Siri Granum Carson

The Fact of Reason

A linguistic-pragmatic approach to the Free Agency Problem

Thesis for the degree of Philosophiae Doctor

Trondheim, December 2008

Norwegian University of Science and Technology
Faculty of Arts
Department of Philosophy

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the degree of Philosophiae Doctor

Faculty of Arts
Department of Philosophy

ISBN 978-82-471-1347-9 (printed ver.)
ISBN 978-82-471-1348-6 (electronic ver.)
ISSN 1503-8181

Doctoral theses at NTNU, 2008:320

Printed by NTNU-trykk

To my daughters Selma and Lisa
– free spirits and precious little pieces of nature.

Contents

Acknowledgements

First and foremost I would like to thank my supervisor, Professor Audun Øfsti, an inexhaustible source of insight, inspiration, and infectious enthusiasm.

I would like to thank the Faculty of Art at NTNU for financing this project.

I would also like to thank Geert Keil for good advice on an early stage of my work with the thesis. Further, I thank Siv Dokmo, Silje Langvatn, Erling Skjei, Bjørn Myskja, Truls Wyller, Ronny Myhre, Kjetil Audsen, Kjartan Mikalsen, and Jørgen Dyrstad for reading drafts and providing me with useful comments, critical remarks, and troublesome counterarguments along the way.

I am grateful for all cheer from my colleagues at the Department of Philosophy. Specifically I thank May Thorseth and Rune Nydal at the Programme for Applied Ethics for giving me the opportunities that you have and for being so delightful to work with.

I am greatly indebted to my parents and to my parents-in-law – I would not have been able to complete this task without you.

Finally, thank you Christian, for all your love and support.

Think (think, think), think about what you are trying to do to me
Yeah, think (think, think), let your mind go, let yourself be free
You need me (need me) and I need you (don't you know)
Without each other there ain't nothing people can do
Oh freedom (freedom), freedom (freedom), freedom, yeah freedom
Freedom (freedom), freedom (freedom), freedom, ooh freedom

(Aretha Franklin)

Chapter 1: Introduction to the free agency problem

'Free will' is the conventional name of a topic that is
best discussed without reference to the will.
(G. Strawson 1998/2004, p. 1)

[F]reiheit charakterisiert eine Seinsweise – die Art wie
handelnde Personen im Raum der Gründe existieren.
(Habermas 2006, p. 675)

1.1 The Free Will Problem

In what sense is this thesis an investigation into the concept of freedom? Let me start by defining what this thesis is *not* about. First of all, it is not an investigation of freedom in the social or political sense: I will not discuss concepts such as democracy, civil liberties, or individual rights. I am concerned with what is commonly known as *the free will problem*. More than being one particular question, this is rather a cluster of related philosophical puzzles that relate to our everyday conviction that we act according to our own freely taken decisions, and regarding what kind of status this conviction has in relation to a world of causally related events.

Both the volume of literature and the "thought density" surrounding this area are enormous and ever-increasing. It seems to be as popular a topic as ever, among philosophers as well as in other academic areas, notably the neurosciences. Thus, it might be too much to expect that I will go so far as to think any truly original thoughts. The best I can hope for is perhaps to combine what I consider are the best available thoughts on the subject in an original manner, and in doing so to shed some new light on this continually perplexing problem.

One of the first moves I would like to make to distinguish myself from the mainstream philosophical debate is to refrain from using the established *free will* terminology. I consider 'free will' to be a somewhat artificial construct derived from the fundamental question of what it means to *act* freely. What I find problematic about this construct is that it creates a false image of 'the will' as a mental unit over and above our ability to act, to which we may – or may not – ascribe freedom. In contrast to this, I would like to emphasise *free agency* as characteristic for a mode of being, viz. an ability to make things happen in the world by acting according to reason. This is what I will later refer to as *performative being*.[1]

Central to the concept of free agency is that it warrants the assigning of *moral responsibility* to agents. As among others Locke and Kant have argued, being free in the morally relevant sense has nothing to do with acting according to one's changing inclinations and wishes. On the contrary, people who behave in such a way appear to be paradigmatically unfree. The word 'will' has 'want' as one of its basic meanings, whereas being free frequently shows itself in an ability to *not* do what I want, in the sense of postponing my immediate needs and acting according to (better) *reason*. According to Schopenhauer, I can *do* what I want, but it does not make sense to say that I can *want* what I want; at any given time I can only want what I in fact want.[2] This is a truth with a modification, however, since the idea of choosing what to want has a clear and reasonable meaning, notably regarding so-called "higher-order wants". A commonly used example is that of an alcoholic who may choose to disregard that she wants a drink in favour of her higher-order want to stay sober.

A traditional view is that free agency is a weaker concept than free will, so that the defence of free agency does not amount to a robust concept of human freedom. According to such a view, acting freely simply means being *unrestrained* in a specific sense – whether from outer/physical or inner/mental forces or obstructions. Thus, if I am locked up in chains, my freedom of agency is severely diminished, whereas my will is as free as ever. Mental restrictions are more difficult. Addiction, obsession, agitation, extortion, ignorance – these are all factors affecting the rationality of my choices as well as the degree to which I am held responsible for my actions. An area of contention is whether these factors affect the freedom of my will or just diminish my range of free agency.

To me, the distinction between a "strong" concept of free will and a "weaker" concept of free agency makes little sense. "The will" cannot be singled out as a particular unit in me

[1] Cf. below, Ch. 5.9. I borrow the concept "performative being" from Albrecht Wellmer (cf. Wellmer 1991, p. 183f.)
[2] Cf. Schopenhauer 1839, p. 542.

over and above my ability to form intentions and act according to these. And my ability to form intentions and act according to these is synonymous with my ability to act freely. Deciding how to relate to one's various inclinations, forming intentions and so on are, as I see it, exactly the kind of rational processes that we *carry out*, and that we call actions. Therefore, I chose to talk about the *free agency problem.*

"Building the will" is in itself a case of intentional agency (in the case of prior intentions) or a part of an action (in the case of intentions-in-actions).[3] In this, I am always subjected to various physical *and* mental restraints and influences, and my freedom never amounts to more than a freedom to make decisions and act *under given circumstances*. "Free will", viewed as the ability to make decisions under prevailing circumstances, cannot be analysed independently from our ability to make things happen in the world. Seen as a purely mental concept, the will can only be free in the negative sense of being unrestrained, not in the positive sense of acting according to reason and thus being responsible. What we are held responsible for are ultimately our *actions* and – among them – our deliberations, choices and decision-making processes. To turn the stronger-weaker distinction around, we could argue that a free will is not enough, but that it must be seen as a derivative component of our ability to act freely.

1.2 The Kantian Problem

The appeal of the philosophical problem area referred to above as "the free will problem" – and that I suggest be renamed "the free agency problem" – comes from the idea of a *conflict* between the freedom we subjectively experience and the world that we experience as functioning according to the laws of nature. Kant demonstrated an acute sense of this conflict in the *Critique of Pure Reasons*. He formulated the conflict as an *antinomy*, i.e. as two contradicting principles, both of which concern reality as we experience it. Thus, Kant points out that we are dealing with a major contradiction in our lives. We may try to dissolve the sheer contradiction – as Kant certainly does – but since the conflict between a subjective and an objective view of ourselves constitutes an important aspect of what it means to be a rational being, it will remain an area of tension in our lives.

[3] On the distinction between prior intentions and intentions-in-action, cf. e.g. Searle 2001, p. 44f.

Albrecht Wellmer argues that a characteristic mark of philosophical activity is the clarification of concepts through the display of inherent contradictions and incoherence. This is not simply restricted to pure logical contradictions, but includes various forms of practically significant oppositions, dilemmas, and puzzles. The persuasive force of a philosophical position shows itself not least in its productive handling of such problems. As a classic example, Wellmer points to Kant's antinomy chapter in the Transcendental Dialectics:

> [B]ei der dritten Antinomie etwa geht es – so interpretiere ich sie – um einen transzendental begründeten Widerstreit zwischen einer 'objektivierenden' (naturalistischen) und einer 'performativen' (normativen) Perspektive auf die geschichtliche Welt und auf uns selbst als Handelnde, einen Widerstreit, der sich zwar als ein Wider*spruch* auflösen, aber als ein praktisch und existentiell bedeutsamer und unvermeidlicher, sowohl in der Lebenswelt als auch in der Praxis der Sozialwissenschaften und des Rechts immer wieder sich meldender Wider*streit* deshalb nicht beseitigen lässt, weil beide Perspektiven immer wieder ihr Recht *gegeneinander* geltend machen, ohne dass sie sich – so scheint es – ohne weiteres in einer umfassenden Perspektive friedlich miteinander vereinigen ließen. (Wellmer 2007, p. 230)

Kant's solution is in a specific sense compatibilistic in that it confirms the possibility of both freedom and determinism. However, as opposed to classic compatibilism à la Hume, he does not refute the contradiction by writing it off as merely apparent. The Kantian formulation of the problem points to a fundamental tension in our lives, due to the opposition between our different perspectives on and accesses to the world. In this way, Kant sheds light on the normative dimension of the freedom problem. The important aspect of the question of freedom is not the possibility of having no strings attached, but the ability to be bound by the right kind of strings, namely *reasons* by which we as autonomous beings bind *ourselves*. In the following, I will argue that to perform an action basically is to enter into a web of commitments, into what Robert Brandom (referring to Wilfred Sellars) calls a "normative space of reasons".[4] This irreducibly normative dimension is the fundament for the ascription of moral responsibility and for the characterisation of actions as free.

The Kantian way of displaying the freedom problem remains of current interest. However, his solution is – as I will argue – dualistic in its representation of man as "citizen of two worlds". One advantage with moving from a "free will problem" to a "free agency problem" is that the concept of rational agency is anti-dualistic at its core, in the sense that an action has both mental and physical properties. Furthermore, Kant does not bring up language

[4] Cf. Brandom 1994 and Sellars 1997. Cf. below, Ch. 4.

as a precondition for rationality, thus of rational agency. In my thesis, I will argue that rational agency must in a specific sense be explained on the basis of communicative action. An adequate approach to human freedom must recognise intersubjective language as a condition for the possibility of free agency.

1.3 The Perspectivity Problem

At the core of the free agency problem lays the (intuitively tempting) idea that an objectivating view of the world can be made absolute or understood as overriding, and that our different subjective, partial views of the world may be regarded as subordinated to an "objective, universal truth". According to Thomas Nagel, these ideas inevitably lead to paradoxes concerning our view of ourselves:

> We can act only from inside the world, but when we see ourselves from outside, the autonomy we experience from inside appears as an illusion, and we who are looking from outside, cannot act at all. (Nagel 1986, p. 120)

Nagel draws the rather pessimistic conclusion that there can be no solving of this paradox, hence no solution to the problem of human freedom.

Nagel distinguishes two aspects of the free agency problem; one having to do with our own freedom (the problem of autonomy) and one with the freedom of others (the problem of responsibility). Both problems boil down to the opposition between an internal and an external view on action, whether our own or the actions of others. Nagel sees the possibility of holding other people responsible as one that is available from within our internal perspective, but which we lose as soon as we view the actions of others "from outside". Initially, though, it seems mysterious that we are able to assume an internal perspective concerning the actions of other people. Nagel seems to think that we infer by analogy:

> In acting we occupy the internal perspective, and we can occupy it *sympathetically* with regard to the actions of others. (Nagel 1986, p. 113, my italics)

I think Wittgenstein's critique of "the realist" applies to this formulation: It makes no sense to *believe* that other people have an "inner life" – in the sense of having it as a hypothesis – since

no conceivable experience would support or weaken this belief.[5] However, if we regard intersubjective language as a condition of both our understanding of ourselves *and* of others as autonomous and responsible, the need for an inference by analogy disappears. Therefore, I propose to substitute the mentalistic distinction between an internal and an external perspective with a linguistic-pragmatic distinction between the expressive and the assertoric mode, or between the performative and the propositional part of speech acts in the Austin-Searle-Habermasian sense.[6]

According to Nagel, the reason we do not hold animals responsible is that we cannot assume their point of view. This might seem to be in line with Wittgenstein's supposition that even if lions *could* speak, we would not be able to understand them.[7] And we are certainly most inclined to judge those who are most like ourselves – as we correspondingly feel entitled to be judged by a "jury of equals". However, the analogy approach can hardly capture the true sense of this expression, since "equal" here means *morally* equal; not actually *alike*, but with the same rights and duties. And the analogy fails utterly when it comes to small children. Nagel seems to suggest that we refrain from judging them because we do not understand them:

> With regard to small children the possibilities of moral judgment are somewhat greater, but we still cannot project ourselves fully into their point of view in order to think about what they should do. (Nagel 1986, p. 121)

Nagel thinks the same goes for e.g. mentally disturbed people or people under the influence of drugs: We cannot judge them, because they are "different from us".[8] However, when we – as Strawson puts it – "suspend our resentment"[9] towards the actions of small children and the mentally ill, this is not because we do not understand them or because we regard them as too much unlike us. It is because we truly do not *hold them* responsible, since their rationality is (as yet – or temporarily) underdeveloped – or underachieving.

Obviously, Nagel has a sound point. Holding each other responsible for our actions depends on our ability to assume the other party's point of view. However, he seems to put the carriage before the horse. It is not because the animal, child or drug addict is too different from us that we cannot assume their point of view, but because they cannot assume the role of

[5] Cf. Wittgenstein 1958, p. 48f.
[6] Cf. below, Ch. 6.3.
[7] Cf. Wittgenstein PU, p. 190.
[8] Cf. Op.cit. p. 122.
[9] Cf. Strawson 1974, p. 77.

full-fledged second persons for us.[10] Holding someone responsible essentially presupposes *co-subjectivity*, and hence *reciprocity*. It is not because we see a resemblance between us and other individuals that we – by analogy or sympathy – hold them responsible. Rather, it is by holding each other responsible that we recognise each other as caught up in the same "web of commitments", and come to see each other as co-subjects. We recognise each other *as* recognisers. Strictly speaking, we can only hold those responsible who in turn are able to hold *us* responsible.

Nagel correctly pins down the free agency problem as one having to do with the changing perspectives we assume towards actions. However, in the following I shall attempt to substitute Nagel's "mentalist" internal-external distinction with a more refined linguistic-pragmatic distinction between first-, second-, and (relative vs. absolute) third-person perspective, connected to the relationship between performative and propositional parts of speech acts. In this way, I hope to attain a more flexible conception of free agency, a conception allowing for transitions between the different perspectives that we may assume towards actions, thereby retaining a monistic concept of action.[11] These changing perspectives are bound together through a system of indexical substitutions. Even objective propositions that render perspective-transcending truths about the world have a perspectival basis.[12]

Nagel observes that the free agency problem arises when we push the objectivation of actions too far. In this sense, the problem is independent of determinism: Our actions seem no more free if regarded as the result of "indeterminist" quantum leaps than they do if regarded as conforming to strict causal laws of nature:

> In either case we cease to face the world and instead become parts of it. (Nagel 1986, p. 114)

I will argue that the problem with many attempts to dissolve or untangle the free agency problem (including Nagel's suggestion that "nothing approaching the truth has yet been said

[10] Cf. below, Ch. 5.4.
[11] Which is what Nagel seeks as well, cf. Nagel 1986, p. 111. I think, however, that Nagel's inner-outer metaphors regarding actions point to a wide spread challenge for the philosophical debate on the free agency problem. Metaphors are not harmless, i.e. not just illustrations of ready-made thoughts; they actually shape the way we think (cf. e.g. Riceur 1977). The inner/outer dichotomy belongs to a set of frequently used metaphors in philosophy (cf. my critique of Sebastian Rödl and Jürgen Habermas below, Ch. 4.4; Ch. 6.9). In my thesis I criticise this metaphor in general for suggesting a too strict dichotomy between a normative I-you-communication and a descriptive third-person view of the world, and for lacking the flexibility necessary in order to perceive free actions as parts of the world, and to speak about other persons as free agents.
[12] Cf. Tugendhat 1976; cf. below, Ch. 6.4.

about this subject"[13]) is that they assume a too-strict dichotomy between the subjective and the objective, between the "internal" and the "external" perspective, or between (normative) participation and (scientistic) observation. An acceptable approach to the free agency problem should ensure that we are able to view subjects facing the world *as* parts of it.

1.4 Free Agency: A fact of reason

Thomas Nagel points out that the problem with a deterministic view of agency applies generally to all attempts to objectivate action. The problem is certainly one of perspective, but it cannot be solved simply by insisting that action must be viewed from an "inner" perspective. A purely "internal" view of agency would make it impossible to see free actions as parts of the world, and would thus lead either to dualism or to an insolvable dilemma à la Nagel.

I think Nagel is right when he argues that our sense of being "the authors of our own actions" cannot really be seen as an intelligible *belief*, but something that somehow *shows itself*. [14] To put it in Kantian terms, free agency should be recognised as a "fact of reason":

> Freiheit ist (…) die einzige unter allen Ideen der spekulativen Vernunft, wovon wir die Möglichkeit a priori wissen, ohne sie doch einzusehen, weil sie die Bedingung des moralischen Gesetzes ist, welches wir wissen. (Kant, KpV 5)[15]

Kant's point can be generalised. Free agency is not only a condition for realising the rationality of the categorical imperative, but a general condition of rationality, i.e. of being able to relate to reasons. Thus, free agency is not something that we can discover with the help of reason, but what *constitutes* reason.

A fundamental element in the approach to free agency proposed in this thesis is an interventionist theory of causality. Von Wright and others have argued that our active interference in the natural course of events is what makes us capable of having a concept of

[13] Cf. Nagel 1986, p. 137.
[14] Cf. Nagel 1986, p. 114.
[15] In a footnote, Kant adds that the relationship between freedom and the moral law can be specified further by seeing freedom as the *ratio essendi* of the moral law and the moral law as the *ratio cognoscendi* of freedom: "Denn, wäre nicht das moralische Gesetz in unserer Vernunft eher deutlich gedacht, so würden wir uns niemals berechtigt halten, so etwas, als Freiheit ist (ob diese gleich sich nicht widerspricht), anzunehmen. Wäre aber keine Freiheit, so würde das moralische Gesetz in uns gar nicht anzutreffen sein" (Loc.cit, fn. 1). More on this below, Ch. 5.5.

causal relations as existing in the world.[16] Action, not causality, is immediately given – although the two refer reciprocally to each other. Von Wright further argues in a Kantian manner that, once the fact of (rational) action is a given, we do not need an additional account of freedom. Freedom is already implied in the concept of action:

> [T]he concept of an action, the ascriptions of actions to an agent, belong to discourse in which 'free will' is taken for granted. (von Wright 1980, p. 78)

Geert Keil formulates this thought by saying that the language of agency already implies a "massive metaphysics of freedom"[17] Viewed in this way, our indomitable experience of being "the authors of our own actions" becomes the first step in the solution of the free agency problem instead of the mysterious 'x' to be explained.

In keeping with this, I will not attempt to argue that the assumption of freedom is true, but rather that it is indispensable. In the following chapters, I will argue that reasoning – even about causal connections in the world – ultimately involves an implicit assumption by the rational subject of herself as a freely acting person. Hence, any attempt to refute free agency will inevitably run into pragmatic self-contradictions or performative inconsistencies.

I will further argue that the *indexicality* of colloquial language, i.e. the possibility for *perspectival multiplicity and change* is a key to resolving the free agency problem. Instead of denying the possibility of objectivating action, I will argue that objectivated action can be seen as a bridge between freedom and nature in Kant's sense – given a non-scientistic objectivation.[18] A complete language equips us with the ability, not only to *perform*, but also to *speak about* free actions.[19] This provides the desired transitions between "internal" and "external" in Nagel's sense, between the performative and the propositional parts of speech acts, and between the normative and the descriptive – since objectivated actions essentially belong on both sides of the "gulf".

However, as Wellmer argues, although we may manage to dissolve the pure, logical contradiction between freedom and nature, the sense of a practical paradox will remain. The conflict will reappear every time we attempt to view ourselves "from the outside". The inner tension within our colloquial language between the performative and the propositional parts

[16] Cf. below, Ch. 3.3, 3.4.
[17] Cf. Keil 2000, p. 12; 2007, p. 89.
[18] Cf. Apel's concept of "secondary objectivation" (Apel 1979, p. 173), cf. below, Ch. 6.7.
[19] In the "relative third person", cf. below, Ch. 6.4; 6.5.

of our speech acts corresponds, it seems, to a fundamental contradiction within human existence.

1.5 Structure of the thesis

Ultimately, my formulation of the free agency problem and the main question of this thesis is this: How can we – practically and epistemologically – specify the ability that rational beings have to act freely, in the sense of making things happen in the world, as well as in the sense of holding each other responsible?[20]

In the next chapter, **Chapter 2: Kantian freedom – between mind and world**, I will attempt to analyse the Kantian approach to free will in order to review the strengths as well as the weaknesses of this theory. I start out with an analysis of the third antinomy in the *Critique of Pure Reasons*, which constitutes the centre of Kant's theory of freedom, whereupon I distinguish between his concepts of transcendental and practical freedom. Kant tries to solve the free agency problem by defending the possibility of a metaphysical space for freedom outside the realm of nature. I will argue that his theory, despite its subtleties, remains dualistic.

Hence, in **Chapter 3: Action and causation – an equation with two unknowns**, I approach the free agency problem from a different angle; namely by arguing that freedom is no "metaphysical possibility" for which we must make room, but rather a presupposition in the form of an analytical component of the concept of action. I start out by looking at the dispute in the philosophy of science known as the *reasons* vs. *causes* debate. Against this background I defend an interventionist account of the relationship between causality and agency and an assumption that if we can make room for a concept of intentional action in our world-view, there is no problem of free agency over and above this. An intentional action is, by conceptual necessity, a free action.

Chapter 4: The normative approach to action builds on the concept of intentional agency suggested in Chapter 3, and aims at capturing the normative dimension of agency. I endorse an anti-reductionist approach to action, rationality, and normativity. This involves

[20] More specifically, I see this as a positive formulation of the free agency problem. The negative side of the problem is what the world must be like in order not to contradict this ability. I largely steer clear of the negative side of the problem; however it is touched on in Chapter 6, in relation to the disagreement between compatibilists and libertarians in the philosophical free will/free agency debate.

understanding action as *irreducibly normative*, in the sense that it cannot be analysed as physical events caused by pro attitudes. I look into respectively Robert Brandom's and Sebastian Rödl's versions of a normative theory of rational agency.

Chapter 5: The intersubjective basis of normativity takes Searle's concept of a "circle of intentionality" as its point of departure, and argues that the basis of normativity is intersubjective language. The chapter ends with a discussion of whether normativity must be explicated on the basis of morality.

In **Chapter 6: Language and world: Two theses of unity**, I attempt to work out some of the details of an anti-dualistic theory of free agency based on a pragmatic theory of language. More specifically, I look at how certain structural features of a complete language render possible an objectivation of action, thereby securing the necessary connection between a performative first-person perspective and objectified true-or-false propositions about the world. In this connection I apply Apel's critique of "the alternative conception" – a form of conceptual dualism of explanation vs. understanding. I endorse Apel's argument that we must emphasise the hermeneutic dimension "between" these two frames in order to avoid a dualistic conception of language as well as of actions and acting persons.

In other words, I defend a thesis of the *unity of language* and of indexical expressions as a key to this unity. Further, I defend a thesis of the *unity of world*, by looking into and criticising Habermas's differentiated world-view. I suggest the concepts of *veritative* and *performative being* as possible replacements for Habermas's concepts of the objective, social and subjective world. In this way, I attempt to avoid a situation where the differentiation of possible relations between language (users) and the world collapses into dualism, as well as to display how different perspectives and relations to the world are integrated in a complete language.

In **Chapter 7: Freedom and first-person priority**, I return to the starting block by looking at the traditional philosophical freedom debate and try to relate my approach to free agency to this debate. I start by pointing out the limits of possible objectivation of the performative perspective and that this perspective must be viewed as superior to observation. Further, I look at the debate between compatibilistic and libertarian defenders of free will/agency. This debate regards the relationship between freedom and determinism, i.e. the question if freedom of action depends on whether or not natural events are the result of deterministic laws of nature. Without concluding with a definite refutation of compatibilism, I point to certain inherent tensions within existing versions of this position, among others within Habermas's attempt at a non-scientistic, anti-dualistic compatabilism.

The emphasis in this thesis lies, however, on a positive account of free agency – i.e. on freedom as a non-circumventable "fact of reason" – and not on a negative account in the sense of the limitations a robust concept of free agency might put on our conception of the world and the laws by which it is governed.

Chapter 2: Kantian freedom – between mind and world

Ein jedes Wesen, das nicht anders als unter der Idee der Freiheit
handeln kann, ist eben darum, in praktischer Rücksicht, wirklich frei.
(Kant, GSM 448)

Der Freiheitsbegriff bestimmt nichts in Ansehung der theoretischen
Erkenntnis der Natur; der Naturbegriff eben sowohl nichts
in Ansehung der praktische Gesetze der Freiheit.
(Kant, KU 196)

2.1 Introduction: Freedom vs. nature

In this chapter I discuss and criticise Kant's theory of freedom as a starting point for my approach to the problem of free agency. In our everyday self-understanding we think of ourselves on the one hand as making decisions and acting upon them, and in this sense acting freely. On the other hand we tend to believe that everything happens as a causal effect of a preceding course of events. To the degree that I view my own actions as entering into the temporal chain of events, how can I at the same time think of them as the results of my free decisions? The metaphysical problem of free agency arises in the light of two coexisting, although apparently contradictory, interpretations of reality.

I think one reason why Kant's theory of freedom continues to be relevant is that it describes this clash in a forceful manner. The mainstream compatibilist chooses a non-confrontational strategy, claiming that there is no real conflict between determinism and freedom, while the incompatibilist denies the possibility of free acts in the case of determinism. Kant agrees with the incompatibilist in maintaining that there *is* a contradiction between freedom and determinism, preserving the intuition that there *is* in fact a philosophical problem to be solved here. His final solution is, however, in a specific sense compatibilistic: Free agency is consistent with universal determinism, given that the two reign in respectively

a noumenal and a phenomenal world. Allen W. Wood refers to Norman Kretzmar, who likens this to saying

> that a married couple is compatible, but only as long as they live in separate houses. (Wood 1984, p. 75)

It seems to me that Kant's solution is more problematic than that. I don't see a problem with a couple choosing to live in separate houses, but if the two never set foot in the same house it seems to be a matter of a *pro forma* marriage, in other words no marriage in the real sense of the word. I will argue that, in spite of all his efforts to avoid it, Kant's solution remains dualistic.

The main reason for the continuing relevance of Kant's theory of freedom is his encirclement of the problem of freedom as one regarding *rationality* and *normativity*. Kant understands the debate on human freedom as having practical reason as its main topic. In this sense, he follows on John Locke, who establishes that freedom cannot possibly be seen as a capacity for ungoverned behaviour, but must be viewed as an ability to be governed by reason:

> If to break loose from the conduct of reason (…) be liberty, madmen and fools are the only freemen. (Locke 1690, 186, § 50)

To Locke, and to Kant, being free is not the same as having "no strings attached." Rather, it is a matter of being bound by the right kind of strings, viz. the "forceless force" of the principles of practical reason. Kant defines freedom as the ability to act according to *reason*.[21] This definition points to an ability *not* to act according to our immediate wishes or motivations, but to suspend these or to distance ourselves from them on the grounds of better reasons. What distinguishes reasons from wishes is an inherent *normative* dimension, as Robert Brandom puts it:

> One of Kant's great insights is that judgements and actions are to be distinguished from the responses of merely natural creatures by their distinctive normative status, as things we are in a distinctive sense responsible for. (Brandom 2000a, p. 33)

[21] "Dieses Vermögen, stets nach Vernunft zu handeln", cf. *Kant's Gesammelte Schrifte* (Akademieausgabe) 28.2,2; 1068 (*Religionslehre Pölitz*).

The normative dimension is established by practical reason: A good reason to act is a *rational* reason, and a rational reason is one we *ought to* act according to.

The title of this chapter refers to a traditional philosophical opposition between the mind and the world. As a starting point it seems clear to me that a satisfactory concept of free agency must be securely tied to both of these concepts. Actions cannot just be considered mental entities, but must at the same time be recognised as entering into the world's actual course of events. My initial suspicion is that, although Kant's concept of freedom has a lot to offer when it comes to specifying the mental dimension of free actions, it is comparatively less suited to clarifying the "worldly" dimension of agency.

Kant contrasts freedom with nature, a dichotomy apparently parallel with the mind-world distinction, and places freedom in an "ideal world" different from the empirical world of phenomena. To what extent and in what way the two worlds are connected is, of course, a major topic of discussion among Kant's interpreters. What I seek is, however, a conception of free agency that allows us to experience actions as entering into the very same world as other events, whether this is understood as an ontological, epistemic or moral demand.

