. This kind of freedom is, in the end, what is usually demanded from will, and it is something which the

---

<sup>50</sup> The main parts of the discussion have been brought together by Vere Chappell in the compilation *Hobbes and Bramhall on Liberty and Necessity*.

agents performing voluntary actions seem to be missing. If all the choices an agent makes are caused by mental states which again are caused by some antecedent conditions, the causal chain goes solely through the agent without giving him any opportunity to make a difference. This is true even without any assumptions about determinism. As long as there is a clear causal chain without any reference to a positively free person, it is hard to see how the very idea of freedom could play any role in the picture. Naturally, the number of seemingly open alternatives may vary from one situation to another, but still the decisions are always “made” by causally effective mental states instead of any positively free entity. On balance, voluntarily acting agent may be free in a sense that he is not, say, tied up or without money to do things, but when it comes to his will, we must deny either the justification of the very concept or at least the possibility of its freedom.

Consequently, I argue that, even if we stubbornly want to use the concept of will in the case of voluntary actions, it would nevertheless be misleading to regard this will as genuinely free, although it undeniably may be free in the negative sense. This is so even if the temptation to think otherwise is also quite understandable. Undoubtedly, since the causal relation between mental states and actions is rather vague, the idea of freedom comes across as the handiest tool for closing the gap. As has been stated, it seems impossible to find strict laws within the social realm, and thus there seems to be an empty space fit for the idea of freedom. This, however, is not right. If we are to consider a particular example where people are supposed to act in a certain way because of some of their mental states, it is highly probable that sooner or later we will have to deal with an exception that does not behave as expected. Now, the easiest way to pass over the problem is obviously to put the blame on the person’s freedom but, as we know, it would not be acceptable especially in the case of explanatory social mechanisms. If we are also willing to truly understand why the person in question did not behave as was expected, as always we need to find the cause of his divergent behavior. In order for an explanatory mechanism to be a proper one, it cannot contain any gaps disregarded by referring to an unexplainable freedom.<sup>51</sup> On the contrary, in voluntary actions there must always be a mental cause and, if the expected cause does not have the desired effects, either the cause is missing or there is a stronger mental state overcoming it. This goes even for the most surprising voluntary actions, and thus the cause must be found in

---

<sup>51</sup> Naturally, we often make compromises as the idea of mechanism does not require that all the causes of divergent behavior should be revealed. Quite the contrary, in Sections 1.3.2 and 1.3.2, I argued that the idea does not even exclude true randomness. However, what I wish to underline here is that it is important to acknowledge that in the case of voluntary action the mental causes must always exist, and thus the ignorance of those mental causes is not the same as the freedom from them.

order to draw the complete picture. The concept of the freedom of will should not be used in these instances since it would be nothing but hiding our own ignorance about the real causal course behind actions.

To my mind, it is undeniable that, from a certain aspect, people are viewed as sole bundles of mental states, and therefore the justification of such a theory cannot be refuted. For instance, agents in social mechanisms are nothing more than combinations of various dispositional mental states by which we are able to explain and predict their conduct. However, in my view, this kind of account of personhood is by no means sufficient for all purposes. We may ask, for example, how many of us can honestly claim that he takes himself as being nothing but a bunch of perceptions or some other kinds of mental states. As Parfit (1987, 21) notes, the first bundle theorist was Buddha who thought that persons are mere combinations of other elements without being separate elements in themselves. Evidently, this type of thinking is popular among certain traditions, yet it is likely that even in their case it would be extremely hard to manage completely without an idea of another kind of self. Even if we are to consider such a notion as illusory, it is nevertheless needed in, for instance, everyday interaction. For example, even for a Buddhist it seems to be true that, while informing us about doing something, the 'I' in the sentence does not refer to a mere bundle of something but to a simpler subject of the action. This confusion is well put by common sense philosopher Thomas Reid (*EIP*, II, IV, p. 341) stating that "I am not thought, I am not action, I am not feeling; I am something that thinks, and acts, and suffers". The non-reductive 'I' seems to be the necessary subject for intentional actions. As also Parfit (*ibid.* 20) admits, this is something even a bundle theorist must accept. There is another idea of self which is necessarily needed in our use of language, and as we know, the role of language in our understanding of the world should not be downplayed.

Hume's criticism of substantial self is in some sense almost insuperable. It comes across as impossible to deny the fact that self is nowhere to be found no matter how we search for it. The criticism is so powerful that even another typical example of an ego theorist, Immanuel Kant, had to accept it. However, even though Kant was forced to admit that our senses are incapable of acquiring any evidence about substantial self, he did not want to give up on the idea (Kitcher 2006, 191). Quite the contrary, Hume's criticism only drove him to find proof of the existence of substantial self from another direction. As self could not be an object for our representations, Kant stated that it must be the factor which makes the representations possible in the first place. After all, it seems true that there must be a subject

for these representations tying them together and constructing knowledge out of them. Therefore, for Kant, substantial self was a necessary condition for our understanding of the world instead of being something about which we could have empirical knowledge. As this conclusion implies, there is not much to say about the nature of this self of which the gaining of information is made impossible. Hence, I will borrow Andrew Brook's (2008 [2004], 15) words as he states that "[...] beyond identifying the need, [Kant] had little to say about what the subject of experience might be like, so we will say no more about it". Anyway, this inaccessibility of necessary substantial self is perhaps most beautifully put by Ludwig Wittgenstein. In *Tractatus Logico-Philosophicus* (5.641) he writes that "[t]he philosophical self is not the human being, not the human body, or the human soul, with which psychology deals, but rather the metaphysical subject, the limit of the world—not a part of it". Here Wittgenstein is ready to follow Kant while claiming that, in the end, self does not belong to the world. What this means is that, when stating 'I', we do not actually refer to anything. However, by denying the reference of the term 'I', Wittgenstein is nonetheless not willing to deny the justification and meaning of the very concept. On the contrary, even though there is no self to be found in the world, there is still the phenomenon of subjectivity. Hans Sluga (1996, 329) determines Wittgenstein's position the following way: "the objective world has to be conceived as a world given to a subjectivity and it is in this that the subject makes it appearance". Evidently, this idea resembles Kant's argument. According to Wittgenstein, self can never be a target for scientific examination which, of course, can only be aimed at things belonging to the world, yet the very notion is needed to complete our picture of the world as an object and a self as a subject. Sluga (ibid. 328) reminds us of Wittgenstein's (*PI*, 304.) portrayal of the sensation of pain according to which "[i]t is not a *something*, but not a *nothing* either" after which he associates the same idea with self in Wittgenstein's thinking: "[...] the I is not a something, but not a nothing either". In the end, one could not agree more with Wittgenstein's (*NB*, 80) humble outcry: "The I, the I is what is deeply mysterious!"

Here I have been arguing in the name of Kant and Wittgenstein that substantial self is a necessary piece in our understanding of the world regardless of whether the thing itself belongs to the world. It is hard to avoid sighs of frustration when talking about an entity which both exists and does not exist, and about which one cannot say much other than that it needs to be taken as given. However, as we have no other option but to accept these confusing features that self embodies, it is time to move on. Now, according to Sluga (1996, 350), Wittgenstein's approach has one extremely interesting consequence. He thinks that,

because Wittgenstein's 'self' is merely imaginary, it cannot possess any causal powers. I am not sure whether we are allowed to infer such drastic implications from what has been said. For instance, in *Notebooks* (80) Wittgenstein states that "[t]he thinking subject is surely mere illusion", yet he adds that "[...] the willing subject exists". This statement seems to suggest the same as my previous claim according to which substantial self is the necessary subject of intentional actions, and thus its existence should not be challenged within the very context. Now, if we accept this thought, we also need to accept that self may possess causal powers. After all, if I am to say that I have done something intentionally, I am undoubtedly implying that I am the cause of that action. Without me, the action simply would not have happened, and therefore I am the one to blame for it. As we will remember from Section 2.3.2, in the case of intentional actions reasons are mere propositions, and hence they serve only as grounds for actions instead of sharing any causal relations. Since also intentional actions must find their place in our causal world view, the only causal power behind them to make sense is substantial self. After all, self is a thing which contemplates and makes decisions in the light of reasons. This parallels with Kant's account. As we will later see, he (*Groundwork*, 4:448) thinks that human beings are necessarily free in their acts, and that this freedom includes the idea according to which the only possible coexisting incentive for intentional action must be something which the agent has made as his maxim. By maxims, Kant means universal rules which share the same role as non-causal reasons, namely the role of being grounds for a decision instead of being its cause. Anyway, the basic idea is that, while an agent acts for propositional reasons, any other cause for the action besides that agent himself cannot exist. This is so, because otherwise the agent would not be free and the reasons would not have any function. Henry E. Allison's (2006, 390) describes Kant's account of agency lucidly while saying that "[...] this idea is that of an uncaused cause, that is, of an agency capable of making an 'absolute beginning', by which is understood the capacity to initiate a causal series that is not itself determined by any antecedent condition". In other words, while acting intentionally, an agent always initiates a new causal chain. If there was a preceding cause for the action, it would steer the causal chain solely through the agent instead of giving him the main part in the play, as happens with voluntary actions. Therefore, only through taking self as the cause of an action, are we allowed to consider it as responsible for what has happened. This idea of responsibility brings us to the last question of the section, namely that of free will.