At the same time it seems unlikely that freedom can be proven in any empirical sense. An action's characteristic of being free cannot be "observed". It is therefore unclear in what sense we can take freedom to *exist in the world*. Certainly not in the same sense as we take objects like tables or mountains to exist. One might think that the difference is one between the "inner" and the "outer" realm, that while tables and mountains exist in the outer realm, we take the reality of freedom and joy, for instance, to be of an inner kind. I think, however, that the traditional distinction between "inner" and "outer" misses the target. If freedom could only be recognised "from the inside", I would have a hard time recognising the acts of other subjects as free in the same sense as my own acts. The important distinctions are rather those between descriptive and normative, and between empirically observable and "acknowledgeable". The existence of objects such as tables and mountains may be observed empirically, whereas the existence of freedom is something that may be *acknowledged* in me and in other subjects.[22] Kant's solution is more sophisticated than the psychologically oriented inner/outer distinction. The Kantian distinction between *intelligible* and *sensible*[23]

[22] Cf. below, Ch. 4 on the normative theory of action. Later in the thesis, in Chapter 6, I attempt to formulate this in grammatical terms, by arguing that the decisive distinction lies not between first-person present tense and all other forms, but rather between first, second and relative third person on the one hand and absolute third person in the other hand (cf. Øfsti 1994, p. 182).
[23] Cf. e.g. KrV A 540f/B568f.

does nevertheless correspond to a certain degree with the inner/outer distinction, and this indicates, I believe, some of the problematic aspects of Kant's theory of freedom.

This chapter does of course not amount to a complete review of all aspects of Kant's theory of freedom, not to speak of the tons of secondary literature on the subject. I have tried to give an overall picture, as well as to pick out certain elements that serve well as background for my approach to the free agency problem. I focus on the metaphysics of the problem, as expressed by Kant in the *Critique of Pure Reason* (KrV). Kant's moral theory, especially from the *Groundwork* (GMS) and the *Critique of Practical Reason* (KpV), will be touched upon to the extent that I find it relevant. I also comment upon certain elements from the *Critique of Judgement* (KU), concerning the attempt to bridge the "broad gulf that divides the supersensible from phenomena".[24]

I start out with an analysis of *The Third Antinomy* (2.2). The Third Antinomy constitutes the centre of the discussion on freedom in the KrV, as well as the basis for Kant's subsequent treatments of the topic and, indeed, for his entire philosophy. As Henry Allison writes:

> It is virtually impossible to overestimate the importance of the Antinomy to Kant's critical project. (Allison 2004, p. 357)

Against this background I distinguish Kant's theory of *transcendental freedom*, viz. the argument concerning the possibility of assuming freedom understood as a causality other than the causality of nature, from his theory of *practical freedom*, viz. his positive account of freedom in the sense of giving "practical proof" of its existence (2.3). In Chapter 2.4 I argue that, despite Kant's efforts to the contrary, his theory of freedom remains dualistic. Finally, in Chapter 2.5 I evaluate certain elements of Kant's theory of freedom concerning action and self-consciousness.

2.2 The Third Antinomy

Kant's theory of free agency is an attempt to rescue the idea of freedom while at the same time maintaining that our actions are determined by natural causes. If our actions are causally determined like other events in nature, how can they at the same time be free in the morally

[24] Cf. KU 195.

relevant sense, meaning that we are responsible for them? At the heart of the Kantian solution lies transcendental idealism, the separation of noumenon and phenomenon. Regarded as phenomena, actions are effects of our empirical character; they are events thoroughly determined by the causal laws of nature. The very same actions may, however, at the same time be regarded as noumena, resulting from of our intelligible character.

A fundamental distinction in Kant's philosophy is the one between *Naturbegriff* – the concept of nature – and *Freiheitsbegriff* – the concept of freedom. In his construction (KrV B 472-480) and resolution (B 560-587) of The Third Antinomy – and throughout his philosophy – Kant attempts to hold two seemingly incompatible theses together: 1) Everything that happens, happens according to strict laws of nature, and 2) Man can act freely and rationally, and be held responsible for his deeds. Kant construes the Antinomy within the frames of the concept of nature, i.e. he tries to show that the Antinomy is generated from the concept of causation itself. The resolution, depending heavily on transcendental idealism and the distinction between noumena and phenomena, is an attempt to theoretically– although only negatively – provide the grounds for freedom. Kant's construction of the Antinomy has been severely criticised, and I think rightfully so. However, even if we consider the construction as unsteady, this does not necessarily render an account of the relation between Naturbegriff and Freiheitsbegriff irrelevant, since freedom is given another – positive – foundation in other places in Kant's work.[25] In other words, Kant's strategy for resolving the Antinomy does not necessarily lose its entire legitimacy even if the Antinomy as it stands must be reconsidered or even rejected. Furthermore, Kant's theory serves as a point of departure, inspiration and guide for innumerable attempts to account for the relation between freedom and nature. In the following paragraphs I discuss the construction and resolution of The Third Antinomy in order to see what I might learn from it before entering into further inquiries into the concept of free agency.

An antinomy (from *anti*-, against, and *nomos*, law) consists, in its literal sense, of two laws, maxims, principles or rules which, when applied, turn out to be contradictory. In a broader sense, it means a paradox or any two well-founded statements that conflict with one another, as in a constellation of a thesis and its antithesis. In the case of The Third Antinomy, the two statements in question are the following:

[25] Cf. e.g. KpV 9; GMS 446ff.

Thesis

Causality in accordance with laws of nature is not the only causality from which the appearances of the world can one and all be derived. To explain these appearances it is necessary to assume that there is also another causality, that of freedom. (KrV A 444/B 472)

Antithesis

There is no freedom; everything in the world takes place solely in accordance with laws of nature. (A 445/B 473)

These two statements and their respective proofs are set up against each other as contradictory. Both sides assume the validity within the realms of experience of a "causality of nature" (as confirmed in the Second Analogy). The question is whether it is also necessary, or permissible, to assume another type of causality, namely transcendental freedom, defined by Kant as "the power of beginning a state spontaneously [von selbst]".[26] The thesis confirms the need to appeal to a causality of freedom, whereas the antithesis denies both the need and the possibility of appealing to an alternative causality. As with the other antinomies, Kant's method of approach is to allow each side to plead its case by demonstrating the impossibility of the alternative. The decisive argument by each side is that the opposing claim is contradictory.

The thesis appeals to the requirements for a complete explanation: In order to fully account for appearances, it is necessary to assume a causality of freedom in addition to the causality of nature. Otherwise we are saying that every state presupposes a preceding state, hence that there can be no completeness of the series of causes. But this conflicts with "the law of nature", namely that everything that takes place has a cause that can be sufficiently determined a priori. Hence, if universalised, the assumption that causality of nature is the only kind of causality is self-contradictory. And because Kant treats causality of nature and causality of freedom as the only two types conceivable to us,[27] he has now established the need to assume a causality of freedom.

Kant's proof of the thesis has been severely criticised from many directions. A major question is why it should be considered unacceptable that the series of causes is interminable. The contradiction is supposedly generated by what Kant calls a "law of nature", namely that "nothing takes place without a cause sufficiently determined a priori".[28] According to Henry

[26] KrV A 533/B 561.
[27] Cf. KrV B 560.
[28] It should be noted that Kant's terminology when it comes to the law(s) of nature is confusing. Geert Keil counts the concept of the "law(s) of nature" seven times in the proof of the thesis, of which five references are

Allison, the thesis is best understood as a polemic against Leibniz.[29] The Leibnizian position is that every occurrence has a sufficient reason, meaning both that it has an antecedent cause and that it must have an ultimate explanation – although accessible only to God. Leibniz rules out the possibility of spontaneity, maintaining that even the divine will itself is determined (although not necessitated) by the divine intellect. The "law of nature" – that "nothing takes place without a cause sufficiently determined a priori" – is a formulation of this dual, Leibnizian requirement. The conflict arises thus between the demand that every explanans be in turn regarded as an explanandum (the "universalisability requirement") and the demand that there be an ultimate explanans in which the series of explanations are grounded (the "completeness requirement"). The advocate of unrestricted causality insists on universalisability, and is thereby led to deny the completeness requirement.

Allison analyses the "law of nature" as a requirement of ultimate intelligibility:[30] The principle of sufficient reason means that we, in theory, must be able to complete any explanation to a point where our thought can rest. The thesis states that there is a conflict between the requirement for such an ultimate resting place for thought – for which no cause can be given – and the requirement for unrestricted causality. It seems, however, easy to avoid this conflict simply by denying Leibniz' commitment to the principle of sufficient reason. Once we abandon Leibnizian metaphysics, it seems that we are free to reject the thesis

> on the grounds that it conflates the requirement for the causal explanation of an occurrence (produce a 'sufficient' cause) with the requirement for the justification of a conclusion (produce a complete set of premises). (Allison 1990, p. 18)

However, Allison points to a fundamental principle underlying the antinomical conflict, namely that reason itself demands a resting place for thought, a complete justification of every explanation, in Kant's words:

---

plural and two are singular. The predominant formulation is that "everything happens in accordance with the laws of nature", where the laws in question presumably are the various *empirical* laws governing events in nature. The two references in the proof to a (singular) *law* of nature point, however, to the *principle of causality* itself, as proven in the second analogy; "everything that happens has a cause". Kant obviously sees a tight connection between the different empirical laws of nature on the one hand and the principle of causality on the other. In the resolution of the antinomy he writes: "That all events in the sensible world stand in thoroughgoing connection in accordance with unchangeable laws of nature is an established principle of the Transcendental Analytic, and allows for no exception". This shows that Kant understands the principle of causality as stating a connection according to unchangeable laws of nature. In other words he assumes the nomological character of causality without ever really arguing for it (cf. Keil 2000, p. 334ff.).

[29] Cf. Allison 1990, p. 14 ff.

[30] Cf. Op. cit., p. 18.

> [W]enn das Bedingte gegeben ist, so ist auch die ganze Summe der Bedingungen, mithin das schlechthin Unbedingte gegeben, wodurch jenes allein möglich war. (KrV A409/B436)

From Kant's point of view, the proponents of the antithesis are committed to this principle, but as the argument for the thesis shows, the principle rules out a complete explanation in terms of causal antecedents.

Kant's point is that this conflict is unsolvable, given the identification of appearances with things in themselves. Given transcendental realism, we necessarily construe the completeness requirement in a dogmatic manner. In other words, what Kant means to provide is not just a reductio of Leibniz's argument, but of the understanding of causality presumed by any transcendental realist.

The antithesis denies the possibility of an appeal to a causality of freedom, and maintains that everything happens in accordance with the laws of nature. The argument shows that the opposite assumption – that there is transcendental freedom – is contrary to the causal law and therefore cannot be encountered in any experience. This reflects the standpoint of the "pure empiricist". The antithesis is generally conceived as less troublesome than the thesis. In fact, the denial of freedom seems to be consistent, given the identification of things in themselves with appearances, in Kant's own wording:

> [S]ind Erscheinungen Dinge an sich selbst, so ist Freiheit nicht zu retten. (KrV A 536/B 564)

The only problem with the antithesis is that it sees the rejection of transcendental freedom in nature as equivalent to rejecting freedom altogether. When misunderstandings of transcendental realism are cleared up, the claim of the antithesis may be restricted to the quite correct assertion – according to Kant – that freedom is impossible *within the realm of nature*.

The aim of The Third Antinomy is to illustrate how transcendental realism fools us into thinking that there is a necessary conflict between exceptionless laws of nature on the one hand and transcendental freedom on the other. Given Kant's transcendental idealism, however, he is able to argue that this conflict is only apparent. What his argument seems to be lacking is strong support for the dichotomy between transcendental realism and transcendental idealism, i.e. the assumption that given the rejection of the first we are forced to presume the second. Kant's well-founded critique of the Leibnizian view on causality loses some of its force if it can be argued that his transcendental idealism is not the only alternative to transcendental realism.

Kant's treatment of the antinomies of pure reason falls into two categories. The first two "mathematical" antinomies, concerning quantity and reality, he resolves by displaying a false presupposition common to them, namely the presupposition that the sensible world is a whole existing in itself. Given this premise, the conflicting claims (in the first Antinomy: that the world is unlimited vs. that it is limited in time and space) are genuine contradictions, of which only one can and must be true. When this presupposition is rejected, however, the apparent contradiction turns into a "dialectical opposition" between contraries, both of which are false. Kant intends the resolution of the mathematical antinomies to amount to an indirect proof of transcendental idealism.[31] Since the resolution of The Third Antinomy depends on transcendental idealism, it is relevant to review this attempted proof. What Kant tries to establish is not only that transcendental idealism provides us with a key to the solution of the Antinomy, but that it is the indispensable key, without which one is bound to fall victim to the "euthanasia of pure reason".[32] Allison reformulates Kant's decisive argument:

> Since transcendental realism necessarily assumes that the world is a whole existing in itself, it likewise must assume that it is either finite or infinite in the relevant respects. But the analysis of the Mathematical Antinomies (…) has shown that this world can be neither finite nor infinite. Consequently, both the conception of the world, which has been shown to be self-contradictory, and the transcendental realism underlying it must be rejected. Finally, given the dichotomy between transcendental realism and transcendental idealism, the negation of the former is logically equivalent to the affirmation of the former. In short, transcendental idealism is true. (Allison 2004, p. 391)

Allison points out that Kant's argument actually rests on an additional premise, namely the "principle of pure reason", as presented in KrV A 307f./B 364f.:[33]

> [W]enn das Bedingte gegeben ist, so [ist] auch die ganze Reihe einander untergeordneter Bedingungen, die mithin selbst unbedingt ist, gegeben, (d.i. in dem Gegenstande und seiner Verknüpfung enthalten).

Kant wants to show that the principle of pure reason, which seems indispensable to human reason, is in fact an illusion when interpreted from the viewpoint of transcendental realism. The principle forces us to consider the world as a whole. When it is combined with transcendental realism – the view that appearances are things in themselves – we get the self-

[31] Cf. KrV A 506f/B534f.; Allison 2004, p. 388ff.
[32] Cf. KrV A 407/B 434: "..sich entweder einer skeptischen Hoffnungslosigkeit zu überlassen, oder einen dogmatischen Trotz anzunehmen (...). Beides ist der Tod einer gesunden Philosophie, wiewohl jener allenfalls noch die **Euthanasie** der reinen Vernunft genannt werden könnte".
[33] Cf. KrV 409/B436, quoted above.

contradictory result presented in the resolution of the Mathematical Antinomies. Allison argues that transcendental idealism from this perspective should be seen as a therapeutic tool in dealing with cosmological problems, and not as the phenomenalistic dogma it is usually taken to be:

> [T]he transcendental distinction, which constitutes the heart of transcendental idealism, is a bit of meta-philosophical therapy rather than a first-order metaphysical doctrine. (Allison 2004, p. 395)

This rather Wittgensteinian reading[34] agrees well with Kant's attack on dogmatism. However, I think the textual support for Allison's reading is weak. Kant clearly does not introduce transcendental idealism simply as a pragmatic device for dealing with philosophical problems, but as a true metaphysical doctrine.[35]

In my view, Kant does not achieve his intended indirect proof of transcendental idealism through his treatment of the first two "mathematical" antinomies. As I argued above, Kant lacks strong support for his dichotomy between transcendental realism and transcendental idealism. Given that there are alternatives other than these two, it seems circular or question-begging to launch the resolution of the antinomies as a proof of transcendental idealism while at the same time using transcendental idealism as a pivotal step in the resolution. However, as long as transcendental idealism is not *refuted*, Kant may well have achieved his goal on the matter of transcendental freedom, namely to show that it cannot be *dis*proved.

If we return to the second pair of antinomies, the two "dynamical" antinomies, we see that Kant treats these differently from the "mathematical" ones, where the competing claims are both shown to be *false*. The assumption underlying the "dynamical" antinomies is that the competing claims are contradictory, and the resolution consists in showing that they really are *compatible*. In the case of The Third Antinomy, Kant tries to show that the competing claims – existence/non-existence of a causality other than that of nature – are subalternates rather than contraries. In other words, he allows for the possibility that both the thesis and the antithesis are correct: The thesis in asserting an intelligible, transcendentally free, first cause (outside of the realm of experience), and the antithesis in refuting such a cause (within experience).

Kant's strategy is to expose the dogmatism of pure empiricism. The pure empiricist infers improperly from the correct premise that all *empirically cognisable* causality must

[34] Cf. e.g. Wittgenstein, PU §§ 133, 255.
[35] Cf. notably KrV A 369.

conform to the laws of nature, to the potentially false conclusion that *all causality* must conform to the same laws. Transcendental idealism provides a solution by creating the conceptual space needed to understand freedom as non-empirical causality. Whether or not this conceptual space is actually filled is something that cannot be established, at least not theoretically. This is why Kant denies having shown the reality of transcendental freedom, claiming only to have shown that freedom and nature are not necessarily in conflict with each other.[36]

We may regard The Third Antinomy as a battle between reason and understanding, with the result being a draw. Reason demands an unconditioned foundation for a series of empirical causes; an absolute beginning. Understanding deems such a thing to be incomprehensible, since an event must be seen as caused by an antecedent state to be cognisable at all. Kant's solution is to let understanding have its unbroken series of causes in the realm of phenomena, while allowing reason to find its resting place in the realm of noumena.

2.3 Transcendental and practical freedom

A frequently used terminology in the philosophical debate on freedom and determinism is that which categorises the debaters as either *compatibilists* or *incompatibilists*. Compatibilists hold that free agency and determinism are compatible. Our actions may be determined by natural causes and at the same time be free in the relevant sense, i.e. concerning moral norms and responsibility. Incompatibilists hold that if our actions are determined by natural causes, then free agency is an illusion. Kant's theory of freedom does not fit neatly into this categorisation. His theory is that actions may be simultaneously free and causally determined. This does not, however, make him a compatibilist in the traditional sense. The mainstream compatibilistic strategy is to reject the existence of a true opposition between freedom and determinism, thus to deny that this constitutes a deep metaphysical problem.[37] Kant, on the other hand, chooses the opposite strategy. He brings the conflict between freedom and determinism to its peak, whereupon he launches the solution that they are compatible only because man belongs to two

[36] Cf. KrV A558/B 586.
[37] Cf. e.g. Hume 1739, part III.

worlds: One noumenal world in which he is a free, moral subject, and one phenomenal world in which he is a determined, natural object. Allen W. Wood says that Kant wants to show

> not only the compatibility of freedom and determinism, but also the compatibility of compatibilism and incompatablism. (Wood 1984, p. 74)

Kant introduces a strict opposition between nature and freedom, but at the same time he insists that the contradiction between them is only apparent.

A central distinction in Kant's theory of freedom is the one between transcendental and practical freedom. *Transcendental freedom* is a purely metaphysical concept, equivalent to a particular form of causality, viz. the causality of freedom. It is defined as "the faculty of beginning a state spontaneously [von selbst]" (KrV A533/B561). *Practical freedom* is the same as free agency, and is what we ascribe to ourselves when we understand ourselves as being morally responsible (KrV A534/B562). Kant further distinguishes two concepts of practical freedom: Practical freedom in the negative sense is our capability of resisting sensuous desires. This capability belongs solely to human beings, whereas animals must always act according to their sensuous impulses. Practical freedom in the positive sense is the power to act morally, i.e. from a thoroughly non-sensuous, purely ideal motive (Loc.cit.).

In order to fully grasp Kant's theory of freedom it is important to understand the connection between transcendental and practical freedom. It is not immediately clear how the cosmological concept of an alternative form of causality is connected with the morally relevant concept of free agency. In the Observation on the Thesis (*Anmerkung zur Thesis*, KrV A448/B476), Kant writes that the transcendental idea is included as an essential ingredient in the mainly empirical "psychological concept" (equivalent to what he later refers to as practical freedom). He localises the problem of freedom to its transcendental aspect, because admitting free agency necessarily means admitting "unconditioned causality". We therefore have to argue the possibility of transcendental freedom in order to enable the possibility of free agency. In KrV A534/B562, Kant writes that the practical concept of freedom is based on the transcendental idea, and that "in the latter lies the real source of difficulty". Allen W. Wood elaborates:

> The free will problem arises for Kant because he believes that practical freedom requires transcendental freedom and that there is no room in the causal mechanism of nature for a transcendentally free being. (…) Practical freedom requires that we be able to determine our actions entirely from within ourselves, through our own legislative reason. Natural causes, however, belong to an endless regressive chain in

> which there is no spontaneous or first cause. We can think of ourselves as practically free, therefore, only by thinking of actions as subject to a transcendentally free cause lying outside nature. (Wood 1984, p. 77)

According to Kant, the morally relevant concept of practical freedom depends on the possibility of a transcendentally free will. Once we have proven the possibility of "spontaneously beginning a state", we have not only shown the possibility of a free act as the absolute first beginning in time, but also the possibility of attributing freedom to other acts entering into the world's course of events.

The desired *proof* of transcendental freedom turns out to be unattainable, however. Kant never claims to have proven freedom in the theoretical sense; in fact he states clearly that such a proof is impossible. Instead he construes a metaphysical theory of freedom which, admittedly, cannot be shown to be correct, but which is allegedly impossible for the opponents of freedom to refute:

> Man muss wohl bemerken: dass wir hierdurch nicht die Wirklichkeit der Freiheit, als eines der Vermögen welche die Ursache von den Erscheinungen unserer Sinnenwelt enthalten, haben dartun wollen. Denn, außer dass dieses gar keine transzendentale Betrachtung, die bloß mit Begriffen zu tun hat, gewesen sein würde, so könnte es auch nicht gelingen, indem wir aus der Erfahrung niemals auf etwas, was gar nicht nach Erfahrungsgesetzen gedacht werden muss, schließen können (...) [D]ass Natur der Kausalität aus Freiheit wenigstens nicht widerstreite, das war das einzige, was wir leisten konnten, und woran es uns auch einzig und allein gelegen war. (KrV A558/B586)

Allen W. Wood compares Kant's task when it comes to freedom to the task of a defence attorney:

> [W]e may assume that freedom is innocent until proven guilty, [and] that the burden of proof lies on those who would undermine our moral consciousness by claiming that we are not free. (Wood 1984, p. 83)

Freedom is presupposed in our everyday morality, and so it should be the task of those attacking it to prove its non-existence. Geert Keil makes the same point in *Handeln und Verursachen*:

> Man kann zeigen, dass der Freiheit *nichts entgegensteht*, aber da nicht zu sehen ist, was als ein positiver Freiheitsbeweis zählen könnte, muss Freiheit, mit Kants Wort, postuliert werden. Das besagt nicht, dass sie eine 'Fiktion' oder eine 'notwendige Illusion' wäre, sondern drückt eben den Umstand aus, dass sie

> nicht bewiesen – aus unabhängig gesicherten Prämissen deduziert – werden kann. Wenn es aber für etwas keinen Beweis *gibt*, liegt der Fehler bei demjenigen, der gleichwohl einen verlangt. (Keil 2000a, p. 472)

Where no (deductive) proof is possible, or even thinkable, it is a mistake to ask for one.[38]

In the first critique, in the solution of The Third Antinomy, Kant settles for a proof of the *possibility* of a causality of freedom. Hereby he attains a *negative* proof in the sense that nothing – transcendentally speaking – stands in the way of freedom. Not until the second critique does he launch a *positive* proof of freedom. Here, he argues that our rational cognition of the practical laws of freedom implies not only the possibility, but the actual existence of freedom.

> Der Begriff der Freiheit, so fern dessen Realität durch ein apodiktisches Gesetz der praktischen Vernunft bewiesen ist, macht nun den *Schlussstein* von dem ganzen Gebäude eines Systems der reinen, selbst der spekulativen, Vernunft aus, und alle andere Begriffe (die von Gott und Unsterblichkeit), welche, als bloße Ideen, in dieser ohne Haltung bleiben, schließen sich nun an ihn an, und bekommen mit ihm und durch ihn Bestand und objektive Realität, d.i. die Möglichkeit derselben wird dadurch bewiesen, dass Freiheit wirklich ist; denn diese Idee offenbaret sich durchs moralische Gesetz.
>
> Freiheit ist aber die einzige unter allen Ideen der spekulativen Vernunft, wovon wir die Möglichkeit a priori wissen, ohne sie doch einzusehen, weil sie die Bedingung des moralischen Gesetzes ist, welches wir wissen. (KpV 4f.)

In the absence of a theoretical proof, he relies on a so-called "practical proof". This is no proof in the empirical sense, but nevertheless a strong reason or – as Kant calls it – a "fact of reason". Kant argues from the thesis that freedom must be presupposed, even if the reality of freedom remains unsettled from a theoretical point of view:

> Ein jedes Wesen, das nicht anders als unter der Idee der Freiheit handeln kann, ist eben darum, in praktischer Rücksicht, wirklich frei. (GMS 448)

[38] One could argue that Keil here presupposes that proofs of the deductive kind are the only valid ones, without allowing for other possibilities, such as Apel's figure of "Letztbegründung". I will not discuss this further here, suffice it to say that Keil's target here primarily seems to be the attempt to deny freedom because no *deductive* evidence may be given for it.

Since we are unable to act except under the idea of freedom, we are bound by the moral law whether or not it can be (theoretically) proven that we are free. In every practical respect we are, therefore, free.

This reminds us that Kant ties the conception of practical freedom tightly to the conception of morality:

> [E]in freier Wille und ein Wille unter sittlichen Gesetzen [ist] einerlei. (GSM 447)

A widespread critique of Kant's theory of freedom is that it only allows us to consider moral acts as free. Only actions motivated by practical reason are free, and necessarily in accordance with the moral law. Apparently, actions motivated (if only partly) by sensuous desires are necessitated by mechanisms of nature, hence unfree. This seems to lead into an absurdity, namely that we are morally responsible only for our good deeds, and that no one can be held responsible for immoral acts. Allan W. Wood argues that this problem is illusory, since Kant clearly states that although sensuous desires have an effect upon us, this does not mean that actions are *determined* by natural causes:

> Since our will is free, our heteronomous actions are performed from sensuous motives without being necessitated by them. (…) [W]e act as practically free beings, and hence are responsible for what we do. (Wood 1984 p. 78)

Practical freedom consists in the *capacity* to act autonomously, although this power may be seldom exercised. Similarly, Henry Allison stresses that according to Kant, all human agency involves spontaneous self-determination, and is never simply determined:

> [E]ven desire-based or, as Kant later termed it, 'heteronomous' action involves the self-determination of the subject and, therefore, a 'moment' of spontaneity. (Allison 1990, p. 39)

Actions may be induced by sensuous motives, but are not entirely determined by them. Spontaneity is a necessary ingredient, hence an act is never simply determined, but is always a result of self-determination. We find a key formulation of this principle in *Religion within the Limits of Reason Alone*, in what Allison calls the *Incorporation Thesis*:[39]

[39] Cf. Allison 1990, p. 40.

> [F]reedom of the will is of a wholly unique nature in that an incentive can determine the will to an action only insofar as the individual has incorporated it into his maxim. (Rel. 6: 24: 19)

What constitutes free agency is not the character of our motives, but our ability to determine what motives to act upon.

Christine Korsgaard is an innovative Kant interpreter who among other things argues that Kant's principle of autonomy must be seen as the ultimate source of normativity.[40] However, she seems to have inherited the problem regarding the status of morally bad action. To Korsgaard, a good action is one that successfully constitutes the agent as a unified subject:

> '[A]ction' is an idea that admits of degrees. An action chosen in a way that more successfully unifies and integrates its agent is more authentically, more fully, an action, than one that does not. (Korsgaard 2002, Lecture 1, p. 17)

This seems to imply that an agent is to a larger degree morally responsible for those actions that are "successful" in Korsgaard's sense. To me, Korsgaard comes close to committing what we might call a "Platonistic fallacy" here, in the sense that immorality and other "defects" are seen as "less real". Normative evaluation is viewed as ontologically significant in the sense that certain things or phenomena, like rational actions, exist *to the degree that* they fulfil some normative standard. In that sense, Korsgaard seems to view normativity "through ontological spectacles". This is quite different from the necessity of viewing reality "through normative spectacles". In Chapter 4 below I argue that *being subject to* normative standards is a distinguishing mark of a rational action. But in the above quoted passage, Korsgaard seems to argue that *the fulfilment of* a certain normative standard is necessary for something to be viewed as a rational action. But what becomes of freedom if we cannot properly choose to do wrong? However, she shares Kant's view that we are responsible for the *incorporation* of motives into our maxims, and in this sense we may be held morally responsible for our immoral actions, although these in themselves do not contribute to constituting us as free agents.

The freedom to be immoral is a perplexing issue. Moral rationalism seems to suggest that immoral behaviour must count as irrational. But in that case it cannot be viewed as free agency, but must be seen as caused by empirical incentives. However, Kant (and Korsgaard) clearly addresses the ability to carry out immoral deeds in full awareness of their offence

[40] Cf. Korsgaard 1996; Cf. below, Ch. 4.8.

against the moral law. Kant calls such actions *radical evil*, and emphasises that these evil acts "muss aus der Freiheit entspringen" in order for us to be able to hold people responsible for their wrongdoings.[41] Natural urges are in themselves morally neutral. But given the Incorporation Thesis, it is the wilful incorporation of such urges into one's maxims – i.e. allowing them to take the motivating place of the moral law – that is blameworthy.

2.4 Kantian dualism

Kant's theory of freedom entails in a certain sense a causal theory of action. Free acts are not uncaused, but rather are caused in a particular way, namely through what he calls the "causality of freedom".[42] Kant distinguishes two concepts of causality as the only two conceivable to us:

> Man kann sich nur zweierlei Kausalität in Ansehung dessen, was geschieht, denken, entweder nach der Natur, oder aus Freiheit. (KrV B 560)

This seems to suggest that everything that happens is caused in either one or the other sense; an avalanche by the causality of nature, a greeting by the causality of freedom. However, actions too are caused in an empirical sense. What distinguishes an action, as opposed to other things happening, is that it is caused in a two-fold sense: Depending upon our perspective, we may see an act as empirically caused by preceding events (causality of nature) or as noumenally caused by the intelligible character (causality of freedom). It is important that Kant does not claim that we are victims of an illusion. It is not the case that it only appears to me that I am free, whereas I "really" am determined; or that it only appears to my neighbour or fellow philosopher that I am causally determined, whereas I "really" am free. It is an (empirical) fact that I am causally determined, and it is equally a fact (of reason) that I am free.

The idea of freedom is something we actively attribute to ourselves and to other subjects.[43] In doing this we remain aware that our actions have causal conditions, but we

[41] Cf. Rel., B 24/A 22.
[42] Cf. e.g. KrV B 472.
[43] A problem with Kant's transcendental-idealistic solution is, however, how we supposedly are able to single out the particular *Erscheinungen* with which we can pertinently associate the transcendental idea of freedom, cf. Øfsti 1976.

maintain that these conditions are only sufficient to explain our actions viewed as empirical appearances. In addition we must think of our actions as absolute beginnings in time, and of ourselves as freely initiating a causal series.

This double concept of causality demands an account of the relation between the causality of freedom and that of nature. Kant's solution to this problem involves separating the subject into its intelligible and empirical character. The intelligible character, he writes, is not given to us as an object of empirical recognition. Therefore, we can never be sure if – or to what extent – a given act is free:

> Die eigentliche Moralität der Handlungen (Verdienst und Schuld) bleibt uns daher, selbst die unseres eigenen Verhaltens, gänzlich verborgen (...) Wie viel (...) davon reine Wirkung der Freiheit, wie viel der bloßen Natur (...) zuzuschreiben sei, kann niemand ergründen, und daher auch nicht nach völliger Gerechtigkeit richten. (KrV, B 579 fn.; Cf. also GMS, 407)

This limitation only applies to the epistemic perspective, however. This perspective can be contrasted with a practical perspective: One is what we can theoretically know; the other is the basis upon which we act. From the practical perspective we may hold an agent responsible. We do this even if we – as Kant does – assume psychological determinism, i.e. believe that human actions, regarded empirically, are caused by the empirical character of the agent:

> Ob man nun gleich die Handlung dadurch bestimmt zu sein glaubt: so tadelt man nichtdestoweniger den Täter, und zwar nicht wegen seines unglücklichen Naturells (...), denn man setzt voraus, man könne es gänzlich beiseite setzen, wie dieser beschaffen gewesen (...), als ob der Täter damit eine Reihe von Folgen ganz von selbst anhebe. (KrV, B 582f.)

The explanation for this peculiarity is that the rationally acting subject *in acting* (performatively) knows himself to be a rationally acting subject, and thereby is conscious of his capacity to act according to the moral law:

> [D]ie Handlung wird seinem intelligiblen Charakter beigemessen, er hat jetzt, in dem Augenblicke, da er lügt, gänzlich Schuld; mithin war die Vernunft, unerachtet aller empirischen Bedingungen der Tat, völlig frei. (Loc.cit.)

Man is “citizen of two worlds”: To the extent that man and his actions appear empirically, he must be considered completely as subject to the causal laws applying to objects in the world of phenomena. Man can, however, conceive of himself as free and independent from this causality of nature, to the extent that he (performatively) conceives of himself as a source of spontaneity.

Kant’s theory says, in short, that one and the same act may be simultaneously free and causally determined, because we belong to two realms. A widespread interpretation of what it means to belong to two realms is that our existence can be viewed from two perspectives. This is known as the “two-aspect” interpretation. The alternative is a “two-object” or “two-world” interpretation, which says that appearances and things in themselves constitute two ontologically distinct sets of entities.

I think there is considerable support for the “two-aspect” interpretation of Kant’s theory. However, even a “weak” dualism, viz. an aspect dualism, suffers from certain unsolvable dilemmas. The quote above, where Kant writes that we view the action “as if” (“als ob”) the actor initiates a series of events, is revealing in that respect. Kant’s transcendental idealism seems to entail that noumenal actions appear illusory as seen from the empirical world, while phenomenal actions become mere phantoms from the noumenal perspective. A standard objection to such a treatment of the question of free acts is that it leads to an unsolvable dilemma. In Allison’s wording:

> Either freedom is located in some timeless noumenal realm, in which case it may be reconciled with the causality of nature, but only at the cost of making the concept both virtually unintelligible and irrelevant to the understanding of human agency, or, alternatively, freedom is thought to make a difference in the world, in which case both the notion of its timeless, noumenal status and the unrestricted scope within nature of the causal principle must be abandoned. (Allison 1990, p. 2)

In short, we seem to be forced to choose between a practically empty and a theoretically unacceptable reading of Kant’s theory.

In the first critique, Kant establishes the possibility of freedom, in the second he claims to prove its actual existence. At the same time he presents a metaphysical model of a hierarchic relationship between freedom in the noumenal realm and the determinism prevailing in the phenomenal realm. According to this model, freedom reigns above determinism, since

> [die] ganze Kette von Erscheinungen in Ansehung dessen, was nun immer das moralische Gesetz angehen kann, von der Spontaneität des Subjekts, als Dinge an sich selbst, abhängt, von deren Bestimmung sich gar keine physische Erklärung geben lässt. (KpV 178)

Although the law of causality is universally valid within the phenomenal realm,

> the entire history of the phenomenal realm could be taken to reflect the free choice of a rational agent in the noumenal realm. (Guyer 1993, p. 29)

Free agency cannot be seen as intervening at any particular moment in the history of the phenomenal world, but ultimately we may view the world as a whole as resulting from free agency. This fits well with the proof of the thesis in The Third Antinomy. At this stage of Kant's philosophy, the relationship between the noumenal and the phenomenal seems quite clear.

However, in the third critique, Kant challenges this relationship, speaking about a "broad gulf" separating freedom from nature:

> Der Freiheitsbegriff bestimmt nichts in Ansehung der theoretischen Erkenntnis der Natur; der Naturbegriff eben sowohl nichts in Ansehung der praktische Gesetze der Freiheit. (KU 195)

Kant seeks a "bridge" over this gulf, and what he finds is the faculty of judgement:

> [D]ie Urteilskraft (…) gibt den vermittelnden Begriff zwischen den Naturbegriffen und dem Freiheitsbegriffe, der den Übergang von der reinen theoretischen zur reinen praktischen, von der Gesetzmäßigkeit nach der ersten zum Endzwecke nach der letzten möglich macht, in dem Begriffe einer Zweckmäßigkeit der Natur. (KU 196)

At the core of the third critique is an attempt to unite the two worlds of man. Despite the solution of The Third Antinomy in the first critique and the practical proof in the second critique, Kant clearly does not consider the problem as solved.

Paul Guyer attempts to clarify the connection between nature and freedom in Kant's philosophical system. He states that the unity between nature and freedom cannot be proven theoretically, but must be postulated.[44] The connection must be realised within nature, since this is the arena for the activity of mankind. We must be able to see nature as a realm where freedom *should* have an influence:

[44] Cf. Guyer 2005, p. 286ff.

> [D]er Freiheitsbegriff soll den durch seine Gesetze aufgegebenen Zweck in der Sinnenwelt wirklich machen, und die Natur muss folglich auch so gedacht werden können, dass die Gesetzmäßigkeit ihrer Form wenigstens zur Möglichkeit der in ihr zu bewirkenden Zwecke nach Freiheitsgesetze zusammenstimme. (KU, Einleitung II, 176)

The unity of nature and freedom is secured through what Guyer refers to as "the concept of the highest good"; a morally necessary concept "for which God is the ground".[45] Guyer argues against a dualistic reading of Kant's philosophy, but according to his interpretation, the "great gulf" separating nature and freedom can only be bridged by an appeal to God. This solution hardly agrees well with the "methodological atheism" of modern philosophy.

2.5 Action and self-consciousness

In the first critique, Kant defends the theoretical possibility of transcendental freedom, i.e. the possibility of a "metaphysical space" for freedom outside of the limits of theoretical reason. The concept of transcendental freedom is a metaphysical concept applying to everything in so far as everything may be viewed as noumenon and hence as independent of determining causes. Thus far, however, human freedom does not differ from "the freedom of a roasting-jack", which, once it is wound up, performs its motions of itself.[46] The distinguishing mark of human beings is our *self-consciousness*, i.e. an ability to be aware of our freedom. This self-awareness is out of reach for kitchen equipment as well as for other things and animals; thus only human beings (to our knowledge) are able to gain autonomy. Free agency is realised only when the agent incorporates his motives into the maxim of the act, an operation we are capable of thanks to "pure apperception":

> [D]asjenige Selbstbewusstsein (…), was, indem es die Vorstellung Ich denke hervorbringt, die alle anderen muss begleiten können. (KrV B 132)

[45] Cf. Guyer 2005 p. 293.
[46] "Freiheit eines Bratenwenders", cf. KpV 174.

Self-consciousness is clearly an important element in any theory of free agency. However, we must bear in mind that the Kantian concepts of action and self-consciousness deviate from a modern understanding of the words. Wolfgang Becker argues that action is a fundamental concept in *Kritik der reinen Vernunft*, not least with regard to its role in explaining other central concepts, such as 'judgement' and 'synthesis'. However, Kant's epistemology cannot be considered to be a theory of action in a modern sense. Kant refers mainly to "acts of the understanding" – *Verstandeshandlungen* – e.g. the passing of a judgement. According to Becker, Kant's concept of action remains

> hoffnungslos mentalistisch und allenfalls metaphorisch. (Becker 1987, p. 42)

The *Verstandeshandlung*en of KrV are not actions in our sense of the word, but should rather be viewed as constitutive performances necessary for acting in a practical sense. They are "acts of pure thought" (cf. KrV B 81), thus non-empirical and not determined in time.[47]

This leads us to a well-known discussion in the Kant literature, concerning so-called "timeless acts". It seems to follow from the Kantian theory of freedom that the only way to conceive of our actions as free is to regard them as outside and independent of the temporal flow of events, just like the reason they originate from:

> Denn da Vernunft selbst keine Erscheinung und gar keinen Bedingungen der Sinnlichkeit unterworfen ist, so findet in ihr, selbst in Betreff ihrer Kausalität, keine Zeitfolge statt. (KrV B 581)

When an action is considered in the flow of time, it falls under strict, empirical laws. Kant is a psychological determinist and holds that, given sufficient empirical knowledge, every human act can be predicted with certainty.[48] Freedom is an aspect of an action only in so far as it is considered independently of its empirical qualities, i.e. as timeless:

> [I]ch bin *in dem Zeitpunkt* darin ich handle, niemals frei. (KpV 169, my italics).

However, other passages pull in the opposite direction:

---

[47] Cf. Becker 1987 p. 49f . Becker further criticises Kant for being unable to come up with a concept of action that is suitable for both his theoretical and his practical philosophy. Kant's interest in the concept of action is primarily practical – it concerns how we may think about freedom without contradiction. However, the concept serves as a fundamental part of his epistemology as well, and so what Kant really needs is "ein Handlungsbegriff, der nicht schon vorwiegend oder gar ausschließlich an Problemen der praktischen Philosophie und der Ethik orientiert ist" (Becker 1987, p. 59).

[48] Cf. e.g. KrV B 578 and KpV 177.

> [E]r hat jetzt, *in dem Augenblicke*, da er lügt, gänzlich Schuld. (KrV B 583, my italics)

In the latter quote, Kant points to the "fact of reason"; that an agent inescapably *must* consider himself as free *in acting*. But the question is: How *can* I view my action as free – and myself as responsible – if the outcome of my action is determined and could have been predicted prior to my decision? Kant famously argued that "ought implies can",[49] but it is not easy to see how this can apply to concrete actions.

Allen Wood takes Kant to mean that while events in time follow a necessary order as determined by their empirical causes, the *whole course* of the world's history is a result of the "timeless choice" of the free subject.[50] The timeless, intelligible choice of the subject is the ultimate reason for everything that happens. He tries to rescue Kant from the most monstrous consequences of this interpretation – such as that I can be blamed for World War I – by stressing that I do not directly select a certain history of actions. My intelligible choice regards my empirical character and with it the "fundamental maxim" for my actions.

It seems to me that if Wood's interpretation is coherent, then Kant's conception of free agency is of little importance to a modern theory of action, since the subject matter seems to be a wholly different one. Henry Allison writes:

> [O]ur conception of our agency is obviously inseparable from our understanding of ourselves as temporal beings. How, then, could it be claimed that the conception of ourselves as timeless beings is somehow required in order to 'model' or regulate our ordinary understanding of ourselves as agents who act and decide in time? (Allison 1990, p. 47)

Allison argues that while Wood's interpretation lends strong support from Kant's own choice of words in the Critiques, it is not in agreement with what he conceives as a central aspect of Kant's theory, namely the above-quoted "Incorporation Thesis". According to this principle, the agent must incorporate his motives into his maxims. Allison argues that this supports a reading that allows for the possibility of assigning a regulative function to the conception of an intelligible character without denying that actual actions take place in time. The Incorporation Thesis allows a (timeless) "moment of spontaneity" to enter into the motives of a temporal act.

---

[49] Cf. e.g. KpV 54.
[50] Cf. Wood 1984 p. 89ff., and above, Ch. 2.4.

Still, it seems to me that such an account remains far from our everyday understanding of agency. A relevant concept of free agency must imply that it is the actual *carrying out* of actions *in time* to which we ascribe freedom, for it is these actions we are held responsible for. Normally, we make our decisions based on the intention to achieve a more or less immediate effect. If Kant's concept of rational choice is something completely different from our deliberations immediately prior to the act, then his theory is not what we ask for in a modern theory of action.

Allison is clearly preoccupied with this problem, as the quote above illustrates, but he manages to find support for a more appropriate account of action elsewhere in the first critique. And obviously, different passages in Kant pull in different directions. However, given the central thesis of transcendental idealism, it seems that the causality of freedom can have no impact on particular, empirically recognisable actions carried out in the flow of time.[51] In that case it becomes hard to see how actions in our sense of the word, i.e. making things happen in the world, can be recognised as free.

Kant's approach to agency is coloured by the *methodological solipsism* of his approach. By methodological solipsism I mean an understanding of human cognition as principally independent of intersubjectivity or of a community of language users.[52] Kant's philosophical system is construed within the bounds of a philosophy of mind, where a common language is not yet recognised as a condition of rationality and agency. Within these bounds it is hard to deliver a basis for acknowledging other people as capable of free agency. Kant obviously considered it possible to hold other people responsible, cf. again KrV B 583, where he writes that we rightfully blame another agent for a malicious lie

> als ob der Täter damit eine Reihe von Folgen ganz von selbst anhebe.

However, there is a reason for his use of the phrase "as if". The possibility of blaming another agent may not *actually* be open for Kant, since transcendental idealism seems to imply that the subject can recognise the actions of other people solely as phenomena. Free agency cannot be recognised "from the outside", but it is hard to see what other access we may have to other people's actions. Furthermore, questions of unity and plurality are inadmissible in the realm of noumena, this realm being logically prior to the categories. The question of whether free

[51] One way to deal with this problem is to introduce a differentiated concept of time, as Peter Rohs does. He uses John McTaggard's concepts of an A-series and B-series of time to distinguish between A-causality and B-causality. Actions cannot be seen as subjected to B-causality, Rohs argues, since freedom is compatible only with A-causality (cf. Rohs 2003, p. 40 f.).
[52] Cf. Granum (Carson) 1999 and 2000.

subjects other than myself exist, is therefore – strictly speaking – devoid of meaning within the Kantian system.[53]

Subjectivity is a fundamental topic in the *Critique of Pure Reason*. Kant emphatically rejects the possibility of identifying a Cartesian thinking substance, i.e. a permanent, un-compounded, individual subject of thought and experience, capable of existing independently of body or matter. The source of this illusion is, as Strawson puts it, that "the unity of experience is confused with the experience of unity".[54] Kant calls this the fallacy of paralogism, resulting in the metaphysical supposition that we may have knowledge about "the soul".[55] The unity of experience, an "I think", which must (implicitly) accompany all my representations, is a necessary condition for the possibility of knowledge.[56] But this does not mean that we can experience a *unity* – such as an immaterial object – undertaking this synthesis.

On the other hand, we need a positive connection between the transcendental subject for experience – the "I think", which is not a concrete person, but an "extensionless point"[57] – and the empirical subject viewed as an identifiable unit among other phenomena. Not least for the purpose of morality – our capability to "blame the agent" as well as the agent's capability to defend herself – we need to justify a positive connection or "identity". The concept of a causality of freedom may be seen as an attempt to solve this problem by stating that the subject of experience is an agent who produces changes in the world of experience through her actions. However, it seems to me that the problem remains at least partly unsolved if we are bound to *either* experience the action in the temporal flow of events, in Kant's words:

> Daher kann keine gegebene Handlung (weil sie nur als Erscheinung wahrgenommen werden kann) schlechthin von selbst anfangen (KrV A 553/B 581),

or to think of the action as the timeless product of the intelligible subject:

> In Ansehung des intelligiblen Charakters (…) gilt kein Vorher, oder Nachher, und jede Handlung, unangesehen des Zeitverhältnisses, darin sie mit anderen Erscheinungen steht, ist die unmittelbare Wirkung des intelligiblen Charakters der reinen Vernunft. (Loc.cit.)

---

[53] To pursue this along the line of Alan Wood: That any *one* intelligible subject should be the cause, not only of its own empirical subject, but in addition of a complete natural history is only coherent given that there is no room for a plurality of subjects in the Kantian universe. Otherwise we would have the problem that every subject causes its own natural history, which would amount to an extreme relativism far from what Kant intended.
[54] Strawson 1966, p. 37.
[55] Cf. KrV B 399ff.
[56] Cf. KrV B 132.
[57] Cf. Wittgenstein, TLP 5.64.

There is a thoroughgoing dichotomy in Kant's philosophy: On the one hand we have the concepts of freedom and a timeless, noumenal realm – and on the other the concepts of nature, space-time and the phenomenal realm. Actions – in our modern sense of the word – do not fit neatly into either of the categories.

In spite of the shortcomings of Kant's theory of freedom, I strongly disagree with Jonathan Bennett, who judges it to be "worthless".[58] Kant's theory is ground-breaking, not least for the reasons Bennett himself mentions, such as Kant's ability to ask the right questions and

> his sensitivity and subtlety of response to conceptual pressures and tensions. (Op.cit. p. 108)

Kant leads the way for later attempts to clarify the relation between the subject and the world by stressing that we must avoid the "Scylla" of rationalism as well as the "Charybdis" of empirical psychologism.

What I would like to accentuate as a guiding line to any theory of free agency is the way Kant indicates a conceptual bond between freedom and the inescapable first-person perspective of a consciousness-in-acting:

> Ein jedes Wesen, das nicht anders als unter der Idee der Freiheit handeln kann, ist eben darum, in praktischer Rücksicht, wirklich frei. (GSM 448)

Questions of freedom relate only to those of us who are *aware* of our own ability to act. To us, moreover, the presumption of freedom cannot be dodged.

However, since Kant remains within the constraints of a philosophy of mind, there are certain consequences of this insight that he is unable to perceive. Therefore, I think that leaving the bounds of a methodic-solipsistic philosophy of mind and turning towards a pragmatic philosophy of language will help us on the way towards a more adequate understanding of free agency.

To sum up this chapter, I see Kant's problem as one of dualism, although he attempts to avoid it. The system of two worlds/realms/aspects blocks the possibility of understanding free agency as making things happen in the world. His concepts of freedom and free agency remain mentalistic. I think a crucial point is that free agency is not only recognised from the

[58] Cf. Bennett 1984, p. 102.

"inside" (and *cannot* only be thus recognised). We recognise free agency in the actions of other people as well, which explains why we may rightfully "blame the agent". Kant acknowledges this, but is unable to account for it within the limits of his philosophy of mind.

I believe Kant is right in assuming that freedom of agency cannot be proven or refuted in any theoretical sense. This is because agency is not something we *de*scribe, but something we *a*scribe, as I will argue below.[59] In the following chapter, I suggest a reversion of the Kantian order: Instead of starting out (negatively) with an examination of the metaphysical possibility of transcendental freedom, we may begin (positively) with the way we actually recognise (our own and other subjects') actions as free. It is not a matter of identifying and/or describing an inner quality, but a matter of mutual recognition and acknowledgement. In the next chapter, I try to replace Kant's attempt to analyse the concept of freedom as a counter-concept to nature with an attempt to view freedom as an analytical component of the concept of rational agency.

[59] Cf. below, Ch. 4.2.

Chapter 3: Action and causation – an equation with two unknowns

The 'freedom' (…) of a man consists in the *fact* that he acts.
(von Wright 1980, p. 78)

[D]as Verhältnis zwischen Handlung und Kausalität [sei] als
eine Gleichung mit zwei unbekannten zu behandeln.
(Keil 2000a, p. 149)

3.1 Introduction: The external view on agency

According to Thomas Nagel, the free agency problem arises whenever we attempt to view action from an objective or external standpoint.[60] From this perspective it becomes difficult or impossible to recognise individual agents as sources of actions. Instead, all actions seem to be sucked up into the temporal flow of events that make up the world. Nagel points out that determinism is a common, but not the only possible outcome when action is seen from the external perspective:

> 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 judgement the inner view of the agent rebels. (Nagel 1986, p. 110)

In other words, any form of objectivism will produce this effect. As we saw in Chapter 1, Kant established the Third Antinomy based on the irreconcilability of freedom and universal determinism, but actually nothing depends on the presupposition of universal determinism. The "great gulf between nature and freedom" springs from the external standpoint itself. However, most of us share a view of people and their actions as a part of nature's order – in this trivial sense we are all (or at least most of us are) naturalists. No world-view accepting

[60] Cf. Nagel 1986, p. 110.

that actions are events and that every event has a cause can avoid the problem concerning the status of human action. In the words of Marcus Willaschek:

> Verhalten muss natürliche Ursachen haben, und so steht man trotz aller perspektivistischen Relativisierung immer noch vor der Frage, warum diese Ursache gerade zu einem solchen Verhalten geführt haben, das sich (…) durch die Wünsche und Überzeugungen des Handelnden begründen lässt. (Willaschek 1992, p. 143)

Any objectivation of agency will produce the free agency problem in the sense that it alienates the agent from her own action. However, as Nagel assures us, the "inner view" of the agent rebels against such objectification. This will for the most part resolve any practical discomfort, whereas the theoretical free agency problem will continue to arise whenever we attempt to *explain* phenomena such as reasoning, language use, or decision making in an objectivistic manner, e.g. by indicating how they enter into a causal chain of events. In this chapter I will look at the dispute within theory of science over whether actions might be said to have *reasons* or *causes* (3.2), and Donald Davidson's theory of action, which states that reasons *are* causes (3.3). From this point of departure I discuss the nature of causal relations (3.4), before I defend an interventionist account of the relation between causality and agency (3.5) in spite of certain difficulties when it comes to explaining "inner action" (3.6). Finally (3.7), I argue that the proposed concept of intentional action has the presupposition of freedom as one analytical component. This means that if we can make room for a concept of intentional action in our world-view, there is no free agency problem over and above this. An intentional action is, by conceptual necessity, a free action. In this I subscribe to Geert Keil's vonwrightian statement, that

> bereits unsere gewöhnlichen Handlungsbeschreibungen eine massive Freiheitsmetaphysik *implizieren*. Wenn es ernst gemeint ist, dass Menschen Handlungsvermögen besitzen, gibt es nicht noch einmal ein *separates* Freiheitsproblem. (Keil 2000a, p. 12)

That ordinary language *implies* freedom of agency does not amount to a theoretical *proof* of freedom. It does, however, as Keil argues, indicate why the demand for such a proof is both misplaced and misunderstood.

3.2 The reasons vs. causes debate

Attempts to theoretically explain action initiated the so-called *reasons vs. causes* debate in the modern theory of science. The central question of this controversy is whether actions may be said to have causes in the way that (other) natural phenomena do. The participants in the debate may be roughly divided into a "positivist camp", which argues that all events, including actions, should be explained according to a common, scientific scheme, and a "humanist camp", which argues that actions demand a totally different mode of understanding than (other) natural phenomena, and are better understood with reference to reasons than to causes.

This debate peaked with the Hempel-Dray controversy in the 1960s, but reaches back at least into the first half of the nineteenth century, when Auguste Comte founded the positivist movement as a protest against speculative romanticism. In 1843, John Stuart Mill, in *A System of Logic*, formulated a view of the human sciences as "immature" compared to the natural sciences, a view that has enjoyed widespread support right up until the present day:

> The backward state of the Moral Sciences can only be remedied by applying to them the methods of Physical Science, duly extended and generalised. (Mill 1843, p. XV)

In the opposite camp, the hermeneutic theorists, notably Schleiermacher, defended the humanities (*Geisteswissenschaften*) as "scientific" in their own right. Towards the end of the nineteenth century, Willhelm Dilthey attempted to establish a methodological foundation for the "human sciences" (e.g. history, law, literary criticism) that showed them to be distinct from, but equally "scientific" (*wissenschaftlich*) as, the "natural sciences" (e.g. physics, chemistry). Dilthey's work built on the distinction between *explanation* and *understanding* drawn by his colleague Johann Gustav Droysen. In the natural sciences, Droysen argued, we seek to *explain* phenomena in terms of cause and effect, or the general and the particular, whereas in the humanities we seek to *understand* phenomena in terms of relations between the part and the whole. The efforts of Schleiermacher, Droysen, Dilthey, and others established the humanities as theoretically and methodologically independent and separate from the natural sciences.

In the wake of logical empiricism, the debate over the relationship between the humanities and social sciences on the one hand and the natural sciences on the other is once again ignited. The logical positivists (Neurath, Carnap, Morris) renewed Comte's efforts to establish a unified science. In 1942, Carl G. Hempel published the article "The Function of General Laws in History", launching a standard model for all scientific explanation, the deductive-nomological method. Karl Popper, although disagreeing with both Hempel and the logical positivists on certain matters, shared their ambition of developing a scientific method of explanation that would help the humanities and social sciences overcome their alleged methodological immaturity.

These were popular viewpoints in the philosophy of science in the 1940s and 50s, but were met with considerable opposition towards the 1960s. The year 1957 marks a turning point in the reasons vs. causes debate, when William Dray and G.E.M. Anscombe published their books *Laws and Explanations in History* and *Intentions*, followed by Peter Winch's *The Idea of a Social Science* in 1958.[61] These writers were all inspired by the late Wittgenstein, and pointed to a distinction between the language of natural science and the ordinary language necessary for imparting e.g. historical knowledge. Dray argued that the deductive-nomological explanation model, which he referred to as the covering law model, cannot be of much use for the historian. This is due to the specific complexity (or "density") of an historic event.

As an illustration, Dray displayed how an attempt to formulate a law covering the outbreak of the First World War in 1914 is bound to be unsuccessful. Either such a law will be too simple, hence easy to falsify ("If an archduke of Austria is murdered, a world war will break out"); or it will be too vague, tautological or in any case methodologically uninteresting ("If an archduke of Austria is murdered in Sarajevo, and if the world is on the verge of war and a sudden murder will provoke this war, *then* the First World War will break out"); or again the law will be too specific, to a degree where there can be only one instantiation of it, namely the event we are trying to explain. ("If there is a tension between two countries, and if a murder of the regent occurs in the capital of one of the countries, and so on and so forth –

[61] An important contribution was also made by Hans Skjervheim in his *Objectivism and the Study of Man* (1959), which was intended as a reply to the objectivistic ideals of Arne Næss's *Erkenntnis und Wissenschaftliches Verhalten* (1936). Skjervheim stresses the non-circumventability of the participant's first-person perspective, and may in that sense be seen as fronting similar ideas as Wittgenstein, whose *Philosophical Investigation* was concurrent with, but unknown to Skjervheim at the time. But Skjervheim's contribution – notably in the essay "Participant and Observer" (originally 1957, cf. Skjervheim 1996) – might also be viewed as an early version of "normativism" à la Robert Brandom (cf. below, Ch.4).

counting everything happening from the Napoleonic Wars until the outbreak – *then* the First World War will break out." Well and good, but we already knew that.)