It is an irrefutable fact that the idea of the freedom of will is included in the idea of intentional action. Here, by freedom I mean true freedom in its most positive sense, namely the ability to choose between alternatives. This was well put by Thomas Reid in one of the quotations above (see page 88–89). As intentional action means action performed in the light of non-causal reasons, it would be extremely hard to find any sense in which these kinds of reasons could have any kind of influence on action, if the agent was not free. Reid compares a motive or reason with advice and exhortation, saying that only the supposition of liberty gives them justification. After all, what good would it do to advise a man to get some exercise, if the man in question had already been forced to run round a track? The most interesting feature in non-causal reasons is that, even though they explain actions, they do not necessitate them the slightest bit. Nagel (1986, 115) gives an illuminating example of this:

When someone makes an autonomous choice such as whether to accept a job, and there are reasons on both sides of the issue, we are supposed to be able to explain what he did by pointing to his reasons for accepting it. But we could equally have explained his refusing the job, if he had refused, by referring to the reasons on the other side—and he could have refused for those other reasons: that is the essential claim of autonomy.

As was said in the preceding paragraph, these reasons are incapable of producing actions by themselves, and thus they need something to do it for them, namely the self of an agent in the shape of choices and decisions. Now, as self does not have any nature in itself on which we could get a grip, there is a large gap between self and reasons on one side and action on the other. Since we are not able to acquire any empirical information about self, neither is it possible to find something that would render the causal relation between self and its actions necessary. The only things to be found are reasons, but even they are mere justifications instead of being something which could put some further element in the causal relation. On the contrary to voluntary actions, the gap in intentional actions is not there because of our ignorance but because otherwise reasons would simply be meaningless. In other words, the gap here is necessary for our understanding of intentional actions, and not even all the knowledge in the world could close it. The empty space in question is for freedom, or to put it another way, the empty space *is* freedom. To conclude, all intentional actions are free by definition. Therefore, as long as people act intentionally, they have the power to choose freely, and this power cannot be taken away. In the context of intentional actions, and only



there, human beings possess will that is free and, within that context, nothing can threaten its place.

In this section, I have compared the notions of self and will between the contexts of voluntary and intentional actions. I have argued that, only with the latter, can will be described as truly free. All in all, it has become clear that the two contexts provide totally different views of many things, including personal identity and the causal nature of action. In the last section of this chapter, I will reach conclusions on what these heavily contrasting contexts have to do with our everyday life. What does it actually mean that we are able to consider one and the same action from various angles? Are we making conscious choices all the time between the contexts or does it all happen without noticing? Moreover, are there certain rules governing which perspective we need in different contexts?

### 3.3.6 Changing the Perspective

This Section 3.3 goes by a name referring to social mechanisms, but little by little I have found myself further and further in the most basic theoretical issues concerning the understanding of actions and their agents. Naturally, I apologize for all the confusion it must cause, yet I am willing to adhere to the decisions I have made. As I have been arguing that we are able to choose between various perspectives through which we may contemplate human action, and that these different perspectives share a great deal of contradicting elements, I guess I should definitely say something more about the very idea of changing perspective. This is necessary not only to complete the whole picture but also to clarify what I mean by saying that in social mechanisms we need to see actions as voluntary. Perhaps the most important question is whether I am saying something general about the nature of human action or not. What does it mean that we can choose altering and even contradicting accounts for human beings and their actions? However, these questions I will contemplate later while making the final conclusions of my work, but before that it is time to pin down the mysterious process of changing the perspective towards an action.

If we imagine some item in front of us, say, a cake, there are various ways in which we may appreciate this object. I will not go as far as claiming that we are able to see it as a telephone, for example, if it does not look like it, but definitely we have the freedom of calling it ‘an unhealthy dish’ or ‘a bakery product’. Moreover, there cannot be any doubt about our freedom of defining how accurate a view we are willing to have of the object. For instance, we may take it as a mere cake or, alternatively, as a big chocolate cake. Naturally, we are still referring to the same thing, but in the latter case we are bringing out more properties making the picture much more accurate. Furthermore, one can always sharpen the image more and more assuming that there are more properties to grasp. For example, it is much vaguer to talk about a mere cake instead of a big chocolate cake which the town baker has prepared for the mayor’s birthday. In this latter picture, we go as far as even explaining the existence of the cake by telling what its cause is and what it has been made for. Now, all that has been said goes to actions and other events as well. For example, we may say that there is a man waving his hand, yet not specify it with more detail. However, as events and actions leave at least as much room for interpretation, instead of the sole finding of new properties in the same action, it is very easy to conceive completely new actions in the same phenomenon. For instance, we may claim that the man is giving us a sign. Now, this is a completely different action to the first, as waving one’s hand is clearly different to giving a sign. In any case, if we are now to focus on the action of giving a sign, we may add further properties, as we did with the cake in the above example. For instance, we may say that the man is giving us a sign with a wild look in his eyes. Now, adding this kind of new property does not say anything about from which perspective we are viewing the action, although most likely the perspective was chosen at the very first moment. If we are to explain the action, on the other hand, in many cases we must reveal our perspective. However, there are also a great number of exceptions. For example, we may say that the man is giving us a sign because there is a car coming, yet outsiders may not know for sure whether we mean that the arrival of the car is causally making him to do it, or that the fact that the car is coming is the reason for his choice of giving the sign. Nonetheless, even though solely by examining the language it may be impossible to find sufficient evidence for any perspective, it is not something we have to contemplate in our everyday life. On the contrary, normally the perspective is clear from the beginning, and thus can be unraveled without any inconvenience. If we are to consider our example once more, it should be obvious that, while examining the behavior of a stranger, we are doing it from the objective perspective, in other words, we are talking about it as voluntary. In the end, the basic distinction here is quite clear. It is also the reason why

we normally do not need such artificial properties as ‘voluntary’ and ‘intentional’ used in the sense I have done. When I talk about adding the abovementioned properties, I actually refer to the procedure of choosing the perspective which we usually do unconsciously, and hence it is understandable that any specific properties are seldom needed. After all, for an action to have a property of ‘being intentional’ or ‘being voluntary’ is nothing other than having certain other properties and tendencies for properties of a particular type. For example, it is to have a certain kind of causal nature. Next I will contemplate the process of choosing the perspective.

Basically, there is only a choice between subjective and objective perspectives which means that either we try to see an action from the agent’s perspective or we do not. An exception to this, though, is the moment we act ourselves, since then we do not have any other options but to look at our action from the subjective perspective. While making decisions, we necessarily see ourselves as metaphysical selves evaluating the alternatives and, within this process, we are doomed in absolute freedom and responsibility. After all, if we merely stayed waiting for something to cause us to act, we would never perform a single action besides the endless waiting. Kant already brought out this point in his own florid way as he (*Groundwork*, 4:448) stated that “[r]eason must regard itself as the author of its principles independently of alien influences; consequently, as practical reason or as the will of a rational being it must be regarded of itself as free, that is, the will of such a being cannot be a will of his own except in the case of the idea of freedom, and such a will must in a practical respect thus be attributed to every rational being”. In this light, it is perfectly understandable that Kant takes the freedom of will as given. After all, as long as we act we must necessarily consider ourselves as free, and after we stop acting we cannot object any more, since it would just be one more action. Of course, as I mentioned at the beginning of this section, there is also the question of how seriously the images these perspectives are drawing should be taken, but this I will consider later. Either way, while we act and make decisions we must take ourselves as acting intentionally which, as we remember, means a certain kind of view of our actions, their reasons and even ourselves. However, as these things are rarely very simple, the possible scope of this subjective perspective is not limited only to our own actions. Evidently, it cannot be denied that it is possible to take it towards people around us as well, given that from outside it is a piece of mere speculation as only the agent himself can possess the proper subjective perspective. We may consider an example where I debate with my girlfriend whether she should accept a certain job offer. Now, as we

are contemplating the *pros* and *cons* of taking the job together, apparently we are both seeing my girlfriend as a substantial self within the process of making free choices, even if we do not pay any attention to the fact. As Reid mentioned earlier (see page 89), it is not worthwhile giving an advice to a person whom we do not regard as free. The same goes for all the people we truly want to understand, not as bare creatures at the mercy of the outside world but as people standing behind their decisions. For instance, I have wondered many times what Knut Hamsun saw in the National Socialism that made him support it so powerfully. This example brings us to the one of the most important factors in intentional action, namely morality.

If we are to explain actions or whatever with causal connections, undoubtedly the causal reasons should not be taken as good or bad in themselves, except if the causal factor is a metaphysical self. For example, let us imagine a situation where our favorite tree has fallen because it has been struck by lightning. In this picture, it would make sense to be angry and to see this natural phenomenon as unfortunate to the highest degree, yet it would be rather odd to put the blame on anyone except God or a similar creature, even though we often do it. We may accuse the lightning of being heartless, yet we know deep inside that our behavior is quite mindless in all its weird rancor. The lightning happened because of some antecedent causes which again had some other causes and, as this chain goes on to infinity, no-one in it has done anything for the sake of, say, evil. What is interesting is that the same thing should come to our minds while thinking about agents acting voluntarily. After all, there is a causal chain bringing about mental states which again cause some kind of behavior and, in the end, there is no one to blame for what has happened. Regardless of this, we often blame those bundles of dispositions. For example, it is quite natural to blame our neighbor's dog for some doggy tricks, yet the blame in these cases should be of a different kind. It cannot be moral blame for a self choosing badly as it is blame on a thing being bad, in this case, a no-good combination of dispositions. At least, this is what we should think, but the problem is that sometimes we are also willing to see non-rational agents as rational. Anyway, to be fully appropriate, this kind of blame should be directed at the ensemble of states which cannot be morally good or bad but on which one can have influence. Similarly, if we were to see criminals acting voluntarily, which means that their actions are results of antecedent conditions, punishments would make sense only in the light of bringing up these people. If there is no metaphysical self making decisions, neither is there a self receiving punishment. This is usually the account of the philosophers who deny the existence of free will (for example, Pereboom 2001, 158–186). If nobody is doing things freely, there is no one to take

the blame. The only thing left to do is the changing of the combinations of dispositions through punishment, education and so forth, so that the person would behave in a different manner next time around.

I believe it is self-evident that the idea of morality is tied to the activity of substantial self. If we are to take somebody as fully blameworthy for his actions, he must have acted freely instead of as a mere automatic reaction to outside causes. In other words, there needs to be an intentionally acting, continuous self to take the blame. This was Kant's (*Groundwork*, 4:447) account, and Wittgenstein (*NB*, 5.8.16.) also put it quite clearly when calling the I "the bearer of ethics", and saying that "[w]hat is good and evil is essentially the I, not the world". Consequently, the aspect of intentional actions is the only one which justifies the idea of retribution solely in itself instead of being mere means of bringing up. This kind of account of punishment is called *retributivism*, and stands for the idea according to which criminals need to be punished first and foremost because of atonement. Here the difference between intentional and voluntary actions is easy to see. Especially, in situations where a person is being blamed for his actions but where somebody is still trying to defend him by picturing the action within the frames of voluntariness. For example, one may be accused of a cruel murder done for vicious reasons, but the defense counsel may still try to bring to light the horrible childhood the accused suffered from. What the lawyer is doing here is trying to hide the self behind the unfortunate causal chain which has befallen the poor murderer. Now, the next question is what dictates the amount of moral blame and praise. Of course, the action itself plays a crucial role here, yet the seemingly same actions can be judged in various ways. The answer is evident as all intentional actions are naturally evaluated in the light of their non-causal intentions and reasons. A piece of a causal chain can never be bad or good in itself as it merely happens, but a propositional reason in the light of which a thing is done can play that part. Similarly, a mental state cannot be wicked or noble as it is solely a state contrary to a propositional intention. This is because the propositional reasons and intentions are something pictured in the eyes of the agent, and thus are not mere events in the world. Quite the opposite in fact; they are something formed or chosen by the agent. This is what G. F. Schueler (2003) means by saying that there is always an evaluative aspect in non-causal reasons. They are something the agent has weighed, and hence he is to be blamed, if he has done his weighing negligently or even wickedly. Only by true deliberate actions must one stand. These deliberate actions are intentional ones made on the grounds of non-causal reasons.

Even with these few exceptions, most actions seen from outside are seen as voluntary. In my study, I have also brought out the aspect of pure physicality but, when human action is concerned, it is seldom one for common people in their normal surroundings. After all, it seems to be only for natural scientists to be interested in the physical causal chains behind bodily movements. Anyway, it is quite natural to see the actions of our fellow citizens in the light of voluntariness instead of intentionality, and this is mostly for two reasons. First, we do not have the will, energy and even capacity to step into the agents' shoes and to see the world through their eyes in most situations. For instance, while watching people running about in the sales or players striving for victory in a soccer match, it is likely that we must conceive them as mere creatures driven by their appetites and beliefs. This is how we understand the rest of the animal kingdom, and hence it also serves perfectly in most situations where human agents are playing the main role. Trying to see the world through the agent's eyes necessarily requires more effort as we must figure out all the various reasons that he may be measuring. Second, it is also natural for us to approach situations as truthfully as it is by any means possible. By this I mean that, after all, the aspect of voluntariness cries out for objectivity, which is completely contrary to the subjective aspect of intentional actions and, in most cases, the objective perspective is valued much higher than the subjective one. Whereas the subjective view is mostly for empathy and understanding people, the objective view is taken for a higher form of knowledge, namely scientific understanding. After all, objectivity means stepping to the other side of the agent's own ideas and revealing the true factors behind his actions. This contrast is easy to understand with the examples of self-deceptive agents. If we are to imagine once again an agent who has a drinking problem but does not admit it to himself, it becomes clear that the knowledge of the problem and, above all, the reasons behind the problem are much more valuable than the illusions of the agent about his drinking habits. As it comes across as natural for human beings to approach the world around them as profoundly as possibly, obviously the more complete causal chains have more importance than those that begin from a mysteriously free agent. Nevertheless, even if the objective information seems to carry more weight than the subjective one, as was presented above, sometimes the subjective side must necessarily be taken. Anyway, one interesting feature in the objective voluntary aspect is that we have the ability also to take it towards actions which are in the first place considered as subjective. For example, when the abovementioned alcoholic finally understands and admits his condition, he may start to contemplate his own deeds as voluntary springing from physical and mental disorders, although at the same time he must always act from the intentional perspective. All in all, the voluntary perspective is

not limited to the mere actions of people outside, just as the intentional perspective is not limited merely to the agent's own deeds.

In this section, I have sifted through the various traditional cases where actions are conceived as voluntary or intentional. Now, in the light of all that has been said, it should be more or less easy to grasp what it means to claim that we are free to choose between the perspectives through which we contemplate actions around us. I believe it is right to say that in some sense the process is free and in another it is not. What I mean by that is that when we take an action as voluntary or intentional, we add the property there automatically. For example, while observing a man driving a car, we do not spend even a second wondering whether we should contemplate it as a voluntary or intentional action. On the contrary, the context completely determines our aspect towards the action. For instance, as I have argued in this chapter, in sociological mechanisms all actions must necessarily be seen as voluntary. However, the freedom still lies there as we often also have the possibility to reconsider. As was the case with the abovementioned alcoholic struggling with the recognition of his problem, he also had the freedom to figure out the true causes of his behavior. Similarly, we usually have the possibility of trying to see the world through the eyes of a voluntarily acting agent. In that case, we separate the substantial self from the mere bundle of mental activity and start to evaluate the reasons on the grounds of which the choices must be made. Hereby, I conclude the final chapter of my study by saying that, even though the context always determines our perspective to actions, in most cases we also have the freedom to reconsider the context and thus to change our perspective on them.