Like the hermeneutic theorists before them, Dray, Winch, Anscombe and the other neowittgensteinians defended the human sciences as methodologically independent and separate from the natural sciences. The Logical Connection Argument (LCA) must be seen in this light, as an attempt to separate causes from reasons, thereby separating the explanation of natural phenomena from the understanding of intentional action. The LCA states that the connection between intentions and the actions rationalised through them is a logical or conceptual connection, unlike the contingent, empirical connection needed to constitute a cause-effect relationship. For the neowittgensteinians, this constituted a decisive argument against the causalist claim that intentional action has causes.

Donald Davidson attacked the LCA on the grounds that a conceptual connection only holds between events *under certain descriptions*, whereas causal relations hold between the events themselves. Reasons are not opposed to, but identical with causes, Davidson argued.[62] Stating the rational reasons for an action is just one way of expressing the intentional attitudes causing a certain bodily movement. In the following section I will look more closely at this connection between action and causation.

3.3 Action and causation

An appropriate start to this subchapter would perhaps be to offer a definition of the concept of action. From there I could proceed with examining how this concept is connected to the phenomenon of causation. However, this is not an obvious route. It seems that we might as well turn the table and try to develop a notion of action from the concept of causation. Geert Keil describes the attempt to explicate the relationship between causation and action as trying to solve an equation with two unknowns.[63]

Let us rather start out with an intuitive distinction between what merely *happens* – and the things people *do*, between events and actions. What distinguishes human actions from mere happenings? According to Davidson and the proponents of a causal theory of action, actions are events of a specific type. All events are caused and can be explained as

[62] I.e. reasons "in themselves" (on the "metaphysical" level) – not as linguistically articulated, *understandable* reasons (cf. below, Ch. 3.4).
[63] Cf. Keil 2000a, p. 3.

instantiations of natural laws. What distinguishes actions from other events is that we can *also* describe them as intentional, i.e. they may be viewed as caused and rationalised by reasons:

> If an event is an action, then (…) under some description it is intentional. (Davidson 1980, p. 61)

This means that a piece of human behaviour can be described as an intentional action if it can be rationally justified by identifying certain attitudes that causally explain the behaviour. The causal theory defines and explains action through certain intentional attitudes, typically consisting of a so-called belief-desire pair: 1) a pro attitude, typically a desire, towards something, and 2) a belief that a certain action will lead to the desired goal.

Davidson's famous 1963 essay "Actions, Reasons and Causes" marked another turning point in the reasons vs. causes debate, with its devastating attack on the LCA argument and the identification of reasons and causes. The causal theory of action, however, pretty soon stood in need of clarification. For an action to be caused by a belief-desire pair, it is not sufficient that the agent has the right belief and desire, and that the corresponding act follows. The action has to be caused *in the right way*, as Davidson added in a later essay:

> [N]ot just any causal connection between rationalizing attitudes and a wanted effect suffices to guarantee that producing the wanted effect was intentional. The causal chain must follow the right sort of route. (Davidson 1980, p. 78)

The paradigmatic case of a causal chain following 'the wrong sort of route' is a so-called *deviant causal chain*: A man wishes to kill another man. He brings the gun to his shoulder, takes aim, when the idea of killing the other man makes him nervous and causes his finger to twitch. The gun fires and kills the victim. Although the movement of the finger was caused by intentional states of the agent, in this case it was not caused *in the right way*. We might specify what has gone wrong by recollecting the distinction between what we do and what merely happens to us. In this case, although the agent has the appropriate attitudes, the movement of the finger is something that merely happens to him. The agent remains passively engaged in his own behaviour, rather than actively engaged. Actually, "agent" seems altogether to be a misnomer in cases like this.

But consider a slightly different example: A man wishes to kill another man. He aims and shoots, misses horribly, but his shot startles a bunch of wild hogs, which run wild and

trample the victim to his death.[64] In this case the man deserves the title 'agent', since he performs an intentional action; he pulls the trigger. And it is this very action that, although by indirect means, leads to the fulfilment of his intention. Still, the agent didn't intend it to happen *this* way.[65]

The basic problem with these counterexamples is that the very same belief-desire pair that explains the deviant causal chain would also explain the genuine action. In these cases, therefore, the belief-desire pair may not be used to distinguish an action from a mere happening. Being able to track the causal history of an event back to the triggering intentional attitudes of an agent is not sufficient to identify the event as an action.

The causal theory of action states that an action is a bodily movement caused by certain intentional attitudes of the agent. Deviant causal chains challenge this statement, as they show that the same intentional attitudes may also cause events that are clearly not actions. But what exactly do deviant causal chains deviate from? Geert Keil discusses this thoroughly in *Handeln und Verursachen* (2000), and concludes that the question can only be answered *ex post actu*.[66] It is impossible to specify in advance what it takes to fulfil the conditions that satisfactorily define an action; this can only be judged in retrospect. The reason is simply that real life is more "dense" than our anticipation of a certain course of events, and so

> weichen abweichende Kausalketten von demjenigen kontrafaktischen Lauf der Dinge ab, der der retrospektiven Präzisierung der eigenen Handlungsabsicht durch den Handelnden entsprochen hätte. (Keil 2000a, p. 102)

What a deviant causal chain reveals is that the definition of action offered by the causal theory is insufficient. Hence, as it has yet to come up with a satisfactory determination of what it means to "cause in the right way", the causal theory of action is incapable of giving sufficient conditions for the identification of an event as "intentionally doing H".[67]

The causal theory draws a distinct picture of the relation between intentions and actions, but without accounting properly for the active role of the agent. Agent causality has

---

[64] This example is borrowed from Keil 2000a, p. 74, but is originally from Daniel Bennett.
[65] Deviant causal chains highlight how (moral and legal) responsibility is attached to intentions as well as to action. In court, both the spastic and the horrible shot would (or at least should) be acquitted of murder, as their intention-in-action failed to achieve its condition of satisfaction. They could, on the other hand, be committed both for attempted murder – because of their *intention-in-action* (the *trying* to shoot their victim), and for conspiracy to murder – because of their *prior intention* (For the distinction between intention-in-action and prior intention, cf. Searle 2001, p. 45).
[66] Keil 2000a, p. 100f.
[67] Cf. Keil 2000a, p. 112.

been a recurring proposal as a way of securing the active role of the agent that is involved in our concept of intentional action.[68] Kant's "causality through freedom" is the archetype of agent causality. In Kant's view, the model of causal connection between two events is unsuitable when it comes to free agency. In addition to the natural causal connection that prevails between events, a separate type of connection is introduced; a connection holding not between events, but literally between agent and event. Roderick Chisholm notoriously defended a similar, "strong" version of agent causality, but there are other, more moderate ways of suggesting that the causal contribution of the agent is not reducible to event causation. Marcus Willaschek argues that the causal role of the agent does not replace "ordinary" event causality, but enters into the picture as a *further* factor necessary to explain behaviour, over and above the agent's beliefs and desires. He argues that having certain intentional attitudes is not enough, since only some of these attitudes – depending on the influence of the agent – will become causes of behaviour:

> [T]he causal role of the person consists in her influence on which of her beliefs and desires will be causally effective. (Willaschek 1998, p. 179)

Willaschek bases a strong, Kantian defence of autonomy on the idea that agents may choose to acknowledge or reject certain motives as entering into their "space of practical reason".[69] This "rational integration of a person's motives" is what makes up a person's *character*, the sought-for "extra" causal factor in addition to intentional attitudes.

The modern concept of cause, however, is not the multiple, Aristotelian one, but has the limited meaning of *causa efficiens*, a triggering cause. A cause explains a change in a condition, i.e. why something happened *at this particular moment in time*. Introducing the character of the agent as a causal factor cannot, in my view, get clear of what Geert Keil calls "Broad's datedness objection":[70]

> The acting person existed before her action took place, and she will live on afterwards. For this reason she cannot be, in a literal sense, the triggering cause of any occurrence. (Keil 2007a, p. 22)

[68] Cf. Willachek 1998, p. 177f.
[69] Op.cit. p. 189. Willaschek's argument may be interpreted as a version of what Henry Allison calls Kant's "Incorporation Thesis", cf. above, Ch. 2.3.
[70] Named after Charlie Broad, cf. Broad 1952.

Causal relations exist between events, understood as changes in time,[71] while the character of an agent is of a more constant kind and must be regarded as an entity of a different category. Willachek deserves credit for highlighting the Kantian aspect of an agent's character as an important factor in analysing and explaining behaviour. And clearly, the character is decisive for what motives we allow to enter into our "space of reason". However, it seems superfluous as well as misleading to qualify the relation between an agent's character and her own motives as a *causal* connection. This does not imply that causal relations are not involved in the shaping of our character, only that these relations may be explicated solely in terms of "ordinary" event causality.

3.4 The nature of a causal relation

The reasons vs. causes debate dealt with whether intentional agency may be explained in terms of causality or not. The nature and epistemological status of the causal relation was more or less left undiscussed throughout this debate. That is, the nomological character of causal connections was largely taken for granted by both sides of the debate – while the central question was whether the connection between intentions and actions may be said to be of a causal kind. Davidson denied the existence of laws of nature governing intentional actions,[72] a view that won widespread support. With the exception of Paul Churchland, few have attempted to formulate strict laws of action since Davidson presented his *Principle of the Anomalism of the Mental*. At the same time he presumed the possibility of causal connections between mental and physical events, e.g. intentions and actions (the *Principle of Causal Interaction*), and argued that wherever there is causality, there are strict, deterministic laws (the *Principle of the Nomological Character of Causality*).

Davidson balances these principles by drawing a line between two levels: Causal laws hold strict and exceptionless reign over the metaphysical level of events "in themselves", i.e. events regarded as concrete entities with an extension in space and time,[73] and regarded as neutral with respect to the possible descriptions they may be subsumed under. At the explanatory level, however – where the same events are described in a non-physical language – "lawlessness" prevails. Davidson assumes a *token-token identity* between mental and

[71] Cf. Davidson 1980, p. 161f. and Keil 2000a, p. 168.
[72] Cf. Davidson 1980, p. 208.
[73] Or, as Davidson would perhaps prefer to put it, with a specific position in the world's "causal chain".

physical events: A mental event (regarded as a singular token) is identical with some physical event (token). This is a *non-reductive* theory in accordance with the principle of the anomalism of the mental. In order to formulate psychophysical laws, on the other hand, we would have to assume a *type-type identity* between mental events of type A and physical events of type B, which would imply that it would be possible to perform a reduction of mental events to physical ones.

Deviant causal chains turned out to be a crux for the causal theory of action, and refuelled an old question of what causal relations really are and how we recognise them. Why do I perceive my persistent tapping on the keyboard as the cause of letters popping up on my screen, instead of comprehending the two events simply as concurrent? According to Hume, an event A is the cause of another event B iff B follows immediately upon A and iff every event of type A is followed by an event of type B. The bonfire causes the water in the kettle to boil, since each time a kettle of water is placed over the fire, it will be brought to boil (provided that it is heated for long enough, there is sufficient fire etc.). However, Hume denied any *necessary connection* between cause and effect, blaming our idea of such a connection on habit, a feature of the human psyche supplying us with the (subjective) expectation that certain types of events are followed by certain other types of events. John Stuart Mill pointed out that Hume's analysis, known as *the regularity theory of causality*, was too weak to capture the essence of the matter.[74] If two events are to be seen as cause and effect relative to each other, it is not enough that actual events of type A are consistently connected to actual events of type B. Night and day do in fact always follow each other, but it is not unthinkable that a situation could arise where night continued infinitely. Consequently we do not understand the phenomena of day and night as causally connected. For something to be recognised as a causal connection, it must be seen as a *necessary* connection in one sense or another: Not only did event A actually follow upon event B, but it *had to* follow given the initial conditions. The challenge is to explain this necessity, and it is among the attempted answers to this challenge that we find Kant's analysis of causality (of nature) as a condition for the possibility of experience, i.e. as an *a priori* category of reason.

A prevailing view since the 1960s is that we must separate between the meaning of causal statements and their truth conditions. This distinction has been the basis of different attempts to explain causality, whether in terms of necessary conditions (Reichenbach), INUS-conditions (Mackie), or by use of counterfactual conditionals (Lewis). However, the nature of

[74] A further point against Hume: His psychological explanation, like the "conditioning" of Pavlov's dogs, actually *presupposes* causality.

causal connections seems to defy a final and satisfactory determination. This prompted some philosophers, prominently Bertrand Russell, to suggest rejecting the concept of causality altogether. Why not ditch the idea, not least since a "mature" science seems to do just fine without it:

> The law of causality, I believe, like much that passes muster among philosophers, is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm. (Russell 1913, p. 171)

Against Russell's view it may be argued that even if "the law of causality" (whatever that is)[75] has no place in modern science, and even if the notion of cause disappears from the language of physics as science develops,

> the fact remains that causal thinking, as such, has not been exorcised from science – and that therefore the philosophical problems about causation continue to be central to the philosophy of science. (von Wright 1971, p. 37)

The concept of causality remains relevant to and controversial in philosophy, along with the debate over how a causal connection is to be understood. In the centre of the discussion we find the much disputed *Principle of the Nomological Character of Causality*. Geert Keil argues against this principle in *Handeln und Verursachen* (2000), by performing a *reductio ad absurdum* of causal laws.[76] What Keil attempts to show, in short, is that the *ceteris paribus*-clause necessary to render a causal law true must be interpreted as an indexical phrase, thereby limiting the scope of the law to applying to "this instance, here, now". But a law that can only have one instantiation is no law in the ordinary sense of the word. He stresses, however, that the alternative to strict causal laws is not to perceive the world as a total mess:

> Es steht nicht Chaos gegen Ordnung, es stehen nichstrikte, störbare Regularitäten gegen ausnahmslose, und die Behauptung lautet, dass die ersteren für die Einheit der Erfahrung genügen müssen. (Keil 2000a, p. 350)

[75] Cf. Von Wright 1971, p. 35, Keil 2000a, p. 166f.
[76] Cf. Keil 2000a, p. 234.

Keil proposes, in other words, that we put up with a "causality without strict laws".[77] More generally, however, he links the concept of causality to our ability to *make things happen*. This way of thinking is rooted in the interventionist theory of von Wright's *Explanation and Understanding* (1971), as well as in the counterfactual analysis of causality by his contemporary David Lewis.[78] Keil ties these theories together, basing our concept of causality in our capability of counterfactual reasoning, i.e. our ability to *do otherwise*. The interventionist account of causality will be the topic for the next section.

3.5 Making things happen – and making sure they don't

Throughout the *reasons vs. causes* debate, as well as according to the causal theorists of action, the existence of causal relations between events in the world was largely taken for granted. The question has been whether intentional action should be understood as instantiations of such causal relations, or as somehow "beyond" causality – at least in surpassing ordinary event causality. Like the causal theory of action, the interventionist theory of causality seeks to clarify the relationship between action and causality. What von Wright basically does, however, is to reverse the relationship between the two. Instead of asking how causality can be used to explain action, he illustrates how intentional action is a necessary element in the explication of causal relations:

> [T]he idea of a causal or nomic relationship can be said to depend on the concept of action. (von Wright 1971, p. 72)

[77] Without delving too far into this debate, it seems to me that although Keil's critique against the nomological character of causality is well-founded, it certainly seems to be the case that we *expect* – and perhaps even *have to* expect – exceptionlessness from (at least some of) the laws of nature, including (some) laws of succession that specifically are under attack by Keil. Keil argues that causality cannot be explained through general laws of nature – for that, the world is "too dense". However, the question remains: What are physicists then aiming for, and according to what principle do they distinguish between mere disturbances and interruptions of the experiment (the oscilloscope tipped over, a bird snatched the falling object) and grave deviations leading to theoretical reconsiderations. Wellmer argues that a disturbance does not qualify as a falsification until it is brought under experimental control, i.e. is reproducible (cf. Wellmer 1967). The reaction to deviations in experimental research is in any case not "Oh, well, it didn't work out well this time", but rather that *either* something has gone wrong with the experiment or the measuring of values, *or* that we must adjust certain law(s), principle(s), or initial conditions at the basis of the experiment.

[78] Cf. David Lewis, "Counterfactuals" and "Causality" (both 1973), reprinted in Lewis 1986.

What we have primary access to, von Wright says, is our own ability to *do* things. By doing something, we bring something about, and this relation forms a crucial precondition for our conception of causal relations. Thus:

> [W]e cannot understand causation, nor the distinction between nomic connections and accidental uniformities of nature, without resorting to ideas about doing things and intentionally interfering with the course of nature. (von Wright 1971, p. 65f.)

Von Wright might be said to move from a *causal theory of action* to an *agency theory of causality*.[79] This is proposed as a general theory of causality in nature. Far from proposing a theory of agent causation, i.e. a separate species of causality unique to intentional action, he introduces a general account of event causation that is derived from our conception of ourselves as agents:

> To say that we cause effects is not to say that agents are causes. It means that we do things which then as causes produce effects. (von Wright 1971, p. 69)

Our understanding of a causal relation involves counterfactuals, von Wright continues, which brings to mind Mill's old critique against Hume that causal relations go beyond mere regularities. To understand an event as the cause of another event involves more than seeing them as following each other on a regular basis. It involves assuming that had the first event not happened, the second would not have happened either. Von Wright puts it thus:

> The assumption (hypothesis) that the concomitance of p and q has a nomic character contains more than just the assumption that their togetherness is invariable. It also contains the counterfactual assumption that on occasions when p, in fact, was not the case q would have accompanied it, had p been the case. (von Wright 1971, p. 72)

But how do we justify counterfactual assumptions? As von Wright points out, it is logically impossible to verify individual counterfactuals. The world consists of everything that is the case, leaving no room for a certain verification of what *would* have or *could* have been the case. However, through our ability to interfere with the way of the world, we can come very close:

[79] Cf. Keil 2007a, p. 20.

> Assume that p is a state of affairs which, on some occasions at least, we can produce or suppress 'at will'. This presupposes that there are occasions on which p is not already there and, we feel confident, will not come to be (on the next occasion), unless we produce it. Assume there is such an occasion and that we produce p. We are then confident that had we not done this, the next occasion would have been one when p was not there. But in fact it is one when p is there. If then q too is there, we should regard this as a confirmation of the counterfactual condition which we could have affirmed had we not produced p, viz. that had p which was not there been there q would have been there too. (von Wright 1971, p. 71f)

It is along this route that von Wright connects causality with action. Our direct experience with making things happen – as well as making sure some things do *not* happen – is the road to cognition of causal connections. Geert Keil comments that this counterfactual element departs radically from empiricism:

> What would have been the case is not part of the observable world. Interventionism demonstrates where science has to go beyond mere observation, and has to take agency and experimentation seriously. (Keil 2007a, p. 24)

Von Wright himself dubs his approach an *experimentalist* notion of causality, as it is based on how we, as agents, can manipulate our surroundings:

> It is *established* that there is a causal connection between *p* and *q* when we have satisfied ourselves that, by manipulating the one factor, we can achieve or bring it about that the other is, or is not, there. We usually satisfy ourselves as to this by making experiments. (von Wright 1971, p. 72)

An interventionist or experimentalist approach to causality offers a solution to Hume's infamous problem, concerning how "one event follows another, but we never can observe any tie between them".[80] *Pace* Hume, it is not a matter of *observing* such a tie, but rather of *pulling* it – and sometimes abstaining from pulling it. Action, and forbearance, the "passive counterpart to action,[81] allow us to explore the tie between causes and effects, whether as a basis for articulating causal laws or non-strict, disturbable regularities.[82]

[80] Cf. Hume 1748, Section VII, Part II.
[81] Cf. von Wright 1971, p. 90.
[82] Cf. Keil 2007a p. 25f. Von Wright's interventionist theory is guided by an attempt to establish exceptionless laws of nature, whereas Geert Keil denies the existence of such laws. However, Keil's approach is still "von Wrightean in spirit" (cf. Keil 2007a, p. 32), in the sense that it supports the assumption of "a *conceptual* link between agency and freedom" (ibid.).

Finally in this paragraph I will consider two problematic features of the interventionist theory of causality. The first is the accusation of *circularity,* which has been raised from several quarters: How can we deduce a general conception of causality from the way intentional action brings about (i.e. causes) changes in the world without already having access to a general concept of causality? A second and – as I will attempt to show – connected accusation against this version of interventionism is that it is a case of *anthropomorphism*, in the sense that reality is being interpreted exclusively in terms of human experience.[83]

As regards the first accusation, von Wright himself admits to a certain circularity in the argumentation:

> No proof can decide, I think, which is the more basic concept, action or causation. One way of disputing my position would be to maintain that action cannot be understood unless causation is already intelligible. I shall not deny that this view too could be sustained by weighty arguments. (von Wright 1971, p. 74)

The notion of *bringing about* is vital to von Wright's explication of causality, and this notion clearly has causal connotations. The question which of the two concepts – action or causation – has explicative primacy seems hard to answer conclusively. Geert Keil blames this on the fact that we are dealing with "basic concepts of our descriptive metaphysics".[84] Such basic concepts are in general connected with each other through mutual implications – Keil mentions "time" and "change" as another example of two concepts linked together this way. It must be possible, Keil argues, to allow for the introduction of basic, metaphysical concepts in *constellations* without being accused of circularity:

> Bei Zirkeln, die offenkundig unvermeidlich sind, muss es sich um hermeneutische handeln. (Keil 2000a, p. 415)

And hermeneutic circles he adds, quoting Heidegger, we should not seek to escape, but to enter in the right way.[85]

---

[83] James Woodward's *Making Things Happen* (2003) is an example of an interventionist theory of causality which defines intervention without reference to free human agency, and in this way attempts to avoid accusations of anthropomorphism. However, Woodward does relate causation to experimentation and counterfactuals. These are both concepts which, as far as I can understand, are conceptually linked to intentional agency.

[84] Cf. Keil 2000a, p. 415.

[85] Loc.cit. Cf. *Sein und Zeit* § 32.

Keil concludes that neither of the concepts may be seen as primary to the other. However, to me it seems intuitively tempting to analyse the conceptual interdependence further, breaking it down to:

1) The *ontological priority of causality: T*he *existence* of causal relations between events in the world must be conceived of as prior to and independent of intentional agency, and;
2) The *epistemological priority of agency*: Our *knowledge* of causal relations is secondary to, and dependent upon agency, i.e. our ability to intentionally interfere with a course of events.[86]

What status the concepts are given depends upon other basic assumptions from different philosophical viewpoints. Habermas has argued that naturalists tend to focus on ontological questions.[87] Accordingly, they will perhaps tend to assign a more basic status to causal relations. On the other hand, more Kantian or pragmatically minded philosophers will tend to focus on how we as intentional agents are actively involved in the cognition of the world.

Conceptual interdependence is supported by the etymological connections between the concepts of cause and action.[88] The notion of cause has its heritage in the language of law and justice, a circumstance that several authors – among them Russell and Quine – saw as an incentive to banish the concept from science altogether.[89] Causality certainly has anthropomorphist connotations, in the sense that it involves interpreting a natural series of events in terms of intentional action, e.g. characterising the cause as "producing" or "bringing about" the effect, causal powers as "operating" in the world, etc. However, I guess that none of the authors mentioned here would deny that experimentation has a central role in the attaining of scientific knowledge. And experimentation arguably has assumptions about causal connections as its basis. Russell may be right that scientific laws move away from the language of causality; as a matter of fact, scientific language in general moves away from ordinary language. Nonetheless, I would argue that science depends upon ordinary language

[86] Cf. my discussion of Habermas' division of ontological and epistemological priority below, Ch. 6.6.
[87] Cf. Habermas 2006, p. 688; Cf. below, Ch. 7.1.
[88] Cf. among others von Wright, cf. 1971, p. 64f.
[89] Cf. Russell 1913, Quine 1974.

as a *starting point*, in the sense that certain basic questions must be formulated in ordinary language in order for science to *get started*.[90]

The interdependency between causality and action can be analysed according to ontological and epistemological priority, all within the limits of an interventionist theory. However, what I see as a challenge for the interventionist account of causation, a challenge that it inherits from the causal theory of action, is how to explain the relation between action and bodily movement.

3.6 Action and bodily movement

"An 'inner process' stands in need of outwards criteria", reads Wittgenstein's famous slogan from the *Philosophical Investigations* (§ 580). In connection with actions this means that we ascribe intentions, beliefs, and desires on the basis of the bodily movements of the agent. At the same time, my intentions are accessible to me in a way that they are to no one else; they are, in a certain sense, private. An appropriate answer to the question of what an action is should account for both of these features.

The causal theory of action takes good care of the "outer criteria", in defining action as a bodily movement which according to some descriptions is intentional. Davidson argues that all actions may be redescribed as "primitive actions", revealing their true identity as

> mere movements of the body – these are all the actions there are. We never do more than move our bodies: the rest is up to nature. (Davidson 1980, p. 59)

What deviant causal chains bring to light, however, is that the causal theory of action is unable to account for the second aspect of action, the "privileged access"[91] or "first-person perspective", which distinguishes the action as *mine*. Geert Keil defends the causal theory to the degree that whenever an action takes place, a bodily movement is involved, not as

[90] Furthermore, the theoretical statements of physics – and e.g. neuroscience – have to be *embedded* in ordinary language in the sense of the *performative-propositional double structure of speech* (cf. e.g. Habermas 1976, p. 334, cf. below, Ch. 6.6): The scientific propositional content must be stated, argued, maintained, challenged, criticised, denied, affirmed, questioned, etc. – and these are all performative verbs of colloquial language.

[91] The well-established phrase "privileged access" has misleading connotations, as it indicates that the difference between my access to my own intentions and to other people's intentions is that I 'observe' my own intentions from a better or more favourable position. Actually, "privileged access" is not based on criteria at all, but consists simply in having a first-person perspective, cf. below, Ch. 6.

something that is identical with the action, but as its substrate.[92] This does not require actions to be a subclass of events, however:

> Was der Klasse der Ereignisse angehört, ist nicht die Handlung, sondern ihr Substrat. Auch wenn jedesmal eine Körperbewegung – und *a forteriori* ein Ereignis – stattfindet, wenn jemand handelt: In der kausalistischen Handlungsdefinition fällt der Umstand unter den Tisch, dass Handlungen vollzogen oder ausgeführt werden. (Keil 2000a, p. 143)

To preserve the aspect that actions are *carried out*, Keil suggests that we give up the essentialist "An action is…" and replace it with "An action takes place when…"[93]

"Every time I move my arm, my arm moves", Wittgenstein commented in *The Brown Book,* regarding the peculiarity that action always seems to involve bodily movement. Causal theories of action and agency theories of causality share the view that action *necessarily* involves bodily movement. However, a category of action that does not fit neatly into this picture is "inner action", including activities such as contemplating, mental calculation, and so forth. Von Wright solves this problem by defining acts without an outer aspect as not belonging to the concept of action.[94] Still, it seems odd that an activity such as the mental solving of a chess puzzle, which undoubtedly is intentionally carried out, should not be counted as an action. In the article "Inneres Handeln", Marcus Willaschek attempts to develop a concept of action broad enough to allow for inner action, defined as

> diejenigen Vorgänge (…), die unter einen Handlungstyp fallen, dessen Ausführung kein charakteristisches (…) körperliches Verhalten involviert. (Willaschek 1992, p. 132)

Willaschek defines action by three necessary, but insufficient conditions:

1) It can not be something merely happening to the agent, it must "originate with herself";[95]
2) It must be rationalisable in light of the agent's beliefs and desires; and,
3) It must be causally related (in "the right way", i.e. not through a deviant causal chain) to the beliefs and desires of the agent.

[92] I.e. identical in the Davidsonian sense of token-token identity.
[93] Cf. Keil 2000a, p. 142.
[94] Cf. Von Wright 1971, p. 87.
[95] "von ihr selbst ausgeht", ibid. p. 140.

He thus avoids using bodily movements as a defining feature of action. However, for the definition to work, we must be able to pick out an action from something merely happening to the agent. It turns out to be hard to analyse the way in which an agent can acknowledge that an inner action "originates with herself". This leads Willaschek to suggest the following foremost distinction between inner and outer action:

> [D]ass nämlich innere Handlungen von niemandem außer dem Handelnden selbst unmittelbar beobachtet werden können. (Op.cit., p. 149)

Against this I would argue that actions – whether inner or outer – may never be "immediately observed" by the agent herself. An action is carried out – or *performed* – from a first-person perspective, and can only be *observed* from another perspective, whether by other agents or the person herself, *ex post actu*.

Wittgenstein's anti-mentalistic demand for "outer criteria" suggests that the conceptual tie between bodily movements and action must be preserved, even in the cases where bodily movements only take place within the brain.[96] However, the dependence between actions and bodily movements should be formulated in a way that does not imply that the agent always in some way "carries out" a bodily movement. When I walk, it seems appropriate to say that I intend a physical course of events to happen. On the other hand, if what I do is to concentrate on my chess puzzle, the movements that take place in my brain, or for that matter the wrinkling of my forehead, are not relevant to or in any way a part of my intention.