## **Conclusions:**

### **A Man on a Bus -Part II**

I guess that it is finally time to get back to where we started from, namely to the man still patiently sitting on the bus. After all, at the beginning of my study, I introduced a rather simple thought experiment which can be considered the starting point of the whole process. The main problem was plain. I imagined there being a bunch of scientists providing perfect predictions and explanations for the event. They were able to both predict that the man is going to take the bus at the time and to explain it completely. Furthermore, the explanations they were offering did not refer to entities like decisions, desires or such like. Now, assuming that the predictions and explanations work as well as is possible, one should start to wonder what role those factors could play which we normally use to explain actions. Moreover, it appears that the predictions are nonetheless dependent on the decisions of the man in question, as evidently the man would not end up sitting on the bus without his own assistance. As I started to contemplate this seemingly innocent case, it gradually emerged that, in order to find a solution to the problems presented, one needs to submerge himself in the most fundamental issues concerning the nature of events and actions as well as the idea of explaining. Here I have travelled down this long road, and thus I believe that I am finally able to provide a solution to the tension springing from the possible existence of various explanations for the sitting on the bus. This solution, however, bears some consequences which must be acknowledged. As I noted, my treatment of the problem happened mostly in the very basic structures of our understanding of the world, and therefore the fundamental statements themselves should be taken seriously. After all, it is quite natural that the most important results of my study should not have much to do with the man sitting on the bus. Clearly, they help us to understand the event profoundly enough so that the tensions should not arise any more, yet they share much more significance on their own. Besides, perhaps even the most crucial question is yet to be answered. However, before more fundamental issues it is time to take the man on the bus under scrutiny for the very last time.

The first chapter of my work was, first and foremost, about the idea of explaining. In pursuance of introducing the idea of scientific mechanism, I took the most fundamental



questions concerning explaining under closer study. However, as we saw, understanding the idea of explaining deeply enough requires understanding of the nature of causality which again cries out for understanding of the most fundamental issues concerning our perceiving and talking about the world about us. In the end, in order to do things properly, all the required steps needed to be taken. Before long, I had to face the question of what events actually are. Now, answering this question turned out to be one of the most crucial pieces for our solving the main puzzles of my work. My reply was, simply, that they are properties of the phenomena of which we cannot say anything other than that they carry the very properties mentioned. Moreover, as I concluded that explaining in science is bringing out the right causal history and that causality is a relation only between properties, it came out that what is at stake in explaining is the reference to relevant properties causally related to the property in need of an explanation. On the other hand, what defines what is “relevant” for an explanation is the context.

Even if everything mentioned are highly important results in the light of the original problem, they are clearly not sufficient in themselves to settle it. Evidently, the tension does not arise, at least to that extent, in all events as the problem seems to lie in the special nature of actions. After all, with actions there appears to be a collision between the various types of explanations which even seem to contradict one another in certain aspects. Hence, the second chapter of my study was dedicated to actions. After a while, I realized that it is simply impossible to give a brief definition of the concept of action without ending up with such a vague thought according to which when acting, “somebody or something is doing something”. As I grew aware of the amount of ways of using the term, I concluded that it is best to distinguish the various aspects of action. In the end, it came across as evident that, while talking about actions, it matters a great deal whether we refer to purely physical movements, voluntary actions explained with mental states or genuinely free intentional actions. Besides, not only the view of action changes but the whole picture, including the agent and explanation by the same token, while shifting from one perspective to another. What serves as a theoretical justification for the possibility of viewing actions from different perspectives is the idea from the previous chapter according to which all events, including actions, are properties, and hence numerous different actions can be found in one single phenomenon. Thus, from all perspectives we see different actions, yet they are possibly referring to the same phenomenon. Now, in the first two chapters I managed to lay a foundation from which the original problem could be met. In the third chapter of my study, I

unified what had been learned in the first part. Here I will demonstrate these conclusions once more by way of our case of the man sitting on the bus.

So, let us imagine that there is a natural scientist capable of providing a complete explanation for the fact that the man is taking the bus. Presuming that the explanation consists only of physical facts, we should call it P. Now, if we are to ask what P is actually explaining, the evident answer would be to say that it is explaining the physical aspect of the phenomenon. After all, it would be hard to imagine any other kind of *explanandum* for P, if there are only physical facts mentioned. However, P cannot be all that is needed. For example, we may also picture there a social scientist asking the natural scientist why the man is sitting there on the bus. Now, if the natural scientist cited P in order to answer the question, it is very unlikely that the social scientist would be satisfied with the answer. Talk about neurons, atoms or whatever does not serve as information for anyone except the specialists of the given field. Either way, even if it were that the natural scientist could provide a P of utmost simplicity and accessibility, it seems unlikely that the social scientist would take it for a proper answer. Plainly, it would not be an answer to the question he posed. This is the crucial point. After all, what is most interesting is that both scientists are seemingly asking the very same question, namely ‘why has the man taken a bus?’, but if the two really were the same question, the same answer would serve both. This is not the case here. As they are both seeking different answers, their why-questions must also diverge in some aspect. Now, if we remind ourselves what has been learned in this study, it is easy to pinpoint the difference. Even if both the scientists are referring to the same phenomenon and even to the same basic action, they are nonetheless interested in some special aspects in them. Therefore, although they are seemingly talking about the same action, they are actually talking about different actions as one is interested in the physical aspect and the other in voluntary acting. As has been said, normally disputes of the kind do not arise as the context determines the aspect. That is also why people do not need to specify their questions under normal circumstances. Anyway, what the social scientist is looking for is an explanation telling us why it was so that the man voluntarily ended up taking the bus. We may call this explanation M as it likely carries some information about the agent’s mental states. Evidently, however, even this is not enough. Let us imagine there another man sitting beside the man on the bus asking him why he is sitting there. Now, clearly by citing P, the man would not give a proper explanation as it is more likely that the questioner would think that the other was making fun of him. Furthermore, in many instances M would also be considered inappropriate. For example, by

saying that people who did not have proper physical education while at school avoid walking as much as they can, he would also confuse the questioner. What this questioner is seeking is not a causal chain but arguments the man could use to justify his way of travelling. Most likely, he is asking what the grounds were on which the man made his decision. For instance, by saying that, although he knows walking would be good exercise, it is so boring that he prefers buses, he would give a perfectly valid answer. Here he is explaining why he took the bus intentionally and, even though the reasons he is giving are not causal themselves, the explanation overall is as it presupposes the causal decision-making force, namely the self.

All in all, my answer to the problem where the whole study began is based on the idea according to which, instead of having various independent explanations for the same action, we have many different *explananda* which all cry out for different kind of *explanantia*. Furthermore, above I only presented one example of all the three types of action introduced in my study, yet numerous others could have been found. For example, within the phenomena there might also be actions of, say, ‘a man sitting quietly on a bus’ and ‘a man watching through a bus window’, and all of them could require different kinds of explanations. Besides, the very action defined in the questions may also be seen in various ways even without taking sides in the question of the aspect. For instance, the latter two explanations in the abovementioned example were evidently answers to the question of why the man took a bus instead of walking. Similarly, the questioners could have been more interested in the question of why he took this bus instead of some other one. In the physical realm, distinctions of this kind are hard to make but, in the case of voluntary and intentional actions, they are highly important. Anyway, as has been said many times, what defines the causal chain asked is the context. Still, sometimes a kind of specification is needed on behalf of the questioner so that the one who is providing the explanation understands what is in question. After all, merely by examining the sentence ‘why is the man taking the bus?’, it is hard to see what is needed for a proper answer. When it comes to predicting, the issue becomes even more interesting. If one is able to find extremely well working causal structures to explain a constantly manifesting action, there is no obstacle in also predicting similar actions yet to come. However, to be strict, this goes only for the physical side of actions. As we remember, in the absence of social laws in voluntary actions the full guarantee cannot be reached. Similarly, as intentional actions always include the idea of freedom, the hope of a watertight prediction is in itself contradictory. Nevertheless, usually one can also make quite good predictions about voluntary and intentional actions. Only perfect certainty is impossible to

reach. So, interestingly enough, in our case only the physical presence of the man on the bus could be validly predicted.