3.7 Free agency – a pleonasm

In Norwegian, the term for the main concept discussed throughout this chapter is *handling*, and the German word is *Handlung*. The etymological source of these words is the word *hand*, which already suggests that what we are talking about is *human* action. After all, only humans, or very human-like animals, have hands. The English term action (from the Latin noun *actio*) is more general, referring to all sorts of activities, of which human action is a subcategory. Intentional action, in turn, is a subcategory of human action, a term that in some

[96] Wittgenstein's demand is connected with the *learnability* of psychological predicates such as the pain predicate, and thus with the Private Language Argument, cf. below Ch. 5.3. It can therefore be argued that "inner action" must be viewed as parasitic upon directly observable behaviour, in the sense that a situation where "inner actions" were the only ones carried out would be unthinkable.

uses includes also unintentional behaviour such as sneezing and seizures; in short, everything we *do*. Handling/Handlung, however, is not fitting for everything we do, but is more clearly restricted to the things we do *intentionally*, whether with a prior intention or with an intention-in-action. In that sense the Norwegian/German term fits better to this problem area. However, in this context, when discussing freedom of agency, I am of course referring to the narrower sense of the term action.

In his essay "Freedom", Thomas Nagel distinguishes "three problems about action":[97] One general, metaphysical problem of the nature of agency, one problem of autonomy (focusing on my own freedom) and one problem of responsibility (focusing on the freedom of other people):

> We may act without being free, and we may doubt the freedom of others without doubting that they act. What undermines the sense of freedom doesn't automatically undermine agency. (Nagel 1986, p. 111)

I have argued that the first two problems collapse into one. If we consider freedom as an analytical implication of the concept of intentional action, an account of intentional agency is tantamount to an account of free agency. In von Wright's words:

> [T]he concept of an action, the ascriptions of action to an agent, belong to discourse in which 'free will' is taken for granted. (von Wright 1980, p. 78)

In other words, the concept of free agency constitutes a *pleonasm*, an excessive use of words in order to describe a phenomenon or express an idea.

In line with Wittgensteinian anti-mentalism, I have argued that we ascribe intentional action on the basis of "outer criteria." This means that – notwithstanding the first-person perspective through which the agent in a certain sense is "privileged" with regards to her own intentions – both the intentions and the actions are (and must be) intersubjectively accessible to other agents. In other words: I always already recognise myself as able to act intentionally, and I can equally ascribe this ability to others. This means that Nagel's three problems merge into one, concerning the nature and epistemology of intentional agency.

Nevertheless, the fact remains that we may be fooled in each individual case: What appears to be an action on the basis of all "outer criteria" may turn out to be nothing but a neurological phenomenon, like a twitch of the finger. When we assume that a person has

[97] Cf. Nagel 1986, p. 110f.

acted intentionally, we ascribe certain qualifications to the person, notably the ability to think rationally about a certain course of events and make a decision. It appears that such an assumption is not based on observation and experimentation alone, but has a distinctive *normative* character. The normative aspect of the concept of action will be the topic for the next chapter.

Throughout this chapter I have to a large extent avoided the discussion as to whether causality is nomological or not, i.e. if the connections between causes and effects can be formulated as strict laws or should be perceived merely as non-strict regularities. The reason is that I do not find this to be of decisive importance to the main topic of the chapter, namely the explanatory relationship between action and causality. In recollection of Thomas Nagel's point from the beginning of this chapter, I think the question of causal *determination* is a subsidiary part of the free agency problem.[98] This problem arises, according to Nagel, whenever we attempt to "view action from an objective or external standpoint."[99] Therefore, in my attempt to dissolve the free agency problem, I continue my investigation of a non-objectivistic concept of action in the following chapter.

[98] Although, as I will argue below, determinism in at least some of its forms seems to create an additional problem in this debate – cf. below, Ch. 7.2, 7.3.
[99] Nagel 1986, p.110.

Chapter 4: A normative approach to action

[T]he normative question is one that arises in the heat of action.
(Korsgaard 1996, p. 91)

Indem man etwas tut, antwortet man auf die ethische Frage.
(Rödl 1998, p. 53)

4.1 Introduction: Action and normativity

Norms and rules are important in the life of a language user. That norms and rules can be obeyed or broken belongs to the very essence of these concepts, as Wittgenstein has pointed out.[100] According to a common view, the distinctions between right and wrong, correct or incorrect, true or false do not apply outside the domain of rational agency. Our actions are judged as more or less right, good, or fitting, as opposed to natural courses of events. According to this view, we cannot reasonably make a claim about a stone that it *shouldn't* roll or that it *ought to* have taken a different route. If a stone rolling down a hill turns out to deviate from the route we anticipated for it, this would either mean that our calculations were wrong, or that we are wrong about the laws of physics. The stone is never to blame. On the other hand, when actions deviate from a given set of behavioural rules, this does not necessarily render the norms invalid: It may be that the agent, not the rule, is wrong.

In this chapter I take as my point of departure that a distinctive feature of being a language user and a rational agent is to be subject to normative standards. Through presentations of what Robert Brandom and Sebastian Rödl have written about this topic, I attempt to give an account of normativity as a distinguishing feature of rational agency. I follow Brandom and Rödl in seeing non-verbal intentional actions as parallel to assertions. Therefore, since the normativity of language use is well explored, I attempt to examine

[100] Cf. e.g. PU § 202.

rational agency from this angle. A non-verbal action and an assertion both have a normative *status*, i.e. they may be judged as right or wrong, and both are expressions of normative *attitudes*.[101]

In the following section (4.2), I pursue what a normative approach to action means. I contrast it with another view, viz. that of action as production, and attempt to analyse these two approaches as two traditional lines of thinking about agency. In 4.3, I explore a contemporary example of a normative approach to action, namely Robert Brandom's. Another current writer on normativity and action – with many similarities to Brandom – is Sebastian Rödl, whose ideas I examine in 4.4.

4.2 Action as production vs. action as expression

Normativity is arguably a distinguishing mark of the *mental*; of concepts and of mental properties and abilities. These concepts, properties and abilities are expressed through a normative *practice*, or as Rödl puts it:

> Geistiger Gehalt ist normative Praxis. (Rödl 2000, p. 762)

Understanding mental content as essentially expressed in intentional states and acts that have a normative status is an approach reminiscent of Wittgenstein in his later works. Robert Brandom describes Wittgenstein's approach as the linguistic explication of an essentially Kantian point, namely that

> to take what we do as judging and acting is to treat it as subject to certain kinds of assessments as to its correctness: truth (corresponding to the world) and success (corresponding to the intention). (Brandom 1994, p. 13)

Saying that someone has performed an intentional action is not simply to assert that certain movements have taken place in the world. It is not a pure description, but involves a normative element.

In the following I propose that agency is to be understood as a normative concept, in the sense that being an agent is equivalent to being bound by norms. Another way to state this

[101] Cf. e.g. Brandom 1994, p. 229ff., and Rödl 2000, p. 767 and 1998, Ch.3.

point is to say that agency, and agents, belong within the "space of reasons". Wilfrid Sellars introduces the concept of a logical space of reasons in order to illustrate how *knowledge* is a fundamentally normative concept:

> [I]n characterizing an episode or a state as that of knowing, we are not giving an empirical description of that episode or state; we are placing it in the logical space of reasons, of justifying and being able to justify what one says. (Sellars 1956, p. 76)

John McDowell stresses that Sellars' point extends beyond the concept of knowledge to the general relationship between mind and world:

> [T]hough Sellars here speaks of knowledge in particular, that is just to stress one application of the thought that a normative context is necessary for the idea of being in touch with the world at all. (McDowell 1996, p. xiv)

Being in touch with the world means that the world somehow "opens up" to the subject. More specifically, it means that a *normative space* opens up between the subject and the world. Experience, in the Kantian understanding, involves making judgements, i.e. experiencing something *as* something. This way of being in touch with the world involves two "directions of fit" in John Searle's sense: When the subject has an experience, her intention adjusts to the world (mind-to-world direction of fit), and when she acts intentionally, the world is adjusted according to the intention of the subject (world-to-mind direction of fit).[102]

Sellars's account constitutes the basis of Robert Brandom's conception of the normativity of action:

> The normative (…) approach to intention and action is rooted in Sellars's discussion of the giving and asking for reasons for action. (Brandom 1994, p. 263)

According to this basically Kantian approach, to be an agent is to have beliefs you take to be true, for which you are responsible, and upon which you act. To have a belief is the same as to be committed to the truth of a proposition, and to be responsive to reasons for or against this belief. Thus, to act *is* to be bound by reasons.[103] In the words of Saul Kripke:

---

[102] Cf. Searle 1983, 96f.
[103] Cf. *Kants Gesammelte Schriften* (Akademieausgabe) 28.2,2; 1068 (*Religionslehre Pölitz*), cf. above, Ch. 2.1.

> The relation of meaning and intention to future action is normative, not descriptive. (Kripke 1982, p. 37)[104]

The "Kantian approach" may be contrasted with a more "Humean approach", which we can recognise in Donald Davidson's causal theory of action, as one example. From this perspective, pro attitudes such as preferences and desires are seen as basic, and as the ultimate motivation for action. Norms governing rational action are *instrumental*, in the sense that their authority stems from these pro attitudes. Brandom's project is a reversion of this approach, in the sense that pro attitudes are seen as derived from rationality, which in turn is considered basic:

> The concepts of desire and preference are (…) demoted from their position of privilege, and take their place as having a derivative and provincial sort of normative authority. Endorsement and commitment are at the centre of rational agency – as of rationality in general – and inclination enters only insofar as rational agents must bring inclination in the train of rational propriety, not the other way around. (Brandom 2000a, p. 31)

Brandom's anti-reductionist approach to agency means understanding actions as *irreducibly normative*, in the sense that they cannot be analysed as physical events caused by pro attitudes.

One alternative to a normative theory of action is the view that action is *production*. Viewing action as production means judging and valuing the action according to its outcome. The following quote by John Stuart Mill may serve as a classic example of one version of this approach:

> All action is for the sake of some end, and rules of action, it seems natural to suppose, must take their whole character and colour from the end to which they are subservient. (Mill, *Utilitarianism*, § 20)

Utilitarians defend productive success – the bringing about of results – as the ultimate and unquestionable parameter for the judgement of action. Considerations of what we *ought* to do are seen as ultimately founded in the (potential or realised) productive success the action in question may have. This suggests that an action can be identified as an action using strictly descriptive criteria, such as the achievement of some external end (which in turn, of course, may be normatively evaluated). In contrast, a normative theory of action maintains that

[104] Cf. Wellmer 2008, p. 111, and Brandom 1994, p. 656, fn. 10.

identifying something as an intentional action *is* – independently of the actual outcome of the action – to ascribe it a normative status.

The alternative to viewing action as production is a conception going back to Antiquity, stating that action is essentially different from production, in the sense that

> while making has an end other than itself, action cannot; for good action itself is its end. (Aristotle, *Nicomachean Ethics* VI. 5, 1140b5)

Actions are qualified, according to Aristotle, not as successful or flawed relating to whether they meet or fail to meet some external end, but as more or less good in themselves. Aristotle distinguishes between *poiesis*, "things made", and *praxis*, "actions done": the former is production that is judged according to rules of art, with the corresponding virtue of *techne*; while the latter is deliberative and discursive, belonging to ethical and political life, with the corresponding virtue of *phronesis*.

The modern concept of action covers a broad range, from ethical and political deeds to those we perform in the pursuit of technical or artistic goals. And of course, when using this modern concept, it makes perfect sense to speak of an action's goal, in the sense that we do certain things in order to achieve something else. However, the Aristotelian distinction between poiesis and praxis is useful in the attempt to specify in what way the concept of action is related to the concepts of rationality and normativity. An action may have certain results, but I claim that it is not *identified* (as to its type, regarding what it is) with relation to its outcome in the way production is, but rather according to its intention. Compare the following propositions:

> Lisa sneezes at the house of cards, and it collapses.
> Selma puffs at the house of cards, and it collapses.

These two sceneries have the same outcome, but our immediate assumption is that only one of the propositions describes an action. We do not confirm that an action has taken place on the basis of whether or not certain results are attained, but on the basis of a (more or less well-founded) normative ascription of rational intention. Two more propositions:

> The girls are whispering.
> The trees are whispering.

Without further context, we understand only one of these propositions as the ascription of action, and the other one as describing something in a poetic or metaphoric way. We assume that only the girls, and not the trees, are *rational*.[105]

Having a mind, or certain mental capacities, is a conceptual part of being an agent. And the mental is typically *expressed* in rational *actions*, for example through speech acts. We may distinguish between two traditional ways of explicating the concept of action:

(1) A *descriptive* concept of action: Action viewed as the *production* of results.
(2) A *normative* concept of action: Action viewed as the *expression* of rationality.

To view action as the production of results is a widespread way of looking at actions that reaches far beyond the example of utilitarianism mentioned above. We recognise it throughout the Hume-Davidsonian tradition previously mentioned, where action is seen as caused by motivating preferences and desires (pro attitudes). The causal theory of action analyses individual actions as events caused by other (mental) events – pro attitudes – which in turn causes other (mental or physical) events. Against this, in Chapter 3, I argued that what belongs to the causal nexus is bodily movement constituting the substrate of the action, while the action itself belongs to a different frame of reference.

Although a normative concept of action may be traced back to Antiquity, it is the Enlightenment's focus on the performances of the human mind that is the real source of any modern theory of the normativity of action. The preoccupation with mental acts – among rationalists and empiricists alike – prepares the ground for what Robert Brandom calls Kant's "normative turn".[106] Kant brings a normative conception of intentionality to light, as opposed to the descriptive approach of his immediate predecessors. Descartes regards having a mind, i.e. having mental representations, as a property that some things in the world exhibit, while others do not. Brandom points out that Kant draws the line differently:

> For Kant the important line is not that separating the mental and the material as two matter-of-factually different kinds of stuff. It is rather that separating what is subject to certain kinds of normative assessment and what is not. (Brandom 1994, p. 9)

[105] In some contexts, e.g. reading a certain type of fiction such as *The Lord of the Rings*, our interpretation would of course be different.
[106] Cf. Brandom 2002, p. 21.

Kant thinks of rational beings as performers of a specifically *conceptual* activity. To Brandom, one of Kant's most decisive contributions to contemporary philosophy is his introduction of the idea that rationality (and intentionality) is distinguished, not by certain descriptive marks, but by its normative character:

> One of Kant's master ideas is that what distinguishes thinkers and agents from merely natural creatures is our susceptibility to certain kinds of *normative* appraisal. Judgments and actions essentially involve *commitments* as to how things are or are to be. Because they can be assessed according to their *correctness* (truth/error, success/failure), we are in a distinctive sense *responsible* for what we believe and do. (Brandom 2002, p. 21)

The epoch-making insight that Brandom attributes to Kant is the understanding of judgements and actions in terms of the way in which we are *responsible* for them. Concepts, to Kant, are not a special group of entities distinguished by certain descriptive criteria. He understands concepts as being *rules*, i.e. specifying how something *ought* to be done. The normative significance of being in an intentional state – e.g. thinking something – or of performing an intentional action lies in its constituting the undertaking (or acquiring) of an obligation or commitment.

Following Kant, *actions* and *judgements* may be viewed as expressions of practical and theoretical rationality respectively. Actions and judgements are the sort of things for which reasons can be given and asked for. Both belong, in the words of Wilfred Sellars, within "the space of reasons". Kant ties normative statuses to normative attitudes:

> [R]ules get their normative force, come to govern our doings, only in virtue of our own attitudes. One is genuinely responsible only for that for which one *takes* responsibility; one is genuinely committed only to that to which one has committed oneself. (Op.cit., p. 219)

In other words: Rules exist only to the degree that our practice of committing ourselves exists.

Brandom points further, however, to Hegel's critique of Kantian normativity. To Hegel, what remains mysterious in Kant's conception is how we can have access to concepts, rules and norms in the first place. Hegel brings in the concept of *reciprocal recognition* to account for the accessibility of determinate norms to which we may commit ourselves:

> Hegel's idea is that the determinacy of the content of what you have committed yourself to – the part that is not up to you in the way that whether you commit yourself to it is up to you – is secured by the attitudes of others, to whom one has at least implicitly granted that authority. His thought is that the

> only way to get the requisite distance from my acknowledgements (my attitudes, which makes the norm binding on me in the first place) while retaining the sort of authority over my commitments that the Rousseau-Kant tradition insists on, is to have the norms *administered* by someone else. *I* commit myself, but then *they* hold me to it. (Op.cit., p. 220)

To Hegel, reciprocal recognition gives normativity a definite form. The strength of Hegel's approach is that normativity becomes "socialised", and thereby acquires a definite discursive character. However, Brandom points out, the Hegelian system suffers from certain inner tensions, notably the status of *Geist* or Spirit as it is at the same time irreducibly social as well as all-embracing, with nothing "outside" of it.[107]

While Hegel's contribution on top of the original Kantian *normative turn* is a *social turn*, Heidegger and Wittgenstein contribute with a *linguistic-pragmatic turn* in order to concretise the view of normativity as socially established. Heidegger's concept of *Dasein* as a genuinely linguistic form of existence and Wittgenstein's approach to language as norm-governed behaviour constitute a conception of norms as being necessarily linguistically founded. Brandom comments on why Wittgenstein should be seen as someone who carries on and develops further Kant's normative turn:

> The starting point of [Wittgenstein's] investigations is the insight that our ordinary understanding of states and acts of meaning, understanding, intending, or believing something is an understanding of them as states and acts that commit or oblige us to act and think in various ways (…) To understand or grasp such a meaning is to be able to distinguish correct from incorrect uses (…) This is one way of developing and extending Kant's point that to take what we do as judging and acting is to treat it as subject to certain kinds of assessments as to its correctness. (Brandom 1994, p. 13f.)

Kant, Hegel, Heidegger, and Wittgenstein thus developed the basis for Brandom's normative theory of language and action, a theory that we will now examine more closely.

[107] Cf. Brandom 2002, p. 227ff. Audun Øfsti (2002) addresses a problem that may be related, cf. p. 104, fn 16: "Hegel hat m.E. ein ernstes Problem mit dem Versuch, gegen das Kantischen Sollen das Ideale als substanzieller Gesellschaftlicher *Zustand* zu denken. Er redet ja zuweilen [in Enz. § 513], als ob die strukturalistische Komplementarität zwischen System ('langue') und Inhalt ('parole') letzlich überschritten werden sollte, wie etwa in einem Kunstwerk, das 'seine eigene Sprache' ist".

4.3 Playing the game: Brandom on rational agency

Robert Brandom's highly influential book *Making it Explicit* (1994) offers an innovative account of language and action. As sketched above, Brandom inherits a view of rationality from Kant and Hegel, and a view of language and mind from Heidegger, Wittgenstein, Frege, and American pragmatism, particularly Wilfred Sellars. These lay the foundation for Brandom's rationalistic, pragmatistic and expressivistic theory, where normative status is seen as the defining characteristic of linguistic behaviour, and of rational agency in general:

> The practices that confer propositional and other sorts of conceptual content implicitly contain norms concerning how it is correct to use expressions, under what circumstances it is appropriate to perform various speech acts, and what the appropriate consequences of such performances are. (Brandom 1994, p. xiii)

The treatment of concepts as norms, and of rational action as normative, is a fundamentally Kantian heritage. Kant introduces a normative view of judgements and actions by distinguishing them as the sort of things we offer *reasons* for. It is in this "rationalistic" spirit that Brandom presents his discursive view on action:

> Giving and asking for reasons for *actions* is possible only in the context of practices of giving and asking for reasons generally – that is, of practices of making and defending *claims* or *judgments*. For giving a reason is always expressing a judgment: making a claim. So practical reasoning requires the availability of beliefs (doxastic commitments) as premises. (Brandom 1994, p. 81)

Action is assimilated to judgement and understood as the application of concepts. Concepts, in turn, are understood as rules determining what commitments agents and knowers undertake by their acts and propositions.

Brandom calls his approach to language *inferentialism*, as it explains concept use primarily in terms of inferential articulation. Inferentialism is contrasted to *representationalism*, the traditional approach in the philosophy of language.[108] According to the representationalist approach to language, the meaning of an expression must be explained

[108] Cf. Brandom 1994, p. xvi: "This semantic explanatory strategy, which takes inference as its basic concept, contrasts with one that has been dominant since the Enlightenment, which takes representation as its basic concept".

in terms of what it represents. In contrast, the inferentialist approach explains the meaning of an expression in terms of its role in inferential relations within language.

Brandom presents inferentialism as a broadening of the scope of representationalism. His aim is not to deny the representational dimension of language – it seems obvious that our propositions, what we say and think, are *about* something – but to point out that this dimension is best understood in light of a superior social dimension:

> The context within which concern with what is thought and talked *about* arises is the assessment of how the judgments of one individual can serve as reasons for another. (Brandom 2000a, p. 159)

In *Making it Explicit*, Brandom develops an idiom with which he can discuss not only language and meaning, but more broadly, action and rationality. The context of the discussion is *normative pragmatics*, a view of language and action as ineliminably normative phenomena:

> No attempt is made to eliminate, in favour of nonnormative or naturalistic vocabulary, the normative vocabulary employed in specifying the practices that are the use of a language. Interpreting states, performances, and expressions as semantically or intentionally contentful is understood as attributing to their occurrence an ineliminably normative pragmatic significance. (Brandom 1994, p. xiii)

The normative dimension of linguistic practice is seen as irreducible, but not as inexplicable. Making explicit what is already implicit in our practices is rather a key to the whole endeavour, as the title of Brandom's main work suggests. Linguistic norms are explicated in terms of how speech acts affect the *commitments* (and *entitlements* to those commitments) of the agents involved.

A lesson Brandom learns from Sellars is that all propositions, even seemingly non-inferential reports like "The ball is red," must be inferentially articulated in order to distinguish genuine concept use from the "mere reliable responsiveness" exhibited by automatic machinery like thermostats or noninferential reporters like parrots.[109] By uttering the sentence "The ball is red", I commit myself, not only to this particular propositional content but, by inference, to a number of further judgements like "The ball is coloured" and "Red is a colour."

Discursive commitments are essentially inferentially articulated, and they stand in inferential relation to each other: Certain things count as evidence for them; they involve

[109] Cf. Brandom 1994, p. 5. On the responsiveness of parrots, cf. Wittgenstein, PU § 346.

certain further commitments; they are incompatible with certain other commitments, and so on. Being rational is thus identified with being a participant in a complex game:

> The overall idea is that the rationality that qualifies us as sapients (and not merely sentients) can be identified with being a player in the social, implicitly normative game of offering and assessing, producing and consuming, reasons. (Brandom 2000a, p. 81)

In line with this, discursive practice is understood as the social practice of giving and asking for reasons or, as Brandom calls it, *deontic scorekeeping*:[110]

> Competent linguistic practitioners keep track of their own and each other's commitments and entitlements. They are (we are) deontic scorekeepers. Speech acts, paradigmatically assertions, alter the deontic score, they change what commitments and entitlements it is appropriate to attribute, not only to the one producing the speech act, but also to those to whom it is addressed. (Brandom 1994, p. 142)

In this way, the inferentialist approach explores how reasons are derived through discursive scorekeeping. This is also the key to explaining the objectivity of concepts. Brandom separates two species of discursive commitment:

> 1) Cognitive/doxastic commitments: "Takings-true" – acknowledging such a commitment is the same as having a belief.
> 2) Practical commitments/commitments to act: "Makings-true" – acknowledging such a commitment is the same as having an intention.

An action may thus be defined as the appropriate response to the acknowledgement of a practical commitment.[111]

The distinction between practical and doxastic commitments brings us back to the analogy between agency and perception.[112] Practical commitments form the basis of actions,

---

[110] Cf. e.g. Brandom 1994, p. 141ff. The scorekeeping metaphor is drawn from baseball, and alludes to the Wittgensteinian concept of language games. A problem with the game metaphor, I think, is that it does not capture the superiority of the phenomenon language. When playing games, we can always step out of the game and consult the rules, take a break (think of cricket) or whatever. What Brandom calls "the normative game of offering and assessing, producing and consuming, reasons" (cf. Brandom 2000, p. 81) is, however, the core aspect of being a subject, and not something we can step in and out of, cf. Wellmer 2002, p. 117: "[D]as Problem im Fall des Wortes 'Sprache' [stellt] sich ganz anders als im Fall des Wortes 'Spiel'. Denn während man sagen kann, dass sich unter den Spielen kein besonderes Spiel auszeichnen lässt, das die Bedingung der Möglichkeit aller anderen Spiele ist, so können wir doch sagen, dass nur jemand, der sprechen kann – der das Sprachspiel beherrscht –, zu dem innovativen Spiel mit den Worten 'Sprache' und 'Spiel' imstande ist".
[111] Cf. Brandom 2000, p. 84.

whereas doxastic commitments arise from observations. Brandom views action and perception as discursive *exit* and *entry* transitions – according to the way they contribute to "changing the score" of a discourse:

> 1) Perception: Discursive *entry* transition – changing the score by acknowledging certain commitments.
> 2) Action: Discursive *exit* transition – changing the score by bringing about certain states of affairs.

Rational agency can thus be defined as "playing the game of giving and asking for reasons". Participation in this game is what separates sapients from sentients:

> [We are] the ones capable of judgment and action. Not only do we respond differentially to environing stimuli, we respond by forming perceptual judgments. Not only do we produce behavior, we perform actions. (Brandom 1994, p. 8)

Brandom accentuates our capacity to acknowledge and act according to practical commitments. By adopting a discursive approach to action, he places intentionality within the limits of a specifically *linguistic* social practice. Agency is understood in terms of linguistic practice rather than the other way around. Brandom seeks thereby to provide a largely Kantian account of the will as a rational faculty. This leads him to explore the *normative* dimension of action with the analogy between practical and doxastic commitment, while the *causal* dimension is explained by the analogy between discursive entry transitions in perception and discursive exit transitions in action. These analogies makes Brandom suggest that

> the rational will can be understood as no more philosophically mysterious than our capacity to notice barns or red things. (Brandom 1994, p. 233)

Brandom's normative pragmatics ties many of the central topics of this thesis together. Intentional action is explored in light of normativity, rationality, and intersubjective language. In one sentence:

---

[112] Searle also bases his theory of intentionality on an analogy between perception (mind-to-world direction of fit) and action (world-to mind direction of fit), cf. table on p. 97 in Searle 1983.

> We are rational creatures exactly insofar as our acknowledgement of discursive commitments makes a difference to what we go on to do. (Op.cit., p. 271)

Brandom understands language as essentially intersubjective – a view I will examine more closely in the next chapter.[113] Inferentialism means that propositions have meaning by virtue of their inferential articulation, i.e. through their role in "the social, implicitly normative game" of giving and asking for reasons.

4.4 Answering the ethical question: Rödl on rational agency

In his book on self-reference and normativity (*Selbstbezug und Normativität*, 1998), Sebastian Rödl deals in an original way with the familiar philosophical problem area of self-reference and self-consciousness. He picks up the thread from authors such as Wittgenstein, Tugendhat, Castaneda, and Perry, and tries to clarify the ways in which the acting and thinking subject can refer to itself by using the term "I".

Like Brandom, Rödl defends a view of action as irreducibly normative. And like Brandom, he operates with a strong parallel between "taking true" and "making true", i.e. between beliefs and actions. Rödl equates acting rationally with "answering the ethical question". The *ethical question* can be formulated thus: "What should I do?" and is seen as parallel to the *epistemological question* "What should I believe?"

Rödl points to the logical connection between action and justification:

> Wenn man etwas tut, kann man es rechtfertigen. Das 'kann' hat hier seine logische Bedeutung. Handeln und Rechtfertigen passen begrifflich zusammen. Handlungen fallen in die richtige Kategorie, um gerechtfertigt zu werden. Daraus folgt zwar nicht, dass man sein Handeln de facto immer rechtfertigen kann. Dass man das nicht kann, ist aber (…) aus logischen Gründen die Ausnahme. Ich spreche dabei aus der Perspektive auf das je eigene Handeln. (Rödl 1998, p. 59)[114]

Rödl pinpoints the relation between self-consciousness and normativity on the one hand and rational agency on the other as a reciprocal dependence between self-reference ("I") and the concept of action:

[113] Cf. below, Ch. 5.3, 5.4.
[114] Cf. Wittgenstein, PU § 345.

> Der bewusste Selbstbezug konstituiert sich im Handeln, und man handelt wesentlich in der 'ich'-Perspektive. (Op.cit., p. 62)

To Rödl, an irreducibly normative aspect of action is displayed by this relation:

> Handlungen schreibt man sich *irreduzibel normative* zu. Normativ: Dass ich etwas tue, hat für mich die normative Bedeutung, dass ich es darin als das bejahe, was ich tun soll. Indem ich sage, was ich tue, sage ich, was (für mich, hier und jetzt) zu tun richtig ist. Ich charakterisiere mich normativ. – Irreduzibel: Ich kann aus der Aussage 'Ich tue dies' ihren normativen Sinn – dass ich darin auf die ethische Frage antworte – nicht herauslösen. Ich kann keinen Anteil ihres Gehalts isolieren. Denn ich habe keine deskriptive Perspektive auf mein eigenes Handeln. Ich kann mich auf das, was ich tue, nicht neutral beziehen, so, dass hinsichtlich der Frage 'Was soll ich tun?' nichts entschieden wird. (Loc.cit.)