Here I have presented my solution to the original problem as explicitly as I can. However, even though it obviously removes the tension between diverse explanations to note that, in the end, they are explanations for different actions, it nonetheless leaves further questions behind. While I have introduced my ideas to other philosophers, I have often faced similar confusion and objections which cannot be ignored. Simply said, my idea is that, in this phenomenon of a man sitting on a bus, an enormous amount of different actions can be found. This is not as much a problem as is the fact that some of these actions present totally different pictures of the agent and course of events, and thus they may even go as far as presenting seemingly contradictory images. As we remember, even Aristotle (see Section 2.3.1) thought that actions as well as all events can be explained in different ways, but in his view there was no collision. For instance, there was no freedom in any of the explanations, and therefore the existence of several explanatory factors was not a problem in itself. However, this is not the case in my account. In my view, if we are to ask on a general level whether agents possess free will or substantial self, both answers, namely 'yes' and 'no', are valid in different perspectives. Similarly, to the question about whether reasons for actions explain actions causally, the answer depends on the aspect chosen. Obviously, this is not a minor problem. In the philosophical tradition to accept both positive and negative answers to one more or less unambiguous question seems to be the same as allowing the existence of clear contradictions. Nevertheless, as we saw, Ludwig Wittgenstein was one exception to this tradition as he appears both to deny and approve the existence of metaphysical self (see page 158). In my view, Wittgenstein was right as he seemed to understand something which unfortunately is usually forgotten.

With concepts like 'will', 'self', 'action' and 'reason for action', it is undeniable that in different instances they may share different meanings. Moreover, it is often so that the people using these concepts are not even very aware of what it is that they are actually talking about. For example, the seemingly simple question of whether people have free will turns out to be far from simple, if we stop to define in detail the concepts used. To my mind, what is crucial in Wittgenstein's thinking in the light our discussion is that he (*PI*) realized that it is extremely hard or even impossible to define many concepts precisely and explicitly, because they can be used in so many ways in our language, and that while using them differently we associate them with different features. As we have seen in my study, all the concepts

mentioned above can be used in very diverse contexts, and within these contexts their meanings can differ to a great extent. So, while asking the question of whether people have free will, our answer should depend totally on the context, that is to say, on the meanings attached to the concepts in the very context. All in all, there is no single general question of the existence of free will. On the contrary, there are at least two profoundly different ways to understand the concept of will, and both these ways share different features. Furthermore, in certain contexts the whole question about will and its freedom loses its meaning as the concept of will may not even have any place within it. Therefore, in this sense, the contradictions in my account are merely illusory.

In any case, I am quite certain that the given answer is not capable of erasing all doubts. Even if one could accept what has been said thus far, there still remains one highly fundamental question to be asked. Now, even if one accepts that, instead of only one, there are many different actions in the case of the man sitting on the bus, one may still ask what it actually means that we are able to see, say, the agent of these actions both as free and not free. If we are able to decide whether to contemplate this man as a genuinely free agent or a mere bundle of mental states merely reacting to certain stimuli, it simply seems that there is too much room left for interpretation. After all, there are evident reasons why this sort of relativism is usually avoided among philosophers. If everything is relative in a sense that all depends on the context, it seems possible that, before long, we will end up in a situation where all the rules disappear and everything is allowed. Besides, it comes across as only a cowardly attempt to avoid all the controversies to say that all the answers are valid in certain contexts. Moreover, with actions there are further implications. For example, as we remember, moral responsibility is usually associated with the freedom of will. Now, by saying that our freedom is relative, we also define moral responsibility as relative, and this is something many cannot bear. If we are always able to see actions and their agents at least from two different perspectives, it follows that it must be perfectly valid also to consider, say, Adolf Hitler both as a tremendously evil man and as a poor victim of the surrounding causal world. To mention one instance, I remember that while I was presenting my account at a conference in England, one person became genuinely worried that my view would erase all the evil from the world. Even if his concern initially sounded somewhat amusing as it would not seem so terrible to get rid of all evil, this kind of concern is rather common and also highly understandable. After all, if we are given the freedom to picture actions differently and if all the perspectives are equally justified, it appears that we cannot say anything definite

about the world any more. If Hitler becomes equally evil and non-evil, in the end it seems that we must face a situation where it is equally justified to judge him as not to judge. Understandably, this is not a situation where we want to spend the rest of our lives.

To people sharing this apprehension I wish to say two things. First, there must be some kind of misunderstanding as, contrary to what has been claimed, by justifying various perspectives we cannot lose evil or good but merely guarantee their existence. After all, since the perspective of intentional action is necessary in a sense that living without it would not even be possible, as long as there are people doing things, both evil and good deeds will be done. However, most likely concern will still arise, as there is a common intuition according to which more objective views are somehow truer. Similarly, people often seem to think that the facts from, say, physics are somehow above everything else and that if there is no freedom, for example, to be found on the most basic physical level, it cannot be found anywhere else either. This is precisely the kind of thinking I have been trying to prove wrong in my study. After all, if concepts like 'will' and 'self' did not even mean anything within the physical realm, how could any physical facts possibly say anything about them? Furthermore, as we saw in voluntary actions, explaining with mental states is a rather questionable procedure as their whole idea is extremely vague. In light of what has been said, I would not consider the perspective of intentional action as any less real than the other ones. Second, even if it would be nice and simple to live in a world which could be easily pieced together, unfortunately this is not the case with the world we live in. On the contrary, if that had been the case, there would not have been much work for philosophers over the centuries. Now, although it undoubtedly would be pleasant to make clear and definite conclusions about the world, we should not do it at the expense of compromising on truthfulness. In the end, it is hard to believe there being anyone who could honestly claim that he is not able to see himself as well as other people as free agents making genuine decisions. Likewise, it would sound odd to claim that there is something wrong in explaining actions by referring to mental states or even to neurons. For example, who could honestly assert that he could not picture Hitler as a man giving horrible orders freely for wicked reasons? As we see ourselves acting, it is easy to picture a self behind those actions measuring reasons and making genuine choices. Similarly, who would dare to claim that it is not adequate to say that, among other things, genes caused Hitler to behave as he did? If we really think of it, it becomes clear that we are also able to consider Hitler as a mere result of antecedent conditions only reacting through various dispositions. In my view, to say the opposite is a clear sign of falsehood. Perhaps we

do not know enough about the conditions, but if we saw the whole causal picture leading to even the most horrible deeds, certainly they would make perfect sense. Even so, the fact does not erase the evil from the actual decisions made by Hitler. All in all, I believe that in some sense we are all able to understand Hitler, but in another I hope none of us is. All that has been said should also be taken into consideration in the question of moral responsibility and punishments. In my view, there is nothing wrong in retributivism (see page 165), yet it should not be forgotten that at the same time punishments should also serve the purpose of educating the criminal as well as the rest of the nation. It is perfectly justified to think that people need to be punished for their bad deeds, but still by giving unfair punishments we are capable of turning things from bad to worse. After all, as people are as much bundles of mental states as they are substantial selves, we should pay attention to what kind of mental states we give emergence.

Before closing my study, there is one more issue which should be taken under scrutiny. In the paragraph above, I stated that the perspectives must be justified as, in all other cases, we would somehow “compromise on the truth”. This is a highly interesting statement as one natural consequence of such a relativistic account as mine is usually taken to be the inflation of truth. I guess it is understandable to assume that, if we allow there to be more than one truth, we no longer have the authority to deny anything. In other words, after accepting more than one truth, we are compelled to accept all of them. Of course, the ideal situation would be one where only one perspective is real and only one truth is held but, as I have been arguing, the world we live in is far too complex for that. At least we human beings are capable of applying different perspectives which all present images impossible to ignore. Anyway, even if I can understand the threat, to my mind it does not seem very real. After all, saying that we are able to answer a question in two totally different ways in different contexts does not imply that we are allowed to answer the question in any way we like. Even if there is more than one causal structure with which we may explain things, that is not to say that there are no rules in these structures. On the contrary, the rules do not vanish as we only accept that there is more than one set of rules. For instance, it is perfectly appropriate to say that Hitler did his horrible deeds because of his genes or his childhood, among other things, yet it is not right to say that he did them because of his ears or the childhood of Nelson Mandela. Similarly, we are allowed to say that he did those things for indefensible reasons, and thus we should not say that he did them for the right reasons. After all, there are certain strict rules in causal language, and even the acceptance of various causal structures do not give us the

privilege to make any causal claims we like. For example, according to my account it is perfectly accurate to say that by throwing stones, we may break windows, but the same does not go the other way round. By throwing windows, we cannot break stones, even in my theory.

The world is a complicated place at least for us who are capable of seeing it in various ways. This ability is both a blessing and curse. While we are able to see the multitude of aspects and properties there are in all phenomena, we share the pain of losing clarity in the midst of all the possibilities. Anyhow, still the world should not be rendered simpler than it is, since as we have seen, various philosophical disputes are developed only because of the stubborn will to deny the possibility of various perspectives. Even if it undeniably would be nice and clear to be able to apply only one perspective to things going on around us, this is not the case, and thus it is plainly false to adhere only to one truth. On the contrary, only by accepting the diversity of the world are we able to truly understand the mess different perspectives and altered meanings for concepts induce and, only by making genuine attempts to clear up this mess can understanding be reached. This is what I have been trying to do here. My main aim has been to clarify the picture the various perspectives are providing and, by the same token, to define the meanings concepts related to action have within the changing pictures. I hope that I have succeeded even slightly in this process. Therefore, it is time to finish my study which I have written both as a free agent on the grounds of sharing my carefully matured thoughts, and as a mere bundle of mental states driven mostly by the compelling academic ambition which my parents have slowly implanted in me since I was a little boy. In both cases, the results are the same.



## Bibliography

- Abbott, Andrew (2007), “Mechanisms and Relations”, *Sociologica* 2: 1–22.
- Acrill, J. L. (1981), *Aristotle the Philosopher*, Clarendon Press, Oxford.
- Ainslie, George (1992), *Picoeconomics: The Strategic Interaction of Successive Motivational States within the Person*, Cambridge University Press, Cambridge.
- Albritton, Rogers (1985), “Freedom of Will and Freedom of Action”, *The Proceedings and Addresses of the American Philosophical Association* 59 (2): 239–251.
- Allen, Garland E. (2005), “Mechanism, Vitalism and Organicism in Late Nineteenth and Twentieth-Century Biology: the Importance of Historical Context”, *Studies in History and Philosophy of Biological and Biomedical Sciences* 36: 261–283.
- Allison, Henry E. (2006), “Kant on Freedom of Will”, in *The Cambridge Companion to Kant and Modern Philosophy*, ed. Paul Guyer, Cambridge University Press, New York, NY: 381–415.
- Anscombe, G. E. M. (1971), *Causality and Determination*, Cambridge University Press, Cambridge.
- Anscombe, G. E. M. (2000 [1957]), *Intention*, Harvard University Press, Cambridge, MA.
- Aristotle (1995 [1984]), *Metaphysics (Met)*, trans. W. D. Ross, in *The Complete Works of Aristotle, Volume two*, ed. Jonathan Barnes, Princeton University Press, Princeton, NJ: 1552–1728.
- Aristotle (1995 [1984]), *Movement of Animals (MA)*, trans. A. S. L. Farquharson, in *The Complete Works of Aristotle, Volume one*, ed. Jonathan Barnes, Princeton University Press, Princeton, NJ: 1087–1096.
- Aristotle (1995 [1984]), *Nicomachean Ethics (NE)*, trans. W. D. Ross, in *The Complete Works of Aristotle, Volume two*, ed. Jonathan Barnes, Princeton University Press, Princeton, NJ: 1729–1867.

- Aristotle (1995 [1984]), *On the Soul (DA)*, trans. J. A. Smith, in *The Complete Works of Aristotle, Volume one*, ed. Jonathan Barnes, Princeton University Press, Princeton, NJ: 641–692.
- Aristotle (1995 [1984]), *Parts of Animals (PA)*, trans. W. Ogle, in *The Complete Works of Aristotle, Volume one*, ed. Jonathan Barnes, Princeton University Press, Princeton, NJ: 994–1086.
- Aristotle (1984), *Physics (Ph)*, trans. R. P. Hardie & R. K. Gaye, in *The Complete Works of Aristotle, Volume One*, ed. Jonathan Barnes, Princeton University Press, Princeton, NJ: 315–446.
- Armstrong, D. M. (1968), *A Materialist Theory of the Mind*, Routledge & Kegan Paul, London.
- Armstrong, D. M. (1983), *What Is a Law of Nature?* Cambridge University Press, Cambridge.
- Augustine (1991), *Confessions (Conf.)*, trans. Henry Chadwick, Oxford University Press, Oxford.
- Augustine (1964), *On Free Choice of the Will (On Free Choice)*, trans. Anna S. Benjamin and L.H. Hackstaff, Prentice Hall, Englewood Cliffs, NJ.
- Bacon, John (2008 [1997]), “Tropes”, in *Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, <http://plato.stanford.edu/entries/tropes/>.
- Baker, Lynne Rudder (2000), *Persons and Bodies: A Constitution View*, Cambridge University Press, New York, NY.
- Bechtel, William & Abrahamsen, Adele (2005), “Explanation: a Mechanist Alternative”, *Studies in History and Philosophy of Biological and Biomedical Sciences* 36: 421–441.
- Bennett, Jonathan (1988), *Events and Their Names*, Clarendon Press, Oxford.
- Biggs, Michael (2009), “Self-Fulfilling Prophecies”, in *The Oxford Handbook of Analytical Sociology*, ed. Peter Hedström & Peter Bearman, Oxford University Press, Oxford: 294–314.
- Boas, Marie (1952), “The Establishment of the Mechanical Philosophy”, *Osiris* 10: 412–541.

- Bramhall, John (1999), “Bramhall’s Discourse of Liberty and Necessity” in *Hobbes and Bramhall on Liberty and Necessity*, ed. Vere Chappell, Cambridge University Press, Cambridge: 1–14. Originally published in 1655.
- Brook, Andrew (2008 [2004]), “Kant’s View of the Mind and Consciousness of Self”, in *Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, <http://plato.stanford.edu/entries/kant-mind/>.
- Bunge, Mario (1997), “Mechanism and Explanation”, *Philosophy of the Social Sciences* 27: 410–465.
- Burrell, Gibson & Morgan, Garreth (1979), *Sociological Paradigms and Organisational Analysis*, Gower, Aldershot.
- Carroll, John W. (2010 [2003]), “Laws of Nature”, in *Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, <http://plato.stanford.edu/entries/laws-of-nature/>.
- Chappell, Vere (1999), “Introduction”, in *Hobbes and Bramhall on Liberty and Necessity*, ed. Vere Chappell, Cambridge University Press, Cambridge: ix–xxiii.
- Charles, David (1984), *Aristotle’s Philosophy of Action*, Duckworth, London.
- Chisholm, Roderick M. (1989), *On Metaphysics*, University of Minnesota Press, Minneapolis, MN.
- Coleman, James S. (1986), “Social Theory, Social Research, and a Theory of Action”, *American Journal of Sociology* 91: 1309–1335.
- Cook, Margaret G. (2001), “Divine Artifice and Natural Mechanism: Robert Boyle’s Mechanical Philosophy of Nature”, *Osiris* 16: 133–150.
- Creath, Richard (2011), “Logical Empiricism”, in *Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, <http://plato.stanford.edu/entries/logical-empiricism/>.
- Dahl, Norman O. (1984), *Practical Reason, Aristotle, and Weakness of the Will*, University of Minnesota Press, Minneapolis, MN.
- Dancy, Jonathan (2000), *Practical Reality*, Oxford University Press, Oxford.
- Davidson, Donald (1963), “Actions, Reasons, and Causes”, *The Journal of Philosophy* 60 (23): 685–700.