Rödl approach to self-reference is to a large extent based on Ernst Tugendhat's critique in *Einführung in die Sprachanalytische Philosophie*, directed against traditional theories of meaning as "aboutness".[115] In this connection, Tugendhat discusses the relation between self-consciousness, consciousness in action and consciousness of objects. He refers to a comment by Kant, stating that concepts like *totality* and *eternity* are comprehensible only on the basis of a concept of successive action. This is why, for Kant, a concept of synthetic action – the synthesis of a multiple according to rules – constitutes the basis for the understanding of consciousness in general, or at least of the cognition of objects, which is the consciousness that Kant calls *experience*.[116] Tugendhat continues:

> Nun ist das Bewusstsein, das jemand von seinem Handeln hat und d.h. von der Regel, die er in seinem Handeln befolgt, wiederum nicht das Bewusstsein eines Gegenstandes. Kant hat also nicht nur auch Bewusstseinsweisen berücksichtigt, die nicht gegenständlich sind; vielmehr wurde für ihn ein bestimmtes nichtgegenständliches Bewusstsein – ein Handlungsbewusstsein – konstitutiv für das Bewusstsein von Gegenständen. (Tugendhat 1976, p. 83)

However, Tugendhat adds, Kant remains oriented towards objectual consciousness. The only consciousness in action that Kant speaks of is the kind that is constitutive for the consciousness of objects and their relations in space and time. He does not attempt to develop this into a general theory of consciousness in action.[117]

---

[115] "Gegenstandstheorie der Bedeutung", cf. Tugendhat 1976, p. 83.
[116] Cf. KrV, B 103.
[117] Cf. above, Ch. 2.5, and Wolfgang Becker's article, "Zum Handlungsbegriff in Kants theoretischer Philosophie", (1987).

According to Tugendhat, Heidegger's *Sein und Zeit* is the first attempt to liberate the understanding of consciousness from its traditional orientation towards objects. A tactical move of Heidegger's is to replace the concept of consciousness with a more connotation-free concept, "Ersclossenheit", in order to free himself from his teacher Husserl's idea of consciousness as intentionally directed towards objects. What Heidegger attempts is to show that self-consciousness should not be seen as a version of object consciousness. He distinguishes *Dasein* as a form of being that is constantly involved in self-interpretation.[118] Normative self-reference is placed at the very core of human existence:

> Dasein ist Seiendes, das sich in seinem Sein verstehend zu diesem Sein verhält.
> (Heidegger, *Sein und Zeit*, § 12)

In *Selbstbewusstsein und Selbstbestimmung* (1979) Tugendhat explores a concept of "Sichzusichverhaltens". His examination of how self-consciousness is expressed in language is, despite critical distance, inspired by Heidegger's *Sein und Zeit* and its thematic focus on the "meaning of being" ["*Sinn von Sein*"].[119] It is in this context that Rödl sees the concept of action as constitutive for self-consciousness. Being a subject simply *is* to act and to be conscious in acting:

> Subjektivität konstituiert sich im Handeln. Handeln ist nicht etwas, das ein Subjekt auch noch kann, keine Möglichkeit, die ein Subjekt auch noch hat, sondern: Als Subjekt dasein heißt handeln. (Rödl 1998, p. 64)

This normative approach to self-consciousness and action is contrasted with what Rödl calls *the metaphysical perspective*, defined as a "view from the side" on one's own thinking and justification:[120]

---

[118] As Charles Taylor points out, such self-interpretation involves self-*evaluation*, in particular what he calls *strong evaluation*, "where we evaluate, that is, consider good/bad, desirable/despicable, our desires themselves" (Taylor 1985, p. 65). Strong evaluation is what shows itself in our ability to condemn an act despite our motivation to do it. Taylor argues further that an intersubjective language is constitutive for our existence as self-interpreting and -evaluating animals (cf. op.cit. p. 68ff.).

[119] Cf. Tugendhat 1979, 8. Vorlesung. "Meaning" is here understood ambiguously. When we speak of the meaning of actions and people we are not just talking about the reference of a word, but more generally about the question: What is the purpose (of this action, of this work of art, of my life)? This is particularly clear when it comes to human existence: "Das Leben eines Menschen ist der Gesamtzusammenhang seines Handelns. Daher fragen wir auch beim Leben eines Menschen, insbesondere beim eigenen Leben – und hier mit einer besonderen Betroffenheit – nach seinem Sinn: ist etwas damit bezweckt, bzw. was ist es, was ich selbst damit will?" (Tugendhat 1979, p. 168)

[120] Cf. Rödl 1998, p. 153.

> 'Metaphysisch' nenne ich den Standpunkt, auf dem man steht, wenn man sein Meinen und Rechtfertigen *von der Seite* anschaut. Es ist der Standpunkt, auf dem man nicht im Rechtfertigen Zugang zur Wirklichkeit hat, sondern sein Rechtfertigen an der Wirklichkeit misst. (Op.cit., p. 153)

This "side perspective" is equivalent to Thomas Nagel's "view from nowhere",[121] within which any self-reference is a reference to "the objective I". The ultimate ideal is to overcome all "I"-talk. Knowledge expressed by the use of indexical expressions is regarded as defective. Husserl's "ideal speech" and Quine's "canonic notation" are examples of index-free languages brought forward as metaphysical ideals. Against this, Rödl argues that the idea of a perspective-free, subject-free knowledge is senseless:

> Der Bezug auf sich als an einem bestimmten Ort ist ein unabtrennbarer Teil des Begriffs eines räumlichen Gegenstands. (…) Die Vorstellung eines perspektivlosen Erkennens ist in sich sinnwidrig. Was als Bedingung ausgegeben wird, unter der das Erkennen *ideal* ist, ist eine Bedingung, unter [der] es *unmöglich* ist. (Op.cit. p. 164)

To ascribe perceptual knowledge to myself involves placing myself in a certain position in relation to the characterised object; thus perception and spatial positioning are inseparable. This means that when I refer to myself, I necessarily refer to a spatial being, a Heideggerian "Leib":

> Meine Identität als erfahrende Wesen und meine Identität als räumliches Wesen können nicht auseinanderfallen, weil der normative Zusammenhang meiner sinnlichen Erfahrung die Kontinuität meines Lebensweges im Raum einschließt. (Op.cit. p. 152)[122]

Like Brandom, Rödl contrasts his own *normative* theory of meaning with a metaphysical theory of meaning as *representation*. According to a common version of the latter theory, to think something is to be in a certain mental state, which according to certain laws or regularities of nature, results from other states.[123] The relation between the thought 'that *p*' and the fact that *p* is reduced to a statistic or causal relation, hence not recognised as a conceptual relation. In this way, Rödl argues, the relationship between truth and justification is dissolved in such a way that all truth talk ultimately becomes senseless.

---

[121] Cf. Rödl 1998, p. 160.
[122] Although Rödl here presents a sound critique against e.g. Nagel, it is imperative to also maintain Tugendhat's (and Brandom's) point, that we *do* have access to index-free, objective sentences – although these are necessarily rooted in a plurality of perspectives. Cf. below, Ch. 6.4.
[123] Wittgenstein's *Tractatus* is an example of an alternative representationalist theory, where the truth conditions of a representation are different from its causal conditions.

We ascribe opinions, reasons, and perceptions to ourselves *normatively*. When I ascribe a belief to myself in propositions like "I believe that p", I do not internally point to certain mental states within myself.[124] Similarly, when ascribing an action to myself, I do not rationally justify this by internally pointing to certain mental states, e.g. impressions or intentions, that have caused it. I ascribe an action to myself from a normative "I"- perspective. Action and belief are the two dimensions of normative self-reference. I *normatively* ascribe actions and beliefs to myself – as opposed to *descriptively* reporting them. Rödl calls this the "inner perspective":

> Man bezieht sich auf das eigene Meinen und Handeln nicht *deskriptiv*, nicht so, dass man es feststellt, konstatiert und darüber berichtet. Meinungen und Handlungen schreibt man sich *normativ* zu. Und dieser Bezug auf ein Sollen ist das, was eine Perspektive zur Innenperspektive macht, was bewussten Selbstbezug und Subjektivität konstituiert. (Op.cit. p. 49)

According to the causal theory of action, behaviour can be described as intentional action if it can be rationally justified by pointing to certain attitudes. As argued above, in Chapter 3, this approach is unable to account for the circumstance that actions are *carried out*.[125] Rödl puts heavy emphasis on this *performative* aspect, the "carrying out", singling this out as our particular form of life, the way we relate to our own existence:

> Ein Haus (…) ist so da, dass es steht, ein Tier so, dass es lebt, ein Mench aber so, dass er seine Existenz vollzieht. Wenn wir sagen, wie ein Mench seine Existenz vollzieht, sprechen wir nicht über Eigenschaften und Zustände, sondern darüber, wie er da ist. (Op.cit., p. 77)

Charles Taylor described human beings as self-interpreting animals. This self-interpretation is not something external to the human being, but its defining characteristic. To Rödl, this is the core of the normative perspective:

> Indem ich mich zu meinem Sein verhalte, betrachte ich nicht mein Sein, ich vollziehe es. Das Selbstverhältnis im menschlichen Sein ist kein theoretisches, sondern ein praktisches Selbstverhältnis. (Loc.cit.)

The error of determinism, according to Rödl, is that it fails to acknowledge that the meaning of the terms "can" and "possible" changes depending on whether one adopts an *assertoric*

[124] Cf. Wittgenstein, PU II, x.
[125] Cf. above, Ch. 3.6; cf. Keil 2000a, p. 143: "In der kausalistischen Handlungsdefinition fällt der Umstand unter den Tisch, daβ Handlungen vollzogen oder ausgeführt werden".

[*feststellenden*] or a *performative* [*vollziehenden*] perspective. "This or that can happen" means that I don't know what will happen (from an assertoric perspective). "I can do this or that" means that I exist in such a way that I take a stand on, and decide about, my own being (from a performative perspective). It is the latter, the *performative* aspect of human existence that we discuss under the heading "free agency":

> Das kann man auch so ausdrücken, dass dieses Sein *frei* ist. Dabei bezeichnet 'frei' keine Eigenschaft des menschlichen Seins, sondern seine Struktur als Vollziehen. (Op.cit., p. 79)[126]

According to Rödl's account, the question of whether one can scientifically prove the existence of human action is a grammatical confusion, on a level with questions such as whether machines can think.

Rödl makes extensive use of the inner-outer metaphor in his approach to the normative. Our attitudes towards our own actions and beliefs belong, as we have seen, to the "inner perspective". However, this seemingly dualistic approach is countered through his critique of what he calls the "Cartesian fallacy"; the presupposition that the I is not a part of the world.[127] Rödl diagnoses the Cartesian fallacy in its undisguised versions, such as in Thomas Nagel's, and in more hidden versions such as those presented by Tugendhat and Strawson. It reveals itself as the underlying premise of recognising the corporeality of the subject by way of recognising other bodies. Rödl repudiates this fallacy in a Kantian manner,[128] by arguing that an ability to refer to objects in space already implies placing oneself in a spatial relation to such objects:

> Die Idee des Raumes hat keinen Gehalt unabhängig von der Idee der eigenen Position im Raum. (Op.cit. p. 48)

Subjects – both oneself and others – can only be referred to as embodied, as "Leib". In this way Rödl stays clear of the crudest charges of dualism. Nevertheless a certain dualistic tendency seems to linger in the way he postulates watertight compartments between the performative and the assertoric perspective:

---

[126] Cf. Wellmer 1991, p. 181: "[D]as Sein des sprachlichen Sinns, der Freiheit, der Wahrheit, der Vernunft [ist] ein performatives Sein […], ein Sein, das sich erst in der performativen Einstellung sprachlich kommunizierender Subjekte konstituiert und nur in ihr sich erhält" – cf. below, Ch. 6.9.
[127] Cf. Rödl 1998, p. 35ff
[128] Cf. e.g. Kant: "Von dem ersten Grunde des Unterschiedes der Gegenden im Raume" (1768) and "Was heißt: Sich im Denken orientieren" (1786) (*Kants gesammelte Schriften*, Akademieausgabe Bd. II and Bd. VIII).

> In der konstatierenden Perspektive lässt sich Freiheit nicht thematisieren. (Op.cit. p. 79)[129]

Rödl argues that the subject of freedom belongs solely within the "performing perspective" of "carrying out", and cannot be looked at from an assertoric perspective. This suggests that freedom can only be recognised from the first-person perspective. But in that case Rödl, like Kant, faces a "methodological-solipsistic" problem regarding how the freedom of other subjects can be recognised.[130]

On the other hand, Rödl argues that my subjectivity is given to me only to the degree that other subjects are equally given to me.[131] This is Rödl's version of the Intersubjectivity Thesis, which will be presented in some of its versions in the next chapter.[132] Like Wittgenstein, Rödl bases this thesis on the demand for independent linguistic standards. Whether there are other subjects is not a puzzle for me to solve from the assertoric perspective, but something I recognise simply by being a subject myself:

> Und deshalb sollte man nicht sagen, ich erkenne, dass jemand ein Subjekt ist. Angemessener ist es zu sagen, dass ich ihn als Subjekt *anerkenne*. (Op.cit. p. 268)

This should clear Rödl from the charge of methodological solipsism. However, it seems to me that he to a certain degree underplays the possibility of *transitions* between what is part of, and what is the limit of experience, in the sense that free agency is not only recognized in direct I-you communication, but also ascribed to other subjects who we speak *about*. This is an insight that is not captured by the inner-outer metaphor, but comes into vision through a closer examination of the grammatical structure of ordinary language. Wittgenstein's paragraphs on the special "infallible" expression in the first person present tense, Searle and Habermas's theory of the performative-propositional double structure of speech, and Tugendhat's display of a system of interchangeable expressions are among the ways of bringing the possibilities of transitions between a performative first-person perspective and other perspectives into sight. This topic will be treated in Chapter 6.

Before moving in that direction, however, I will take a closer look at some underlying elements of a normative approach to action. In Chapter 5, I take Searle's concept of a "circle

[129] Cf. below Ch. 7.3; cf. Habermas 2006, p. 671: "Für den Beobachter ist die Frage, ob die Person auch anders hätte handeln können, kein Thema".
[130] Cf. above, Ch. 2.5.
[131] Cf. Rödl 1998, p. 268f.
[132] Cf. below, Ch. 5.3.

of intentionality" as my point of departure from which to look at the relationship between normativity and intersubjectivity.

Chapter 5: The intersubjective basis of normativity

> Die logische Geltung von Argumenten kann nicht überprüft werden,
> ohne im Prinzip eine Gemeinschaft von Denkern vorauszusetzen, die zur
> intersubjektiven Verständigung und Konsensusbildung befähigt sind.
> (Apel 1973, p. 399)

5.1 Introduction: Understanding normativity

In Chapter 4, I argued that a distinctive feature of being a language user and a rational agent is to be subject to normative standards. I will follow up this argument here, by arguing that the use of language presupposes a community of language users, in other words that all language is intersubjective. The "logical space of reasons" (Sellars) is a space that opens up between a plurality of rational subjects and the world. Karl-Otto Apel's phrase "Verständigung über etwas" expresses the complementarity between understanding and agreement: An observational subject-object relation is only attainable for subjects within a community of language users, who are able to reach an agreement between them on questions of true and false, right and wrong.

A further question I pursue in this chapter is whether it is possible to view the "normative space of reasons" as a purely *logical* space, or whether *morality* must somehow be viewed as fundamental to understanding. According to Kant, a "people of devils" – a community where no one is bound by ethical norms or morality – may be capable of acting rationally, in the sense of establishing a good political order, "if only they have intelligence".[133] However, if epistemic normativity must be explicated on the basis of the moral law - as Kant in other passages comes close to suggesting - this scenario seems to

[133] Cf. *Kants Gesammelte Schriften* (Akademienausgabe) Bd. VIII; p. 366 ("Zum ewigen Frieden").

become an impossible one, or at least would have to be viewed as a borderline case of rationality.

I begin by looking at what Searle calls "the circle of intentionality", thereby trying to encircle the way concepts such as mind, normativity, and intersubjectivity stand in relation to each other (5.2). In this section, I also comment on how a normative approach to language and agency confronts the ideals of analytic naturalism. In 5.3, the thesis of the intersubjectivity of language is examined in its different versions, prompting a closer look at the role of the second person (5.4). The chapter ends with a discussion of the sources of normativity, and specifically the relationship between epistemic normativity and morality (5.5).

5.2 Will the circle be unbroken? – Searle on intentionality and normativity

Intentionality, in its minimal sense of directedness, may be seen as a feature common to all living creatures as well as to a great deal of complex machinery. Examining such intentionality in its various forms and expressions constitutes substantial and interesting fields of study; think of the similarities and differences between intentional states in different life-forms, or the similarities and differences between "live" intentionality and the form we find in robots. However, in this particular inquiry, the focus is on intentionality in the sense of discursive practice, the propositional kind that has been labelled by Robert Brandom "the fanciest sort of intentionality".[134] The aim of such an inquiry is, as he reminds us, not to deny or diminish the relevance or value of lower grades of intentionality. Rather:

> The aim is to understand ourselves as judgers and agents. (Brandom 1994, p. 7)

Higher-grade intentionality is the common denominator for several of the key concepts of this thesis: Freedom, rational agency, mind, language, and intersubjectivity. These are phenomena, capacities, or qualities that are co-instantiated in human beings. The concepts can only to a limited degree be explicated independently of each other. We could say that they must be explicated *holistically*. This holism can be expressed with reference to what John Searle has called "the circle of Intentionality":

> Any attempt to characterize Intentionality must inevitably use Intentional notions. (Searle 1979, p. 90)

[134] Cf. Brandom 1994, p. 7.

Searle has further stated that intentionality has a normative structure.[135] This is in line with John McDowell's thesis of the normative relation between mind and world:

> To make sense of the idea of a mental state's or episode's being directed towards the world, in the way which, say, a belief or judgement is, we need to put the state or episode in a normative context. (McDowell 1994 p. xi)

Searle seems to diagnose intentionality and normativity as constituents of a hermeneutic circle – one that we should not attempt to break, but rather to enter in the right way in order to allow the concepts to clarify each other.[136] However, he himself has apparently made several attempts to break the circularity that he diagnosed as inevitable. In *Intentionality* (1983), he makes the following programmatic comment, connecting intentionality to an altogether different kind of circularity:

> [C]onsciousness and Intentionality are as much a part of human biology as digestion or the circulation of the blood. (Searle 1983, p. ix)

In *Rationality in Action* (2001), Searle elaborates on how intentionality and normativity should be viewed as entirely biological phenomena, since animals are as bound by it as human agents:

> [I]f you think about matters from the point of view of sweaty biological beasts like ourselves, normativity is pretty much everywhere (…) If an animal has a belief, the belief is subject to the norms of truth, rationality, and consistency. If an animal has intentions, those intentions can succeed or fail (…) From the point of view of the animal, there is no escape from normativity. The bare representation of an *is* gives the animal an *ought*. (Searle 2001, p. 182f.)

Searle has argued (against among others Daniel Dennett[137]) that all conscious creatures are capable of what he calls *intrinsic* intentionality. Intrinsic intentionality is the distinguishing mark of conscious states, and is contrasted with the *derived* intentionality of other things, such as sentences. Only conscious minds are capable of intrinsic intentionality, whereas e.g. complex computers or robots only possess intentionality derived from the intrinsic

[135] Cf. Searle 2001, p. 182.
[136] Cf. above, Ch. 3.5.
[137] Cf. Searle 1990.

intentionality of conscious minds. Conscious states are, according to Searle, distinguished by their "first-person ontology".[138] To Searle, the thesis of first-person ontology serves as a safeguard against objective, scientific reductions of consciousness, so as to steer clear of the *Scylla* of materialism. At the same time, however, he maintains that consciousness must be regarded as a completely natural, biological phenomenon, and thus attempts to avoid the *Charybdis* of dualism. However, it is not clear to me how Searle intends to solve the mind-body problem simply by insisting that the mind is both irreducibly subjective and a biological phenomenon. He must somehow show how these two facts come together.

In *Selbstbezug und Normativität*, Sebastian Rödl criticises Searle's concept of intrinsic intentionality, and identifies it as an instance of what he calls "Cartesian ontology". Searle's argument for the irreducibility of a first-person perspective is based on the idea that being in a certain mental state is something to which I have a direct access. My mental states are *transparent* to me – as opposed to the mental states of others. Rödl argues that this way of thinking ultimately leaves Searle defenceless against solipsism as well as scepticism. Since the intrinsically intentional states of other people can be experienced only indirectly by me, I must base my assumptions on an inference by analogy; from my own "inner life" to the existence of other people's "inner life". However, no facts or experiences can count as evidence or even as support for such an assumption, hence the ascription of mental states to other subjects is meaningless. The idea of intrinsic intentional states also suggests that my access to reality through e.g. visual experience can somehow be *justified*. However:

> Die Vorstellung, ich müsste begründen oder sonstwie ausweisen, dass ich im visuellen Erleben Zugang zur Wirklichkeit habe, dass ich also wirklich sehe, ist leer. Die Selbstzuschreibung visueller Eindrücke hat einen irreduzibel normativen Sinn. Visuelle Eindrücke vermitteln als solche Erkenntnis. Das ist aber unvereinbar mit der Idee, der Gehalt solcher Eindrücke sei durch innere Natur bestimmt, sie seien intrinsisch intentionale Zustände. (Rödl 1998, p. 245)

In his 1994 article "Searle, Leibniz and 'The First Person'", Audun Øfsti points out that Searle, despite his brilliant observations on the first-person perspective in some contexts, paradoxically seems to fall back on an objectivist outlook in others. Øfsti's point of departure is a passage in Searle's *Intentionality*, where he presents Leibniz's argument in *Monadologie* (§ 17), saying that if we could enter into a "thinking, feeling and perceiving" machine as we enter into a mill, we still would not find anything by which to explain perception. Searle

[138] Cf. Searle 1998, p. 52.

accuses Leibniz of "proving too much"; the same could be said about water, he claims, since "the behaviour of H2O molecules can never explain the liquidity of water".[139] Øfsti argues that Searle overlooks Leibniz's sound point. This does not, as Searle seems to suggest, concern the incompleteness of our scientific methods. Rather, it concerns the "non-circumventability" of the first-person perspective in the sense that this can never be accounted for from an "external", third-person perspective:

> Searle's attempt to illustrate the relationship between physical and mental, 'outside' and 'inside' (third person and first person) remains within the frame of a third person perspective, as a matter of speaking of 'it' on different levels, without having recourse to the first person. (Øfsti 1994, p. 687) [140]

Searle appears to show a similar neglect of the first-person perspective in a recent debate with Habermas. While Habermas argues that the "language game of responsible agency" requires a participant's perspective and thus cannot be completely replaced by an objectivistic "language game of neuroscience"[141], Searle counters that the two language games "give different levels of descriptions of the same system".[142] To this it can be replied that a characteristic mark of the participating perspective is that it is *not* a description of a state of affairs.[143]

Although Searle emphasises the irreducibility of the first person, he seems to – at least partly – misconstrue this in an objectivistic sense. To me it seems that Searle fails to appreciate the "infrastructure" necessary in order to establish a first-person perspective. Being intentionally directed at something is not sufficient to constitute a genuine first-person perspective. In Chapter 6, I will argue that a first-person perspective is possible only within a

---

[139] Cf. Searle 1983, p. 268.
[140] Øfsti also compliments Searle's anti-dualist efforts. According to Øfsti, however, a certain dualism is appropriate, just not the troublesome dualism *within* the world that Searle attacks: "The fundamental dualism (which is perhaps not so troublesome after all, when seen in the proper light) is the dualism between the world (as a whole) and (the) 'we' living in it, that is once more the dualism between 'the first person' and 'the third person' (the world)" (op.cit, p. 686). In the adjoining footnotes, Øfsti elaborates on how the meaning of psychological verb phrases necessarily has two roots, and that both a first-person perspective and a third-person perspective are necessary elements of a complete language. I will come back to this decisive argument in Chapter 6. The term dualism seems, however, to point in the wrong direction here, since what Øfsti wants to display is a necessary *unity* of language in the form of a system involving declination between first and third person, i.e. *transitions* between different perspectives, cf. Øfsti 2002, p. 126, fn. 46: "Der Preis der Dualismus darf nicht bezahlt werden. Oder besser: Der Dualismus zerstört ja auch auf seiner Weise den entscheidenden Punkt (durch 'Entspannung')". On the problem of avoiding dualism, cf. below, Ch. 6 and 7.
[141] Cf. Habermas 2007a; cf. below, Ch. 7.1.
[142] Cf. Searle 2007, p. 69.
[143] Cf. also Habermas' reply to Searle in Habermas 2007, p. 89ff.

full-fledged language where propositional content can be maintained through different speakers' perspectives.[144]

Robert Brandom differentiates between *sapience* and *sentience*:

> Sentience is what we share with non-verbal animals such as cats – the capacity to be *aware* in the sense of being *awake*. Sentience, which so far as our understanding yet reaches is an exclusively biological phenomenon, is in turn to be distinguished from the mere reliable differential responsiveness we sentients share with artifacts such as thermostats and land mines. Sapience concerns understanding or intelligence, rather than irritability or arousal. One is treating something as sapient insofar as one explains its behavior by attributing to it intentional states such as belief and desire as constituting reasons for that behavior. (Brandom 1994, p. 5)

John Haugeland, Brandom's Pittsburgh colleague, draws the line in question between intrinsic and *ersatz* intentionality. Haugeland argues that

> animal intentionality is ersatz because (or to the extent that) animals do not (…) submit themselves to norms. (Haugeland 1998, p. 303)

A dog cannot, in other words, evaluate its own intentional states according to normative standards such as truth or validity. Being intentionally directed at something – e.g. believing something – is not tantamount to being bound by norms. Although it makes sense to say that a dog *believes* that a bone is buried in a certain place, and although the dog's belief might be said to be true or false, it is not the dog, but us – the concept users – who apply the normative standard of truth.

The unbreakable circle of intentionality may be seen as the antithesis to a programme of analytic naturalism, which in turn may be defined broadly as the thesis that normative concepts can be explained in terms of natural properties.[145] Jerry Fodor, a well-known critique of holism, thus expresses the interest naturalism would have in "breaking the circle":

---

[144] Cf. Tugendhat 1976. Cf. also Carson 2002, p. 60f. on the role of evaluation and normative statements in a complete language.

[145] Such attempts are often categorised as examples of *naturalistic fallacy*. The concept stems from G.E. Moore's *Principia Ethica* (1903). Moore argues that any attempt to analyse the concept "good" constitutes the committing of a naturalistic fallacy. This has later been constructed as the confusion of describing and evaluating (see e.g. Hare), and broader with the "is-ought problem" commonly traced back to a section in David Hume's *Treatise*, and reappearing in numerous versions in post-Humean philosophy. I avoid using the concept "naturalistic fallacy" here, since the approach of analytic naturalism, however problematic, does not in itself seem to constitute a *fallacy* (cf. Keil 2004, p. 16).

> Given any (…) suitable break of the intentional circle, it would be reasonable to claim that the main *philosophical* problem about intentionality had been solved. (Fodor 1990, p. 52)

If sufficient conditions for an intentional phenomenon could be stated using purely non-intentional terms this would count as a successful breach of the circle, whereupon the naturalist could count upon a veritable domino effect.[146]

Against Searle's attempt, as well as other efforts to "break the circle", this, I suggest that action is *irreducibly* normative. This implies that normative phenomena cannot be *naturalised* in the analytical sense, i.e. cannot be redefined in non-normative terms. We could easily get the impression that a fruitful discussion between analytic naturalism and the position defended in this chapter ends here. It is not easy to see how we can move on from this point. A *naturalistic*, and admittedly even a *natural* response to the claim that the intentional circle cannot be broken, is to conclude that it puts a stop to not only this discussion, but to explanation in general. From a naturalist perspective, it would seem to constitute a case of "apriorism", i.e. of answering prematurely what should be left to further scientific research. If the intentional circle cannot be broken, then how do we analyse normativity further? A great part of the philosophical debate on this topic seems to consist of the attacks on and defences of different versions of naturalism. Nevertheless, real contributions to knowledge may sometimes arise from such apparently dead-locked controversies. Geert Keil comments on how a proper critique of naturalism can improve our insight into the phenomena in question:

> Der Erkenntnisgewinn einer solchen Kritik des Naturalismus (…) besteht in einer verbesserten Einsicht in die Größe des intentionalen Zirkels, in die Menge der Phänomene, die sich in ihm verfangen. (Keil 2000b, p. 202)

Thus, a positive outcome of an anti-naturalistic insistence on the ineliminable dimension of normativity is hopefully an "improved insight" into the size and substance of the intentional circle.