- Davidson, Donald (1980 [1978]), “Intending”, in *Essays on Actions and Events*, Clarendon Press, Oxford: 83–102.
- Davidson, Donald (2006 [1982]), “Paradoxes of Irrationality”, in *The Essential Davidson*, Clarendon Press, Oxford: 138–152.
- Davidson, Donald (2001 [1975]), “Thought and Talk”, in *Inquiries into Truth and Interpretation*, Clarendon Press, Oxford: 155–171.
- Demeulenaere, Pierre (2011a), “Causal Regularities, Action and Explanation”, in *Analytic Sociology and Social Mechanisms*, ed. Pierre Demeulenaere, Cambridge University Press, Cambridge: 173–197.
- Demeulenaere, Pierre (2011b), “Introduction”, in *Analytic Sociology and Social Mechanisms*, ed. Pierre Demeulenaere, Cambridge University Press, Cambridge: 1–30.
- Dennett, Daniel C. (1986 [1969]), *Content and Consciousness*, Routledge & Kegan Paul, London.
- Dennett, Daniel C. (1987), *Intentional Stance*, The MIT Press, Cambridge, MA.
- Descartes, René (1984), *Meditations on First Philosophy*, trans. John Cottingham, in *The Philosophical Writings of Descartes, Volume II (CSM II)*, ed. John Cottingham, Robert Stoothoff & Dugald Murdoch, Cambridge University Press, New York, NY: 1–62. Originally published in 1641.
- Descartes, René (1985), *Principles of Philosophy*, trans. John Cottingham, in *The Philosophical Writings of Descartes, Volume I (CSM I)*, ed. John Cottingham, Robert Stoothoff & Dugald Murdoch, Cambridge University Press, New York, NY: 177–291. Originally published in 1644.
- Descartes, René (1991), *The Correspondence*, in *The Philosophical Writings of Descartes, Volume III (CSMK III)*, trans., ed. John Cottingham, Robert Stoothoff, Dugald Murdoch & Anthony Kenny, Cambridge University Press, New York, NY.
- Descartes, René (1985), *The Passions of the Soul*, trans. Robert Stoothoff, in *The Philosophical Writings of Descartes, Volume I (CSM I)*, ed. John Cottingham, Robert Stoothoff & Dugald Murdoch, Cambridge University Press, New York, NY: 325–404. Originally published in 1649.

- Dihle, Albrecht (1982), *The Theory of Will in Classical Antiquity*, University of California Press, Los Angeles.
- Downing, Lisa (1998), “The Status of Mechanism in Locke’s Essay”, *The Philosophical Review* 107 (3): 381–414.
- Ducasse, C. J. (1926), “On the Nature and the Observability of Causal Relation”, *Journal of Philosophy* 23: 57–68.
- Elster, Jon (1989), *Nuts and Bolts for the Social Sciences*, Cambridge University Press, New York, NY.
- Giddens, Anthony (1993 [1976]), *New Rules of Sociological Method*, Polity Press, Padstow, Cornwall.
- Gilbert, Margaret (1996), *Living Together: Rationality, Sociality, and Obligation*, Rowman & Littlefield Publishers, Lanham, MD.
- Gilbert, Margaret (1989), *On Social Facts*, Routledge, London.
- Gilbert, Margaret P. (1990), “Walking Together: A Paradigmatic Social Phenomenon”, *Midwest Studies in Philosophy* 15: 1–14.
- Glennan, Stuart (2008), “Mechanisms”, in *The Routledge Companion to Philosophy of Science*, ed. Stathis Psillos & Martin Curd, Routledge, London: 376–384.
- Glennan, Stuart (2002), “Rethinking Mechanistic Explanation”, *Philosophy of Science* 69 (Supplement): 342–353.
- Grice, Paul (2001), *Aspects of Reason*, ed. Richard Warner, Clarendon Press, Oxford.
- Hall, A. Rupert (1954), *The Scientific Revolution: 1500–1800*, Longmans, London.
- Hampshire, Stuart (1959), *Thought and Action*, Chatto & Windus, London.
- Harman, Gilbert (1976), “Practical Reasoning”, *Review of Metaphysics* 29: 431–463.
- Hedström, Peter & Swedberg, Richard (1998), “Social Mechanisms: An Introductory Essay”, in *Social Mechanisms: An Analytical Approach to Social Sciences*, ed. Peter Hedström & Richard Swedberg, Cambridge University Press, New York, NY: 1–31.
- Hedström, Peter & Udehn, Lars (2009), “Analytical Sociology and Theories of the Middle Range”, in *The Oxford Handbook of Analytical Sociology*, ed. Peter Hedström & Peter Bearman, Oxford University Press, Oxford: 25–47.

- Hedström, Peter & Ylikoski, Petri (2010), “Causal Mechanisms in the Social Sciences”, *Annual Review of Sociology* 36: 49–67.
- Hempel, Carl G. (1965), “Aspects of Scientific Explanation”, in *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, The Free Press, New York, NY.
- Hempel, Carl G. & Oppenheim, Paul (1948), “Studies in the Logic of Explanation”, *Philosophy of Science* 15: 137–175.
- Hitchcock, Christopher Read (1996), “The Role of Contrast in Causal and Explanatory Claims”, *Synthese* 107: 395–419.
- Hobbes, Thomas (1962, [1914]), *Leviathan*, J. M. Dent & Sons, London. Originally published in 1651.
- Hobbes, Thomas & Bramhall, John (1999), *Hobbes and Bramhall on Liberty and Necessity*, ed. Vere Chappell, Cambridge University Press, Cambridge.
- Hume, David (2000), *A Treatise of Human Nature (T)*, ed. David Fate Norton & Mary J. Norton, Oxford University Press, Oxford. Originally published in 1739–1740.
- Hume, David (1975), *An Enquiry concerning Human Understanding (EHU)*, in *Enquiries*, ed. P. H. Nidditch, Clarendon Press, Oxford. Originally published in 1777.
- Kakkuri-Knuuttila, Marja-Liisa (2006), “Kausaalisuhteet ja selittäminen tulkitsevassa tutkimuksessa [Causal Relations and Explaining in Interpretive Research]”, in *Soveltava yhteiskuntatiede ja filosofia [Applied Social Sciences and Philosophy]*, ed. Kristina Rolin, Marja-Liisa Kakkuri-Knuuttila & Elina Henttonen, Gaudeamus, Helsinki: 54–87.
- Kant, Immanuel (1998), *Critique of Pure Reason (CPR)*, ed., trans. Paul Guyer & Allen W. Wood, Cambridge University Press, Cambridge. Originally published in 1781.
- Kant, Immanuel (1996), *Groundwork of the Metaphysics of Morals (Groundwork)*, trans. Mary J. Gregor, in *Practical Philosophy*, ed. Mary J. Gregor, Cambridge University Press, Cambridge: 37–108. Originally published in 1797.
- Kant, Immanuel (2007), *Idea for a Universal History with a Cosmopolitan View (Idea)*, trans. Allen W. Wood, in *Anthropology, History, and Education*, ed. Günter Zöllner &

- Robert B. Loudon, Cambridge University Press, Cambridge: 107–120. Originally published in 1784.
- Kant, Immanuel (2002), *Prolegomena to Any Future Metaphysics That Will Be Able to Come Forward as Science (Prolegomena)*, trans. Gary Hatfield, in *Theoretical Philosophy after 1781*, ed. Henry Allison & Peter Heath, Cambridge University Press, Cambridge: 29–169. Originally published in 1783.
- Kant, Immanuel (1996), *Religion within the Boundaries of mere Reason (Religion)*, trans. George di Giovanni, in *Religion and Rational Theory*, ed. Allen W. Wood & George di Giovanni, Cambridge University Press, Cambridge: 39–215. Originally published in 1793.
- Kenny, Anthony (1963), *Action, Emotion and Will*, Routledge & Kegan Paul, London.
- Kenny, Anthony (1979), *Aristotle's Theory of the Will*, Duckworth, London.
- Kenny, Anthony (1975), *Will, Freedom and Power*, Basil Blackwell, Oxford.
- Kim, Jaegwon (1973a), “Causation, Nomic Subsumption, and the Concept of Event”, *The Journal of Philosophy* 70: 217–236.
- Kim, Jaegwon (1973b), “Causes and Counterfactuals”, *The Journal of Philosophy* 70: 570–572.
- Kim, Jaegwon (1993 [1976]), “Events as Property Exemplifications”, in *Supervenience and Mind: Selected Philosophical Essays*, Cambridge University Press, New York, NY: 33–52.
- Kim, Jaegwon (1989), “Mechanism, Purpose, and Explanatory Exclusion”, *Philosophical Perspectives* 3: 77–108.
- Kitcher, Patricia (2006), “Kant's Philosophy of the Cognitive Mind”, in *The Cambridge Companion to Kant and Modern Philosophy*, ed. Paul Guyer, Cambridge University Press, New York, NY: 169–202.
- Kuorikoski, Jaakko (2009), “Two Concepts of Mechanism: Componential Causal System and Abstract Form of Interaction”, *International Studies in the Philosophy of Science* 23 (2): 143–160.
- Lagerspetz, Eerik (2006), “Ennusteiden ja todellisuuden vuorovaikutus yhteiskuntatieteissä [The Interaction between Predictions and Reality in Social Sciences]”, in *Soveltava*

- yhteiskuntatiede ja filosofia* [Applied Social Sciences and Philosophy], ed. Kristina Rolin, Marja-Liisa Kakkuri-Knuutila & Elina Henttonen, Gaudeamus, Helsinki: 36–53.
- Lange, Mark (2008), “Laws of Nature”, in *The Routledge Companion to Philosophy of Science*, ed. Stathis Psillos & Martin Curd, Routledge, London: 203–212.
- Lewis, David (1973), “Causation”, *Journal of Philosophy* 70: 556–557.
- Locke, John (1975), *An Essay concerning Human Understanding*, ed. Peter H. Nidditch, Clarendon Press, Oxford. Originally published in 1690.
- Machamer, Peter; Darden, Lindley & Craver, Carl F. (2000), “Thinking about Mechanisms”, *Philosophy of Science* 67 (1): 1–25.
- Malcolm, Norman (1968), “The Conceivability of Mechanism”, *The Philosophical Review* 77 (1): 45–72.
- McCann, Hugh J. (1998), “The Formation of Intention”, in *The Works of Agency: On Human Action, Will, and Freedom*, Cornell University Press, Ithaca, NY: 147–169.
- Melden, A. I. (1961), *Free Action*, Routledge & Kegan Paul, London.
- Mele, Alfred R. (1997), “Introduction”, in *The Philosophy of Action*, ed. Alfred R. Mele, Oxford University Press, Oxford: 1–26.
- Mele, Alfred R. & Moser, Paul K. (1994), “Intentional Action”, *Noûs* 28 (1): 39–68.
- Merton, Robert K. (1949), *Social Theory and Social Structure: Toward the Codification of Theory and Research*, The Free Press, Glencoe, IL.
- Merton, Robert K. (1936), “The Unanticipated Consequences of Purposive Social Action”, *American Sociological Review* 1 (6): 894–904.
- Miller, Seumas (2011 [2007]), “Social Institutions”, in *Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, <http://plato.stanford.edu/entries/social-institutions/>.
- Montague, Richard (1974 [1969]), “On the Nature of Certain Philosophical Entities”, in *Formal Philosophy: Selected Papers of Richard Montague*, ed. Richmond H. Thomson, Yale University Press, New Haven, CT: 148–187.
- Nagel, Thomas (1986), *The View from Nowhere*, Oxford University Press, New York, NY.



- Opp, Karl-Dieter (2005), "Explanations by Mechanisms in the Social Sciences. Problems, Advantages and Alternatives", *Mind & Society* 4: 136–178.
- Parfit, Derek (1987), "Divided Minds and the Nature of Persons", in *Mindwaves*, ed. Colin Blakemore & Susan Greenfield, Basil Blackwell, London: 19–26.
- Pereboom, Derk (2001), *Living without Free Will*, Cambridge University Press, Cambridge.
- Pianesi, Fabio & Varzi, Achille C. (2000), "Events and Events Talk: An Introduction", in *Speaking of Events*, ed. James Higginbotham, Fabio Pianesi & Achille C. Varzi, Oxford University Press, New York, NY: 3–47.
- Plato (1997), *Protagoras*, trans. Stanley Lombardo & Karen Bell, in *Complete Works*, ed. John M. Cooper, Hackett Publishing Company, Indianapolis, IN: 746–790.
- Plato (1997), *The Republic*, trans. G. M. A. Grube, rev. C. D. C. Reeve, in *Complete Works*, ed. John M. Cooper, Hackett Publishing Company, Indianapolis, IN: 971–1223.
- Reid, Thomas (1969), *Essays on the Active Powers of the Human Mind (EAP)*, The M. I. T. Press, London. Originally published in 1785.
- Reid, Thomas (1969), *Essays on the Intellectual Powers of the Man (EIP)*, The M. I. T. Press, London. Originally published in 1788.
- Ryle, Gilbert (1949), *The Concept of Mind*, Hutchinson & CO, London.
- Salmon, Wesley C. (1989), *Four Decades of Scientific Explanation*, University of Minnesota Press, Minneapolis, MN.
- Salmon, Wesley C. (1984), *Scientific Explanation and the Causal Structure of the World*, Princeton University Press, Princeton, NJ.
- Schueler, G. F. (2003), *Reasons and Purposes: Human Rationality and the Teleological Explanation of Action*, Clarendon Press, Oxford.
- Searle, John (1990), "Collective Intentions and Actions", in *Intentions in Communication*, ed. Philip R. Cohen, Jerry Morgan & Martha E. Pollack, The MIT Press, Cambridge, MA: 401–415.
- Sluga, Hans (1996), "'Whose House Is That?' -Wittgenstein on the Self", in *The Cambridge Companion to Wittgenstein*, ed. Hans Sluga & David G. Stern, Cambridge University Press, Cambridge: 320–353.

- Spinoza, Baruch (2002), *Ethics (E)*, trans. Samuel Shirley, in *Complete Works*, ed. Michael L. Morgan, Hackett Publishing Company, Indianapolis, IN: 213–382. Originally published in 1677.
- Steward, Helen (2009), “Sub-intentional Actions and the Over-mentalization of Agency”, in *New Essays on the Explanation of Action*, ed. Constantine Sandis, Palgrave MacMillan, New York, NY: 295–313.
- Stinchcombe, Arthur L. (1991), “The Conditions of Fruitfulness of Theorizing About Mechanisms in Social Science”, *Philosophy of the Social Sciences* 21: 367–388.
- Stocker, Michael, (1979), “Desiring the Bad: An Essay in Moral Psychology”, *The Journal of Philosophy*, 76 (12): 738–753.
- Thalberg, Irving (1972), “Can One Intend the Impossible”, in *Enigmas of Agency: Studies in the Philosophy of Human Action*, Georg Allen & Unwin, London: 105–114.
- Thomas Aquinas (1948–1949), *Summa Theologiae (ST)*, trans. Fathers of the English Dominican Province, Benziger Bros, New York, NY.
- Thornton, M. T. (1982), “Aristotelian Practical Reason”, *Mind* 91: 57–76.
- Tolstoy, Leo (1952), *War and Peace*, trans. Louise and Aylmer Maude, Encyclopædia Britannica, Chicago, IL. Originally published in 1869.
- Tuomela, Raimo (1984), *A Theory of Social Action*, D. Reidel Publishing Company, Dordrecht.
- van Fraassen, Bas C. (1980), *Scientific Image*, Clarendon Press, Oxford.
- Velleman, J. David (1992), “The Guise of the Good”, *Noûs* 26 (1): 3–26.
- Velleman, J. David (1989), *Practical Reflection*, Princeton University Press, Princeton, NJ.
- von Wright, Georg Henrik (1971), *Explanation and Understanding*, Cornell University Press, Ithaca, NY.
- Watson, Gary (2004), “The Work of the Will”, in *Agency and Answerability*, Oxford University Press, New York, NY: 123–157.
- Westfall, Richard S. (2007 [1971]), *The Construction of Modern Science: Mechanisms and Mechanics*, Cambridge University Press, New York, NY.

- White, Harrison C. (1970), *Chains of Opportunity: System Models of Mobility in Organizations*, Harvard University Press, Cambridge, MA.
- Winch, Peter (1958), *The Idea of Social Science and Its Relation to Philosophy*, Routledge & Kegan Paul, London.
- Wittgenstein, Ludwig (2003 [1997]), *Movement of Thoughts: Diaries 1930–1932, 1936–1937*, in *Public and Private Occasions*, ed. James C. Klagge & Alfred Nordman, Rowman & Littlefield, Lanham: 3–256
- Wittgenstein, Ludwig (1997 [1953]), *Philosophical Investigations (PI)*, 2<sup>nd</sup> ed. trans. G. E. M. Anscombe, Blackwell, Oxford.
- Wittgenstein, Ludwig (1961), *Tractatus Logico-Philosophicus (TLP)*, trans. D. F. Pears & B. F. McGuinness, Routledge & Kegan Paul, London. Originally published in 1921.
- Woodward, James (2009 [2003]), “Scientific Explanation”, in *Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, <http://plato.stanford.edu/entries/scientific-explanation/>.
- Wright, Larry (1976), *Teleological Explanation: An Etiological Analysis of Goal and Functions*, University of California Press, Berkeley, CA.
- Ylikoski, Petri (2011), “Social Mechanisms and Explanatory Relevance”, in *Analytic Sociology and Social Mechanisms*, ed. Pierre Demeulenaere, Cambridge University Press, Cambridge: 154–172.