[146] Cf. Keil 2000b, p. 196.

5.3 The Intersubjectivity Thesis

Intersubjective understanding may be seen as the common denominator of the phenomena constituting the circle of intentionality, stitching together concepts such as language and mind, rationality and normativity. My approach to free agency is based on what we may call the *Intersubjectivity Thesis*,[147] which we may tentatively formulate thus:

IT 1: All language is intersubjective.

The Intersubjectivity Thesis can be viewed as an attempt to express the essence of Wittgenstein's so-called Private Language Argument (PLA). Stated plainly, a private language is a language that "another person cannot understand".[148] Throughout the *Philosophical Investigation*, Wittgenstein argues that such a language is unthinkable, since it would be impossible for the speaker to maintain a "private" meaning of the linguistic expressions.

The content and extent of the PLA has been the source of philosophical controversy for half a century. The dividing line is found between proponents of the weaker thesis: "A language that *in principle* cannot be understood by more than one speaker is impossible", and proponents of the stronger thesis: "A language that *de facto* is not understood by more than one speaker is impossible". Wittgenstein's own formulations on private languages are notoriously ambiguous. In numerous passages he seems to be launching a strong Intersubjectivity Thesis;[149] however, he never formulates such a thesis explicitly. On the contrary, some passages rather seem to point towards a relativisation of the thesis.[150]

Many Wittgenstein commentators draw the conclusion that all that the PLA demands from a language is that the words refer to something in the world so that other speakers *could* understand them, whether or not such speakers actually exist. This view corresponds to a "weak" Intersubjectivity Thesis:[151]

---

[147] Cf. Wellmer 2004, e.g. p. 114.
[148] Cf. Wittgenstein, PU § 243.
[149] Cf. e.g. PU §§ 77 and 560, on the *learnability* of language.
[150] Cf. e.g. PU § 199: "It is not possible that there should have been only one occasion on which only one person obeyed a rule" is ambiguous, but seems to suggest that an isolated person – a "Mowgli" – possibly *can* follow a rule – as long as he does so more than once.
[151] Cf. Wellmer 2002, p. 109.

IT 2: All language is *potentially* intersubjective, since all language use *in principle* must be interpretable by other language users.

Defenders of IT 2 often emphasise certain passages in *Philosophical Investigation* that pinpoint the importance of external criteria for meaning, such as § 293 on the beetle in the box.

An alternative group of Wittgenstein interpreters, including Kripke and Davidson, emphasise paragraphs on rules and the normative aspect of meaning and understanding, such as § 202:

> And hence also 'obeying a rule' is a practice. And to think one is obeying a rule is not to obey a rule. Hence it is not possible to obey a rule 'privately': otherwise thinking one was obeying a rule would be the same thing as obeying it.

Saul Kripke analyses sentences that specify the meaning of linguistic expressions, such as "By 'and' X means: plus."[152] Basically, he emphasises the normative aspect of linguistic meaning: To understand a linguistic expression means to know how to apply it correctly. Kripke promotes the interpretation of the PLA that a private language is impossible since in order to mean something you must also be able to understand, i.e. to interpret other speakers. The reason why a language requires a plurality of speakers is that the speaker of a language necessarily is an interpreter of other speakers.

Arguably, Kripke's version of the PLA goes well beyond what can be read out of Wittgenstein's own formulations, giving rise to widespread criticism as well as to the nickname "Kripkenstein". Albrecht Wellmer agrees that there are certain problematic aspects of Kripke's Wittgenstein reading. Nevertheless he argues that Kripke's emphasis on the normative aspect of sentences that specify meaning constitutes a decisive reinforcement of the PLA, in sound opposition to Wittgenstein's own hesitation regarding the crucial issue of intersubjectivity.[153]

Donald Davidson is also a defender of a strong reading of the PLA.[154] Basically, like Kripke, Davidson investigates the reciprocal structure of language; however, his argument takes a different approach. Following on Tarski's theory of meaning and Quine's theory of radical translation, Davidson arrives at a general theory of linguistic understanding as

[152] Cf. Kripke 1982, p. 7ff.
[153] Cf. Wellmer 2004, p. 110f.
[154] Cf. e.g. Davidson 2000/2001a.

interpretation. Whereas Wittgenstein emphasises that language is a common practice, Davidson focuses on the plurality of perspectives within language. In order to understand another language user you must always *interpret* her expressions, and the perspective of the interpreter is always different from that of the speaker. Understanding therefore requires a "principle of charity", i.e. an (implicit) assumption of a maximum of truth of the speaker's beliefs.

The principle of charity is, we might say, the analytic philosophy's equivalent to a hermeneutic principle à la Gadamer's, stating that the understanding of another speaker requires a normative "anticipation of completeness". Davidson, however, moves beyond a basic hermeneutic principle and argues that each of us, in order to understand other speakers, must "radically translate" their expressions. It is not only the interpretation of utterances in foreign languages, but all understanding that ultimately follows the pattern of radical translation. According to this account, Wittgenstein's idea of a common language seems to dissolve into a myriad of private languages sharing a common set of interpretation rules.[155]

Against this, Wellmer makes it clear that the paradigm of communication is mutual understanding *within* a language, and not the radical translation of two interpreters with separate languages. He modifies Davidson's principle of charity by dismissing the assumption of (mostly) true beliefs, and instead embraces an assumption of the rationality and linguistic competence of one's co-speakers. He also generalises Davidson's approach, in the sense that he moves from a theory of the meaning of propositions to a theory of the pragmatic meaning of speech acts. On a general level, however, Wellmer endorses the hermeneutic premise that serves as the foundation of Davidson's theory as a safeguard against sceptical arguments.

Davidson argues that distinctions between "true" and "false", "right" and "wrong" only make sense within a space of mutual interpretation. Thus, like Kripke, he arrives at a strong version of the Intersubjectivity Thesis:

> IT 3: Language is *essentially* intersubjective, since communication with other language users is a condition for the use of language.[156]

---

[155] According to Brandom, the Davidsonian approach ultimately displays a Kantian dualism, cf. Brandom 1994, p. 615: "Essential elements of Kant's dualistic conception of concepts are still with us today. They are the basis for the suspicion evinced by some (for instance Davidson) that talk of concepts inevitably commits us to a picture in which they play the role of *epistemological intermediaries*, which stand between us and the world we conceptualize and forever bring in to question the very possibility of genuine cognitive access to what lies beyond them."

[156] Cf. Davidson 2000, p. 407: "[E]s (kann) keine private Sprache geben (…), d.h. keine Sprache, die nur ein einzelnes Wesen *versteht*. (…) Was (…) erforderlich ist, ist Interaktion zwischen mindestens zwei Sprecher-Interpreter" (my italics). In Davidson's version, the PSA explicitly states the necessity of a plurality of speakers,

This strong version of the Intersubjectivity Thesis underlies my approach to action and normativity in the last chapter and in this one. According to this approach, reasons are conceptually public. Reasons are public both in the sense that 1) the question of what reasons the agent should commit to must be determined relative to our experience as agents in the world, and 2) in the sense that a reason to believe something represents a commitment to rational standards that are in principle sharable.

Wellmer defends IT 3, and a strong reading of the PLA, stating that the normativity of language is conditioned by a social practice:

> [D]ie Normativität des Bedeutens (muss) als die Normativität einer sozialen Sprachpraxis mitgedacht werden. Durch diese Normativität konstituieren sich Bedeutungen allererst. Wittgenstein verwendet zur Kennzeichnung dieser Normativität der sozialen Sprachpraxis auch den Begriff der Regel. (Wellmer 2004, p. 57)

He argues that there are certain "holistic-normative conditions for the ascription of beliefs".[157] In the ascribing of beliefs in propositions like "X believes that p", we recognise a peculiar reciprocity: *In a strict sense* I can only ascribe beliefs to a being that can equally ascribe beliefs to me.[158]

Wellmer relates his defence of IT 3 to Brandom's argument that an intersubjective content of beliefs requires a plurality of perspectives. Brandom examines the transitions between *de dicto* and *de re* ascriptions of beliefs in language, i.e. between attributing the belief *that* a saying (*dictum*) is true and attributing the belief *of* or *about* some thing (res). He also examines the role of *anaphora*, expressions referring to other expressions, such as deictic expressions. He argues that the de dicto-de re transitions express the *difference* between various scorekeeping perspectives, whereas anaphoras secure the *connection* between them. These linguistic mechanisms make it possible to refer to *the same* belief or proposition from different perspectives. This is constitutive for intersubjective, situation-independent propositional content, as well as for successful communication. In this way, the representational dimension of language and the social or communicative dimension reveal

---

as opposed to Wittgenstein's more carefully formulated thesis on the impossibility of a language that another person "nicht verstehen *kann*" (cf. PU § 243).

[157] Wellmer 2004, p. 142f.

[158] "In a strict sense" we cannot, in other words, ascribe beliefs or intentions to dogs, cf. above, Ch. 5.2. The dog's intentionality is derived in the sense that it cannot itself apply the normative standards, e.g. of truth.

themselves as complementary, and not as two self-sufficient and opposed perspectives. Rather they are seen as two sides of the same coin:

> The context within which concern with what is thought and talked *about* arises is the assessment of how the judgment of one individual can serve as reasons for another. The representational content of claims and the beliefs they express reflect the social dimension of the game of giving and asking for reasons. (Brandom 2000a, p. 159, cf. above, Ch. 4.3)

Brandom's inferentialism illustrates that a plurality of speakers' situations is what makes situation-independent content possible, explicating the intersubjective conditions for "truth talk":[159]

> The representational dimension of propositional contents is explicated in terms of the social-perspectival character of discursive scorekeeping and the substitutional substructure of its inferential articulation. In this way it is possible to understand what is involved in assessments of judgments as objectively true or false – as correct or incorrect in a sense that answers to the properties and relations of the objects they are about, rather than to the attitudes of any or all of the members of the community of concepts users. (Brandom 1994, p. 649)

To Wellmer, Brandom's approach constitutes a new way of formulating the strong Intersubjectivity Thesis:

> Brandoms These ist (…) eine neue Version der starken Intersubjektivitätsthese (…). Sie besagt, dass die Idee eines von einzelnen Sprechakten unabhängigen propositionalen Gehalts von Äußerungen – dasjenige also, was einen Streit über die Wahrheit von Äußerungen erst möglich macht – eine Pluralität von Perspektiven voraussetzt, aus denen solche propositionalen Gehalte sich artikulieren lassen. (Wellmer 2004, p. 131)

One aspect of the Intersubjectivity Thesis is that objective truth presupposes a plurality of perspectives from which a situation-independent propositional content may be articulated.[160] Normative status – truth, rightness, goodness – is based on the intersubjective relations between speakers of a common language:

---

[159] Cf. Brandom 1994, p. 17: "The business of truth talk is to evaluate the extent to which a state or act has fulfilled a certain kind of responsibility".

[160] In this sense, Tugendhat 1976 may also be viewed as presenting a version of the Intersubjectivity Thesis, cf. below, Ch. 6.4. There are strong parallels between Brandom's inferentialism and Tugendhat's display of a system of interchangable expressions.

> Die Fähigkeit zur Unterscheidung zwischen richtigen und falschen Wortverwendungen [ist] eine Bedingung der Möglichkeit der Unterscheidung zwischen wahren und falschen Äußerungen (…) Das heißt, dass die Gemeinsamkeit der Sprache und die mit ihr verbundenen Unterscheidungen zwischen 'richtig' und 'falsch' nur die Voraussetzung bilden für eine normative Unterscheidung anderer Art: die Unterscheidung zwischen berechtigten und unberechtigten, zwischen wahren und falschen Äußerungen. (Op.cit, p. 123f.)

An intersubjective language, with its necessary distinctions between the "correct" and "incorrect" use of words, is a condition for normative judgement as such, including the distinction between true and false. A thesis of the intersubjectivity of language is, in other words, a necessary element in the explanation of our ability to make normative distinctions. The Intersubjectivity Thesis evolves from a linguistic-pragmatic approach to truth and to normative statuses in general. The idea that normative statuses are instituted by normative attitudes is a thought that Brandom traces back to the Enlightenment:

> Enlightenment conceptions of the normative are distinguished by the essential role they take to be played by normative *attitudes* in instituting normative *statuses*. (Implicit social contract theories of political obligation are a case in point.) It does not make sense to talk about commitments and entitlements, responsibility and authority, apart from our practices of taking or treating one another *as* committed or entitled, responsible or authoritative. This thought should be understood as another holist, reciprocal sense dependence thesis. (Brandom 2002, p. 54)

The "holist, reciprocal sense dependence thesis" may count as a version of the circle of intentionality, and provides the basis for a view of language as intersubjective practice. Language must thus be conceived of in terms of language use, in other words as a subcategory of, as well as a condition for, rational agency.

5.4 The second person

The Intersubjectivity Thesis states that language necessarily is a social construct. To Wellmer and Brandom, the intersubjectivity of language makes true statements possible. A plurality of perspectives is required in order to render possible a propositional content that is independent of the prevailing speech situation.[161] Brandom criticises the traditional view of

[161] Cf. Tugendhat 1976; cf. below, Ch. 6.4.

intersubjectivity, which he calls "I-we sociality".[162] This view focuses – misleadingly – on the contrast between a limited individual perspective and the privileged perspective of the community, the "we". The approach is defective, Brandom claims, because it excludes the possibility that the privileged perspective is wrong. It implies that

> what the community *takes* to be correct *is* correct. (Brandom 1994, p. 599)

Brandom's alternative is the "I-thou sociality", where no single perspective is privileged when it comes to truth:

> [A]ccording to the I-thou construal of intersubjectivity, each perspective is at most *locally* privileged, in that it incorporates a structural distinction between objectively correct applications of concepts and applications that are merely subjectively taken to be correct. (Op.cit., p. 600)

Brandom attempts to re-construe the concept of objectivity: from referring to certain non-perspectival or cross-perspectival propositional *contents* to concerning a particular perspectival *form*:

> What is shared by all discursive perspectives is that there is a difference between what is objectively correct in the way of concept application and what is merely taken to be so, not what it is – the structure, not the content. (Loc.cit.)

Brandom's approach to intersubjectivity and the possibility of objective norms has attracted the attention of Jürgen Habermas. In *Truth and Justification* (2003, German edition 1999), Habermas refers to *Making it Explicit* as

> a landmark in theoretical philosophy comparable to that constituted in the early seventies by *A Theory of Justice* in practical philosophy. (Habermas 2003, p. 131)

Habermas emphasises that Brandom's programme, despite its "new pragmatic vocabulary" and "rare combination of speculative impulse and staying power" (Loc.cit.), already has been sketched by others, referring specifically to his own universal pragmatics as well as the transcendental pragmatics of his colleague Karl-Otto Apel.[163] Apart from the divergent vocabulary, the parallels to "Apelmasian" pragmatics are striking, although Brandom's work

[162] Cf. Brandom 1994, p. 599.
[163] Cf. Habermas 2003, p. 131 and fn.1.

contains no references to this tradition.[164] They share predecessors like Kant and Hegel (e.g. on rationality), Wittgenstein (e.g. on meaning as use) and Frege (e.g. on propositions as the basic unit of language and thought). All three relate explicitly to American pragmatism as well: Apel is primarily preoccupied with the founder of American pragmatism, C.S.Peirce; Habermas refers frequently to G.H. Mead; and Brandom focuses on Wilfred Sellars' combination of elements from American pragmatism with analytic philosophy. In my view, Habermas, Apel and Brandom all belong within the framework of what Brandom calls *rational pragmatism*.

Despite his approval of the major direction of Brandom's work, Habermas has weighty critical remarks. The main point of Habermas' critique concerns Brandom's oscillation between idealist and realist ontological commitments. Another problematic aspect, according to Habermas, is that despite his good intentions, Brandom fails to establish a symmetric I-thou relation as the fundament of an intersubjective language. The reason is that he fails to do justice to the grammatical role of a genuine second person perspective:

> [Brandom] analyzes the attribution of validity claims, and their evaluation, without taking into consideration the complex interconnections of the first-, second-, and third-person perspectives. He actually construes what he calls the 'I-thou relation' as the relation between a first person who raises validity claims and a third person who attributes validity claims to the first. (Habermas 2003, p. 162)

Habermas does not think that Brandom's "I-thou relation" satisfies the conditions of a truly intersubjective communication. Brandom's "thou" is not an "addressee who is expected to *reply to the speaker*", but rather someone who merely "assesses the utterance of the speaker" (ibid.). In the terminology of Hans Skjervheim,[165] Habermas accuses Brandom of adopting an objectivist view of communication by representing the "thou" as a *spectator to*, and not as a *participant in* the discourse. To Habermas, Brandom's choice of examples is revealing, notably the one where a prosecutor and a defence attorney partly debate each other, and partly address a court room audience:

> Interestingly, Brandom singles out the indirect communication of the speakers with the spectators who are listening to them – rather than the communication of those directly involved – as the paradigmatic case. (Op.cit., p. 163)

[164] In Brandom 2000, p. 3, Brandom mentions how his focus on distinguishing discursive creatures from non-discursive creatures *separates* his project from a.o. earlier American pragmatism. This is one of the features, it seems to me, that *assimilate* his project to the "Apelmasian" versions of normative pragmatism.
[165] Cf. Skjervheim 1959 and 1996, cf. also Habermas 1981, p.163f.

Habermas argues that the audience in Brandom's example are *listeners*, assuming the role of third persons waiting to see what happens, and not *hearers*, directly involved in the discourse. As listeners, as opposed to hearers, they do not adopt the *performative* attitude of a first person toward a second person.

Habermas assimilates Brandom's approach with Davidson's theory of radical interpretation, which is contrasted with hermeneutic interpretation. This is in contrast to Wellmer, who as we have seen praises Davidson's (as well as Brandom's) theory as a fundamentally hermeneutic approach, contributing with a necessary radicalisation of Wittgenstein's Private Language Argument, cf. above (5.3). However, as mentioned, Wellmer argues that Davidson's theory must be adjusted, so that it becomes clear that the paradigm of communication is mutual understanding *within* a language, and not the radical translation of two interpreters with separate languages.[166]

Like Brandom, Davidson defends a strong version of the Intersubjectivity Thesis and the structural importance of a second person.[167] However, it seems to me that both Brandom and Davidson portray the second person primarily as an interpreter, i.e. as an assessor of the utterances of the first person, rather than as a *participating* and *responding* discourse partner. In contrast to this tendency, Habermas stresses the "coordinating" aspect of communication:

> Communication is not some self-sufficient game in which the interlocutors reciprocally inform each other about their beliefs and intentions. It is only the imperative of social integration – the need to coordinate the action-plans of independently deciding participants in action – that explains the point of linguistic communication. (Habermas 2003, p. 164)

Habermas rightly emphasises mutual understanding and social integration as the key elements of language use. However, I find it misleading to speak of this as the "point of communication" or to speak of rationally motivated agreement as the "goal of communication" (op.cit. p. 165). Thus, I think Brandom is right when he counters in a reply to Habermas:

> I deny this – not because I think that linguistic communication has some other point, but because I think it is a mistake to think of it as having a point at all. (Brandom 2000b, p. 363)

[166] Cf. Wellmer 2002, p. 206ff.
[167] Cf. Davidson 2001a, p. 107-121: "The second person".

In particular situations, we may use language in order to achieve some goal, but it does not make sense to speak of language use in general as *instrumental*, i.e. as having a goal outside of itself; rather it should be conceived of as *expressive*, i.e. as constitutive of our way of life or as a central element of our lifeworld.

As I see it, the crucial point in Habermas's critique of Brandom (which should be retained in spite of Brandom's meta-critique), as well as in his critique of Davidson, is that a genuine, participating second person is necessary in order to establish normativity. Commitments paradigmatically arise between subjects striving towards coordination of action-plans or mutual agreement – not between subjects reciprocally interpreting and assessing each other's speech acts.

5.5 Sources of normativity

Following the arguments of Kant, Wittgenstein, Brandom, and Rödl, among others, I have placed normativity at the very core of subjectivity and rational agency. But what is at the core of normativity? To Kant, the capacity to make normative judgements, and thus to be bound by rational reasons, arises from autonomy. The rational subject's ability to submit to the norms she lays down for herself is what constitutes her capability of normative judgement. As Brandom puts it:

> To be a self, a knower and an agent, is, according to Kant's original normative insight, to be able to take responsibility for what one does, to be able to undertake commitments. It is to be bound by norms. According to the autonomy thesis, one is in a strict sense bound only by rules or laws one has laid down for oneself, norms one has oneself endorsed. What *makes* them binding is that one *takes* them as binding. (Brandom 2002, p. 21)

Locating the source of normativity within the subject must not, however, collapse into a subjectivistic "whatever seems right to me *is* right". As Wittgenstein points out, only an intersubjective practice can secure binding rules, otherwise

> thinking one was obeying the rule would be the same thing as obeying it (PU § 202),

in which case normativity would break down.

The linguistic-pragmatic transformation of philosophy gives a definite form to the intersubjective basis of normativity, a form that arguably is missing in Kant's philosophy of mind. Intersubjective language, and the discursive commitments constitutive of it, is a condition for normative statuses in general – including the true-false distinction.

Both Habermas and Brandom seek to formulate the principles of a formal pragmatics by making explicit what is implicit in our everyday practices, thus to localise the genesis of normativity in the evolution of these practices. Brandom's conception of normativity is thoroughly epistemic, prompting certain critical remarks from Habermas.[168] Brandom, however, counters that there is a reason for his omission of a specific discussion of moral commitments:

> [T]he understanding of conceptual normativity has been hampered by the fact that theorists of normativity have typically focused on moral norms. (Brandom 2000b, p. 371)

Brandom remains agnostic about the relation between his epistemic normativity and the normativity of moral concepts, but claims that it is natural to hope for an enlightenment of moral norms on the basis of a clarification of "the more fundamental class of discursive ones" (Loc.cit.).

However, even if we agree that we must trace the origin of normativity to discursive norms, i.e. to the development of a community of language users involved in a "web of commitments", we might still ask what the source of normativity is in another sense. What, we might ask, is it that ultimately *makes norms binding* to us? In one sense epistemic normativity is most easily displayed as non-circumventable, e.g. by showing in a discursive or formal-pragmatic manner that in arguing you are always already caught up in a normative commitment regarding the truth of your statements. However, although truth is normative for belief, it does not seem to point to an irreducible normative standard for rational agency. It is not tautological to say: "You should not lie". From this it may be argued that we need a more fundamental idea of normativity than the one we can derive from epistemology, perhaps one relating to agency as "answering the ethical question" in Sebastian Rödl's sense.

Karl-Otto Apel argues that "answering the epistemological question" in one sense must already be viewed as "answering the ethical question", as Rödl would put it. In the last essay of *Transformation der Philosophie* (1973), Apel argues that Kant in a certain sense is mistaken when he suggests that a "people of devils"– i.e. a group of people co-ordinating their

[168] Cf. Habermas 1999, p. 147ff.

actions according to instrumental rationality alone – may be capable of acting rationally in the sense of establishing a good political order "if only they have intelligence"[169]:

> Zwar lässt sich nicht bestreiten, dass der logisch richtige Verstandesgebrauch als bloßes Mittel von einem bösen Willen in Dienst genommen werden kann. Insofern ist die Logik als Theorie des normativ richtigen Verstandesgebrauchs eine moralisch wertfreie Technologie (…). Es lässt sich insofern auch nicht sagen, dass die Logik eine Ethik logisch *impliziert*. Dennoch kann auch behauptet werden, dass die Logik – und mit ihr zugleich alle Wissenschaften und Technologien – eine Ethik als Bedingung der Möglichkeit *voraussetzt*. (Apel 1973, p. 398f.)

A decisive aspect of Apel's argument is that, in order to check the validity of arguments, we must always already presuppose a community of thinkers capable of reaching understanding and agreement. Even a "lone thinker" depends *in principle* on the possibility of justifying her arguments within a community of language users. In such a community, Apel argues

> ist die wechselseitige Anerkennung aller Mitglieder als gleichberechtigter Diskussionspartner vorausgesetzt. (Op.cit., p. 400)

According to Apel, this mutual recognition is to be conceived of as a fundamentally *moral* demand on all members of a community of language users.[170]

The question of the ultimate source of normativity has occupied many philosophers in recent decades, among them Christine Korsgaard. Like many Kant-inspired philosophers, she looks to the categorical imperative for an answer. Much like Brandom, she stresses the Kantian idea of autonomy; that being a free, rational agent consists in the capacity to be bound by the norms you lay down for yourself. She argues further that the product of such self-legislation is the categorical imperative, commanding us to act only on a maxim that we could will to be a law.[171]

---

[169] Cf. *Kants Gesammelte Schriften* (Akademienausgabe) Bd. 8; 366 ("Zum ewigen Frieden").

[170] At this point there is a controversy between Apel and Habermas. According to Habermas, the norms that are implicit in all argumentation must be viewed as independent from morality: "[T]he moral principle of according the interests of all equal consideration cannot be justified by appealing to the normative content of presuppositions of argumentation alone. One can invoke this rational potential implicit in discourses in general with this goal only when one already knows what it means to have obligations and to justify actions in moral terms. Knowledge of how to participate in argumentation must be *joined* with knowledge drawn from the experience of a moral community" (Habermas 2008, p. 87). Habermas's motivation for holding back on this point is to avoid having to subordinate the principle of democracy to the moral principle. Apel seems to have a sound point, however, in that it is hard to see how a "principle of discourse", stating the equal consideration of the interests of all those affected, can be viewed as morally neutral, at least as long as the basic moral principle is considered to be an abstract principle of universalisation.

[171] Cf. Korsgaard 1996, p. 98.

Although Kant is ambiguous at this point, his concept of a "Fact of reason" seems to offer a clue to his position. To repeat one of his key formulations on the relationship between freedom and morality:

> Freiheit ist (…) die einzige unter allen Ideen der spekulativen Vernunft, wovon wir die Möglichkeit a priori wissen, ohne sie doch einzusehen, weil sie die Bedingung des moralischen Gesetzes ist, welches wir wissen. (Kant, KpV 5).

As I argued in the introduction, Kant's point may be extended in the sense that free agency is a general condition of rationality. However, it may also be viewed as a clue to a discussion on the source of normativity. Kant adds in the famous footnote to this passage that freedom must be viewed as the *ratio essendi* of the moral law in the sense that without it there would be no moral law (since "ought implies can"[172]). On the other hand, Kant writes, the moral law is the *ratio cognosciendi* of freedom, in the sense that were it not for our "clear thought" of the moral law, we would have no possibility of justifying the assumption of freedom. That the moral law in fact makes a claim on us is what makes us aware of our freedom in the first place, viz. our capacity to act according to reason. In the sense that being a free agent is equivalent to "playing the normative game of giving and asking for reasons" (Brandom), Kant seems to argue that the categorical imperative is the ultimate basis for explicating normativity.

Normativity and morality are not equivalent terms, and not all norms are *moral* norms. Clearly, however, the terms are closely related, and may in many situations replace one another. Is it possible to specify one term as basic in relation to the other? Geert Keil suggests that morality is to be viewed as

> ein Teilphänomen von Normativität und zugleich deren Explikationsbasis. Der explikative Primat des moralischen Sollens kann m.E. in zwei Schritten begründet werden: Erstens ist zu zeigen, dass Normen stets direkt oder indirekt auf Handlungsaufforderungen verweisen, zweitens ist zu zeigen, dass sich das 'Du sollst' nicht von hypotetischen Imperativen her aufklären lässt, sondern nur vom kategorischen. (Keil 2004, fn. 42)

Like Korsgaard, Keil points to the categorical imperative as an ultimate explicative basis for normativity. This is also in line with Marcus Willaschek, who argues that universal ethics might be seen as "the ultimate, all-inclusive level of normative standards", thus as the

[172] Cf. e.g. KpV 54.

explicatory basis of rationality as such.[173] From this perspective, a "people of devils" would have to be considered as a borderline case of rationality, made possible only against the background of a larger community of rational beings capable of making moral decisions.

Willaschek points to the "existentialistic" idea that we cannot *avoid* being autonomous. To the degree that we are rational agents, we are condemned to freedom, as Sartre phrases it. Given Kant's "reciprocity thesis",[174] this implies that we cannot act except on the basis of a fundamental morality. Hence, the normative evaluation that is crucial to instrumental rationality ultimately rests on morality, not – like ethical naturalism would have it – the other way around:

> If the justification of moral demands was limited by considerations of instrumental rationality, we would have to reject as unjustified moral demands that require acting against one's instrumentally defined interest. But in morally wrong action, we often have to admit that a moral demand on ourselves in fact was justified, even if we have decided for instrumental reason not to fulfil it. This, I believe, can only be understood if we regard moral claims as more fundamental than instrumental ones. (Willaschek 1998, p. 199)

Willaschek, like Keil, argues that hypothetical imperatives, or instrumental rationality, must be explicated by reference to a categorical imperative. Hypothetical imperatives give only relative reasons for acting. I may reason: "To achieve a, do b", but the question remains: Why try to achieve a? Every instrumental reason can be challenged – even a reference to the maintenance of my life: "But why should I aim at surviving?" In Sebastian Rödl's words, what we stand in need of is a *last reference point* of action:

> [E]in letzter Bezugpunkt der ethischen Frage (…) ist einer, zu dem es keinen übergeordneten gibt. (Rödl 1998, p. 50)[175]

Rödl pinpoints the normativity of action by equating acting with "answering the ethical question". Instrumental rationality is not enough, he argues, in order to explain the normativity of action. The last reference point of action is always the ethical dimension, namely that by acting I answer the question "What should I do?" In this sense, to the degree

[173] Cf. Willaschek 1998, p. 199.
[174] I.e. that freedom and morality are reciprocal concepts, cf. e.g. GMS 446f.; cf. KpV 52, §6. Cf. Allison 1990, Ch. 11.
[175] Rödl specifies this further by (in the spirit of Tugendhat) distinguishing between an *ethical* and a *moral* sense of this last reference point, respectively "happiness" and "the simply good". It would be too much of a diversion to go into this discussion here.

that something is a rational action, it is always *in principle* an attempt to "do the right thing", although such attempts frequently fail.[176]

However, in order to steer clear of what Rödl diagnoses as "the Cartesian fallacy"[177], it is important that the last reference point is not understood in a methodological-solipsistic sense, but is given an intersubjective interpretation. Rödl argues that

> die Dimension des Sollens, in der ich mich zu mir verhalte, mir nur offen steht, indem ich mich auf einen anderen beziehe, der sich in dieser Dimension zu sich verhält. (Op.cit., p. 269)

In other words, in order to view the moral law as the ultimate explication basis of normativity, its status as a Kantian "fact of reason" must be linguistic-pragmatically reinterpreted. Apel suggests that it should be understood in terms of "the a priori of a community of language users".[178] A main aim for me in this thesis is to argue that the approach to free agency must be reconstructed on the basis of intersubjective language in order to avoid dualism. In the final two chapters I take a closer look at the challenge of dualism with regard to the free agency problem.

[176] Cf. the Kant-Korsgaard discussion above, Ch. 2.3.
[177] Cf. above, Ch. 4.4.
[178] Cf. Apel 1973, p. 429: "'Faktum der Vernunft', sofern es als *Apriori der Kommunikationsgemeinschaft* begriffen wird".

Chapter 6: Language and world – Two theses of unity

> [O]ur language of deliberation is continuous with our language of assessment,
> and this with the language in which we explain what people do and feel.
> (Taylor 1989, p. 57)

> Die raumzeitliche Welt ist nur einmal da, und verschiedene Disziplinen
> untersuchen verschiedene Aspekte oder Eigenschaften dieser einen Welt.
> (Keil 2007b, p. 46)

6.1 Introduction: Agency, language and world

In Chapter 2, I argued that within a Kantian frame it is impossible to recognise the concrete action of any concrete subject as free. Kant's philosophy is a mentalistic philosophy of mind, in the sense that it does not take into account language as a precondition of thought and subjectivity. A transition to a pragmatic philosophy of language is needed. However, a transformation from a methodical-solipsistic philosophy of mind to a pragmatic philosophy of language is not enough. Even within a linguistic-pragmatic idiom, careful steps must be taken in order to steer clear of dualism. Not much is gained by replacing the conceptual divide between "Naturbegriff" and "Freiheitsbegriff" with a distinction between two languages or language games.

Dualism – in all its varieties – is a threat to any viable theory of free agency. Or, in a positive formulation: A robust conception of free agency is the key to avoid a dualist scheme. However one tries to make sense of a divide between two worlds, kingdoms, conceptual frames, languages, or language games; whether one draws the line between an empirical and an intelligible world, a concept of freedom and a concept of nature, or a language game of intentionality and a language game of causality, intentional agency seems to belong on both sides of the "gulf".

Agency is therefore a starting point from which to explore the connection between mind and world. The very existence of intentional agency seems to constitute a strong argument against dualism. Searle talked about the mind-world connection as having two "directions of fit": mind-to-world (perceptual experience) and world-to-mind (intentional agency).[179] If the mental and the physical were to belong to two different worlds or spheres, then how could we account for this interaction between mental and bodily events? The causal aspect of agency cannot be explained away.[180] Thus, intentional agency may be seen as a problematic concept, one that must somehow be legitimised within our existing world-view. A more promising alternative, however, is to start out from intentional agency as the most immediately given, being as it were our access to the world. This is the ingenious move of an interventionist theory on causality; a move that has a potential of being generalised.

In this chapter, I start out from Apel's identification and criticism of "the alternative conception" (6.2), which can be regarded as a form of dualism between the conceptual frames – or language games – of explanation and understanding. Apel argues that we must assume a hermeneutic dimension "between" these two frames, holding them together in order to avoid a dualistic conception of language, or a conception of language as divided into language and meta-language. In the *Tractatus*, Wittgenstein shows an acute sense of this problem, but then ends up evicting both meta-language and the language of understanding in one and the same move. Against this, in 6.3, I defend a thesis of the *unity of language* – in the sense that a complete language functions as its own meta-language, and that indexical expressions are a key to this function. Certain systematic features of a "complete" language are examined in 6.4, 6.5, and 6.6. In 6.7, I look into and criticise Habermas' theory of a "differentiated world-view," and argue further in 6.8 for a *unified concept of the world*. In 6.9, I sum up the chapter and discuss how the concepts of *veritative* and *performative* being may be employed in order to address some of the sound intuitions behind Habermas's suggestion of a differentiated world-view.

Against my criticism of Habermas' "multiple worlds", it might be countered that it is a waste of effort, since these are innocent metaphors used to shed light on the need to differentiate between various kinds of relations between subjects and the world. However, metaphors are not always innocent. Sometimes they point in the wrong direction and thereby lead to epistemic, ontological, or even practical misconceptions. An example I have already

[179] Cf. Searle 1983, p. 96f.
[180] Cf. Rödl 1998, who also stresses the causal sense of perceptual experience, cf. p. 130: "Der kausale Sinn von 'wahrnehmen' ist irreduziebel".

discussed is the inner-outer metaphor,[181] and as I will show, this image plays a part in Habermas's multiple world allegory as well.

The important differentiation of relations between language (users) and the world easily collapses into a version of dualism. This is a tough balance for any theory of agency, given the unavoidable tension between different perspectives on action.[182] Dualism should, however, be avoided – including any kind of property dualism, aspect dualism, or language (game) dualism. A certain kind of "perspective plurality" is necessary, however it is vital to display 1) how these perspectives can be integrated in *one language*, and 2) that these perspectives all constitute ways of relating to *one world*.

6.2 A critique of "the alternative conception"

Kant's attempt to determine the conditions for the possibility of objectively valid experience while maintaining the idea of transcendental freedom led him to carefully distinguish the area of causal explanation (the phaenomenal realm) from the area of "causality through freedom" (the noumenal realm). I argued in Chapter 2 that this amounts to a strict dualism, in the sense that Kant excludes any mediation between "Freiheitsbegriff" and "Naturbegriff". It appears, as Karl-Otto Apel writes,

> als ob zwischen den Menschen als Adressaten seiner Ethik und den Menschen im Sinne einer empirischen Anthropologie überhaupt keine Identität bestünde. (Apel 1988, p. 76)

This conceptual boundary between the areas of "ought" and "is" corresponds to what Apel has called "the alternative conception".[183] This is closely related to a "scientistic fallacy" that Apel attributes to Kant and subsequent Kant-inspired epistemology, as well as the "unity of science" tradition within philosophy of science:

> Aus der 'Kritik der reinen Vernunft' (…), d.h. aus der von Kant in ihrem Rahmen kritisch begrenzten Gegenstands-Konstitution möglicher objektiv gültiger Erfahrung, konnte – und kann – man nur wissenshaftstheoretische Konsequenzen im Sinne eines radikalen *Szientismus* ziehen, d.h.

[181] Cf. above, Ch. 4.4.
[182] Cf. above, Ch. 1.3.
[183] Cf. Apel 1979 p. 204 and 216ff.

> Konsequenzen im Sinne der Einschränkung der Idee der Erkenntnis auf neuzeitliche Naturwissenschaft (science) und – allenfalls – auf im Sinne ihrer Kategorien reduzierte Quasi-Naturwissenschaft. (Apel 1979, p. 61)

"The alternative conception" may be illustrated with reference to the way we talk about actions. Following this conception, there are methodologically speaking only two relevant types of "why?" questions to be asked regarding actions: 1) Questions and answers in the first and second-person present tense, regarding acts and their legitimacy – where we ask for *reasons*, and 2) Questions and answers in the third person, regarding the explanation of events in the objective world – where we ask for *causes*.

> [M]an hätte nur zu unterscheiden zwischen Fragen einer normativen Wissenschaftslogik und Ethik (oder deontologischen Logik?) in bezug auf das, was *gilt*, und Fragen der empirischen (Natur-) Wissenschaft in bezug auf das, was 'der Fall ist'. (Op.cit., p. 216)

What is missing from this picture, according to Apel, is the possibility to perceive the person(s) we talk about in the third person as contemporaries or historical subjects, i.e. as "*virtuelle Interaktions- bzw. Kommunikationspartner*"[184] whose acts we are trying to *understand*, not causally explain. Apel wants to make room for a hermeneutic understanding of other people, not as objects for scientific observation, but as "co-subjects".

Apel discusses "the alternative conception" in light of the dispute within theory of science known as the "explanation vs. understanding controversy" since Dilthey, or as "the reason vs. causes debate (cf. above, Ch. 3.2), with prominent participants such as Karl Popper, Carl Hempel, William Dray, Peter Winch, and G. H. von Wright. The scientistic fallacy is easily recognised in Popper's and Hempel's one-sided emphasis on causal-nomological explanation, but Apel points to certain scientistic prejudices even in the opposite camp. These prejudices result, according to Apel, from disregarding the *hermeneutic* dimension of understanding. Von Wright, as one example, gives an account of linguistic communication as just another case of "hypothetical" and "provisional" verification of the existence of intention. He thereby assumes that interlocutors in communication consider the efforts of each other as "verbal behaviour", i.e. as objects of teleological explanation which

> does not in principle afford more direct access to the inner states than any other (intentional) behaviour. (von Wright 1971, p. 113)

[184] Cf. Apel 1979, p. 216.

According to Apel, this is an example of a scientistic fallacy due to its neglect of the hermeneutic dimension of understanding. If von Wright was right, Apel argues, it would be possible to examine other people's intentions simply by *observing* them, without presupposing an always already functioning understanding (Verständigung) *with* them. However, this ultimately implies methodological solipsism: The only way towards understanding other subjects would be to make them (i.e. in principle all possible communication partners) objects of explanatory hypotheses on the basis of observation – but in that case the subject of cognition must be able to understand at least her own hypotheses or thoughts as *pre-linguistic* or *private-linguistic*.

The alternative is to assume that my own thoughts – including all my explanatory hypotheses about other people's intentions – are in principle always already linguistically mediated. Hence the forming of hypotheses (about natural events as well as the actions of other people) presupposes a linguistic communication that in turn is not normally understood on the basis of explanatory hypotheses. Apel wants to radicalise von Wright's own approach to action and consciousness. As presented above (Chapter 3.5), von Wright argues that our understanding of ourselves as agents intervening in the course of events is a conditions for the possibility of causal explanation. Apel now points out that a basic (linguistic) understanding with other subjects is a condition for the understanding of actions in general, including our own intervening actions:

> Das *Verstehen der Mitteilungen anderer* scheint mir (…) in ähnlicher Weise die Bedingung der Möglichkeit der *intentionalen Erklärungen* zu sein wie das *Verstehen der eigenen Interventions-Handlungen* die Bedingung der Möglichkeit von *Kausalerklärungen* ist. (Apel 1979, p. 178f, fn. 112a)

Von Wright seems to be committing a scientistic fallacy by assuming a principally pre-linguistic consciousness-in-acting. In contrast, the rationalistic-pragmatic approach to language and action represented by Apel and Habermas systematically stresses the reflective aspect of language, and argues that rational agency must be explicated on the basis of specifically communicative action. Cf. e.g. the following passage in Apel 1979:

> Durch die *Reflexionsperspektive des Gesprächs* werden Kommunikationshandlungen im engeren Sinn – Sprechakte – und verstehbare menschliche Handlungen überhaupt sowohl unterschieden als auch zueinander in Beziehung gesetzt (…) Gäbe es keine *Kommunikationshandlungen*, für die die Intention

> des Verstandenwerdensollens und damit die Selbstreflexion konstitutiv ist, so gäbe es überhaupt keine verstehbaren menschlichen *Handlungen*. (Apel 1979, p. 217f.)

Brandom states a similar point in this way:

> [R]ational agency, on which instrumental behavior is modeled, depends essentially on specifically linguistic practices, including asserting. It follows that simple, non-linguistic, instrumental intentionality can not be made fully intelligible apart from considerations of the linguistic practices that make available to the interpreter (but not to the interpreted animal) a grasp of the propositional contents attributed in such intentional interpretations. (Brandom 1994, p. 155)

Habermas, Apel, and Brandom all view communicative action – where a basic understanding with my communication partners is always already anticipated – as a necessary basis for the understanding of rational agency in general.

The Neo-Wittgensteinian tradition that von Wright belongs to has been criticised for wanting to replace Cartesian substance dualism with a "new dualism", distinguishing two "language games" or categorical frames that mutually exclude each other: 1) The language game under which we talk about things, events, and causes, subsumable under the laws of nature, and 2) The language game of people, actions, intentions, reasons, and rules.[185] This distinction resembles the Kantian distinction between the two categorical frames of *Naturbegriff* and *Freiheitsbegriff*. However, Apel's view is that the Neo-Wittgensteinian, linguistic-analytical version is more neatly "transcendental-pragmatically transformed" than Kant's transcendental idealism. This means that it is more easily adapted to an idea of complementarity. Apel wants to show that the language games of explanation and understanding mutually exclude and presuppose each other. Again, he views this as an extension of von Wright's theory. Von Wright's argument – that the concept of a causal necessity of changing states within an adequately isolated system is preconditioned by a concept of experimental action – is generalised to an argument regarding the complementarity of language games or conceptual frames.

Through a concept of communicative experience, Apel wants to make room for a dimension of hermeneutic understanding "between" objectivated nature and the subjective-intersubjective conditions for the objectivation of nature. This means an extension of the concept of science. It also implies a transformation of Kant's – and Husserl's – concept of the transcendental subject as a condition of the possibility of objectivated knowledge of nature.

[185] Cf. e.g. C. Landemann (1965-66): "The New Dualism in Philosophy of Mind".

The concept of a transcendental subject can no longer be that of a pre-linguistic and pre-communicative "synthetic unity of self-consciousness":

> Denn es ist davon auszugehen, dass, ebenso wie das intersubjektiv gültige Sinnsverständnis von *etwas als etwas* – auch schon das *Selbstverständnis des Ich* immer schon *sprachlich artikulierbar* und insofern auch durch die *hermeneutische Synthesis der Kommunikation* vermittelt sein muss. Der *Sinn* aller Gedanken muss insofern *öffentliche Gültigkeit* besitzen und kann daher nicht als *noematische* Leistung einer prinzipiell einsamen *intentionalen Noesis* verstanden werden. (Loc.cit., p. 326f.)

The transcendental subject cannot be understood as a self-sufficient and closed "synthetic unity of self-consciousness"; it is always already engaged with other members of a communicative community in a "Verständigung über etwas".[186]

According to Apel, the "understanding vs. explanation controversy" in the theory of science can be (dis)solved if it is recognised that not every transition from the first and second person to the third person is a transition to an observational standpoint. Apel distinguishes between two types of transitions: 1) The transition to an observational standpoint with a theoretical relation between the subject and a third-person object (objectivation), and 2) The transition from "I-you" communication to communication with virtual co-subjects in the third person (secondary objectivation).[187] From this perspective, the cognitive interest of hermeneutic understanding appears to be complementary to, instead of competing with, the cognitive interest of causally explaining natural events.

6.3 The unity of language (and meta-language)

"The scientistic alternative conception" is in many cases the outcome of an attempt to secure the language of science from the confusions and obscurities of colloquial language. Traditionally, many empiristically oriented philosophers have tried to specify the conditions under which a language could be a precise means of rendering true, scientific facts. From John Locke and onwards, the hope of many empiricists was to overcome difficulties and misunderstandings in philosophy and science by precisely defining the meaning of a word by

[186] Apel's phrase "Verständigung über etwas" expresses the *complementarity* between a purely observational subject-object-relation and a communicative subject-co-subject-relation, and covers both the *understanding* of meaning and the *agreement* on standards of validity (questions of truth and falsity).
[187] Cf. Apel 1979, p. 323 and p. 173.

reduction to a "simple idea". The problem with this endeavour is how to ensure a common meaning of words, since the meaning of words had to be traced back to "private ideas" in the mind of the language user. How can one speaker know that other speakers – assuming that they mean something at all – speak about the same thing, i.e. connect the same private ideas with their words as she does herself?

Wittgenstein's *Tractatus Logico-Philosophicus* (TLP) is an attempt to overcome this problem, by connecting Locke's theory of a reduction to private ideas with Leibniz's view of language as an a priori intersubjective calculus. Wittgenstein's basic idea is that the perfect logical form of a (universal) language is hidden under the rough surface of colloquial language. The form of language is for every language user a priori identical with the form of the world. Hence, for every state of affairs there is an ideal sentence uniquely corresponding to it. To Apel, the *Tractatus* represents a standard case of "the alternative conception".[188] However, in this case only one side of the divide is equipped with a language. Every possible true proposition belongs to the totality of natural science (cf. TLP 4.11). Anything else – including the sentences of the *Tractatus* itself – is strictly speaking meaningless and unspeakable. Philosophy constitutes a "ladder" for us to climb in order to understand the relation between language and world. The "ladder" should be thrown away once considered, whereafter silence reigns within the realm of freedom (cf. TLP 6.54 and 7). Only the trivial – the scientific picturing of the world (everything that is the case) – can meaningfully be said.

Wittgenstein analyses sentences of the form "A thinks that p", "A believes p", and "A says p" as really being of the form "'p' says p" (TLP 5.542). As such they are meta-linguistic, in other words strictly speaking meaningless, since they do not depict a state of affairs in the world. This way, Wittgenstein avoids getting mixed up in a hierarchy of meta-languages that led some of his predecessors into an unfortunate regress, or into presuming and leading everything back to an unexplained, already understandable meta-language or "meta-thought". We cannot go outside of language to explain language. According to Wittgenstein's "showing-saying" doctrine, the logical form that is common to language and the world – and that thereby renders possible the logical depiction of facts through utterable sentences – cannot itself be depicted. It cannot be *said*; it can only *show itself* in our language use:

> Der Satz kann die logische Form nicht darstellen, sie spiegelt sich in ihm.
> Was sich in der Sprache spiegelt, kann sie nicht darstellen.
> Was *sich* in der Sprache ausdrückt, können *wir* nicht durch sie ausdrücken.

[188] Cf. Apel 1979, p. 216.

> Der Satz *zeigt* die logische Form der Wirklichkeit.
> Er weist sie auf. (TLP 4.121)

Wittgenstein's sound, Kantian point is that the perspective from which experiences are had, cannot itself be part of the experience, but is "the limit of experience". Hence the perspective – the "I think", if you wish – cannot itself be depicted in the act.[189] However, a prohibition of meta-language cannot take care of this transcendental-philosophical intuition in a satisfactory manner. Rather, it paradoxically leads to a kind of logical naturalism that made the *Tractatus* – together with Hume's *Treatise* – favourite reading material among logical empiricists. Although the *point* of the saying-showing doctrine is a thoroughly non-naturalistic one, this produces no results on the pragmatic level. Hence, on *this* level, the radical transcendentalism of the *Tractatus* actually coincides with radical empiricism.

Wittgenstein's efforts to protect the transcendental limit of the world from being depicted as a fact in the world dramatically limit the scope of language. Only one type of sentence is allowed within this system, namely the depiction of facts in propositions, while the performance of the subject disappears into the unspeakable realm of logical form. To Wittgenstein at this point, it seems that the only way to protect the subject from naturalisation is to deny it linguistic articulation altogether. Hence, the actual, hermeneutic understanding *between* subjects becomes unaddressable too. Under the *Tractatus* construction, a plurality of subjects – if conceivable at all – contributes nothing to the logical form of language. Wittgenstein reflects upon this himself:

> Here we see that solipsism strictly carried out coincides with pure realism. The I in solipsism shrinks to an extensionless point and there remains the reality coordinated with it. (TLP 5.64)

Strawson elaborates in *Individuals*: For the solipsist the subject-place in language – the "I" – is simply redundant.[190] In a logically perfect language as is outlined in the *Tractatus*, the I disappears from language in the sense that it coincides with the world, and hence – logically speaking – with language. This means that in language all subjects – if more than one is conceivable – must a priori be in total agreement. When the subject vanishes from language,

[189] Cf. e.g. Kant, KrV B 132: "Diese Vorstellung aber ist ein Aktus der Spontaneität, d.i. die kann nicht als zur Sinnlichkeit gehörig angesehen werden".
[190] Cf. Strawson 1959, p. 73: "[T]he true solipsist is (…) one who simply has no use for the distinction between himself and what is not himself". In his later writings, Wittgenstein pointed out that "I" – according to its rules of use – must have "neighbours", cf. Tugendhat 1976, p. 94f.

so does the possibility of perspectival change within language through indexical shifts.[191] Accordingly, the possibility of ascribing intentions and intentional actions to other subjects in the world disappears, and with it the prospect of conceiving possible other subjects *as* subjects. In the *Blue Book*, Wittgenstein calls himself to account, criticising the solipsistic implications of a *Tractatus* relation between language and world:

> When I made my solipsist statement, I pointed, but I robbed the pointing of its sense by inseparably connecting that which points and that to which it points. I constructed a clock with all its wheels, etc., and in the end fastened the dial to the pointer and made it go round with it. And in this way the solipsist's 'Only this is really seen' reminds us of a tautology. (Wittgenstein 1958, p. 71)

The possibilities of perspectival multiplicity and transitions are excluded from the *Tractatus* language. This is, however, attended to by Wittgenstein in his later works. In the *Philosophical Investigations* (PU), Wittgenstein shows that the meaning of predicates regarding consciousness and intentional action necessarily arise from two roots:

> 1) The *performative* or *act constitutive* use: The use of mental and action predicates in the first-person present tense does not *depict*, but *constitutes* the act itself,[192] and
> 2) The *assertoric* or *act descriptive* use: The use of the predicates in other temporal and personal forms *describes* who is doing or has done something, or who is in a certain mental state.

The two-rooted meaning of the predicates is – in accordance with the Private Language Argument (PLA) – accessible to a plurality of subjects. This is accomplished by the possibility of indexical shifts from first to third person, from present to past tense, and so on, and makes it possible to ascribe intentions and actions to each other. Act performance and act description are in this way displayed as two necessary elements of a complete language. Thus, indexicality comes into sight as a core element of a complete language.

---

[191] Cf. Carson 2002, p. 59.

[192] This covers both purely *expressive* utterances such as "I am in pain" – which do not *describe*, but *replace* the "natural" pain expression (cf. PU § 244) – and paradigmatic *performative* utterances in Austin's sense, e.g. "I promise you…" The latter do not replace anything, but are *essentially* linguistic acts. Thus, only language users can make promises, whereas most animals can in some sense express their pain. In the following, I mostly do not distinguish between expressive and performative utterances, as they fulfil more or less the same function within the view of language I suggest, and since they are in my view overlapping, or constitute a continuum, cf. a sentence such as "I beg you".

Wittgenstein emphasises that the "being in pain" predicate has a uniform meaning, even if it serves a different purpose in the first type of sentence ("I am in pain") than it does in the other ("She is in pain", "I was in pain"). The uniformity of meaning is not least a condition for the learnability of these predicates.[193] The relation of predicates in the first-person present tense and in other forms cannot be regarded as one of translation: It is not an inter-linguistic, but an *intra*-linguistic relationship.[194] It can, however, be regarded as a *meta-linguistic* relationship, in the sense that one of the sentences ("She is in pain") can function as a report of what is expressed in the other ("I am in pain"). This is why colloquial language – unlike the ideal language proposed by Wittgenstein in the TLP – does not stand in need of an independent meta-language to explicate the meaning of its sentences. This is a characteristic mark of colloquial language: Everything that can be *done* in such a complete language can also be *described* in that same language.

In "Act Performance and Description" (1985), Audun Øfsti criticises different versions of objectivism within the study of human agency. His approach is to show that the scientistic ideals of the objectivist programme involve an illegitimate absolutation of the third-person perspective:

> [M]y criticism of the objective-psychological attempt can be formulated as a criticism of an illegitimate reduction: the reduction involved in a reconstruction of our full-blown colloquial language within one of its dimensions, namely the descriptive 'third-person' dimension. (Øfsti 1985, p. 15)

Against this, Øfsti defends a thesis of the *unity of language*:

> What natural languages allow, but which would be lost through an objective-psychological reconstruction of intention-*attributions*, is the possibility of keeping the different 'person positions' (first, second and third) within the *same* language, including the logicality of 'substituting the first person.' (…) That this unity obtains means that the language of the observed performer/'intender' and the language of the reporting observer do not fall apart in language and metalanguage as separate languages. The language of the other and the 'metalanguage' in which I describe him – or the other way around: my language (in terms of which I express myself and perform the speech acts I perform) and the 'metalanguage' in which I am reported – remain one single language and a common medium. (Op.cit. p. 16)

[193] As Wittgenstein stresses – see e.g. PU 244ff. To me, it seems that this argument is closely related to his Private Language Argument and to what I have referred to as the Intersubjectivity Thesis, see above, Ch. 5.3.
[194] This against e.g. Quine's concept of "radical translation" and to a certain degree also against Davidson's concept of "radical interpretation", cf. above, Ch. 5.3.

The "glue" holding a complete language together – keeping it from falling apart into language and meta-language – is the function of indexicality, making transitions between different speaker positions and perspectives possible.

6.4 The horizontal and the vertical indexical system

Øfsti distinguishes between two "indexical systems" of a complete language: the horizontal system and the vertical system.[195] The horizontal indexical system corresponds to Ernst Tugendhat's approach in *Vorlesungen zur Einführung in die Sprachanalytische Philosophie* (1976). Tugendhat argues that indexicality is the perspectival basis necessary for having perspective-transcending truth. Hence, indexicality is not only consistent with objectivity; indexical expressions are rather an indispensable condition for our ability to refer objectively to things and states of affairs in the world. Indexical expressions are situation-relative, but they are nevertheless needed in order to establish true or false, "eternal" (index-free) propositions. Indexicals are an integral part of a system of reciprocal substitution, which permits a connection between situations of use – where assertions are made – and situations of verification/falsification – where their truth value can be settled.

In his *Vorlesungen*, Tugendhat outlines an extensive system of expressions – including indexical expressions such as demonstratives, time-and-place adverbs, personal pronouns and verb inflections, but also "objective" location descriptions (related to a fixed origo, e.g. calendar time, latitude and longitude), and the sign of equation (linking e.g. indexical expressions with names, in order to build sentences such as "This = a"). This system allows shifts in the speaker's position to be "compensated for" by altering the indexical expressions in the sentence, so that a constant (true or false) propositional content can be maintained from different speaker's perspectives. The system makes it possible to have a meaningful use of predicates even in situations where the qualities referred to are not immediately present. This is in contradistinction to the more limited signal languages of animals and small children, who are only able to use "quasi predicates," e.g. by making a certain gesture or saying "bow-wow" whenever a "dog quality" is present.[196]

[195] Cf. Øfsti 1997, p. 153ff. He emphasises that the two systems cannot be held strictly apart, but are interconnected and partly overlapping.
[196] Cf. Tugendhat 1976, p. 208: "Das Charakteristische der Quasiprädikate ist, dass bei ihnen die Verwendungssituation und die Erklärungssituation von derselben Art sind."

Tugendhat argues that if language did not carry with it the possibility of preserving the same propositional content through different perspectives, we would not be able to transcend our own private and momentary perceptual field, and we would have no access to *truth*. The indexical system that he articulates constitutes our very access to the world in terms of "veritative being".[197] Øfsti sums up:

> Mit diesem System öffnet sich die Welt in emphatischem Sinne, mit ihm ist die Ebene der konditionalregulierten Quasiprädikation verlassen. (Øfsti 1997a, p. 157)

The *horizontal indexical system* allows us to step out of the "here-and-now" perspective and gain access to objectivity and truth. However, this is a "horizontal" movement, in the sense that we move on the level of *describing* the world, whereby we grasp the world in the sense of "veritative being". But indexical mechanisms can also be seen as a condition for a complete language in another sense, namely in making possible the shifts between use and mention. By this I mean the shifts between *acts* and *description of acts* – where the latter of course takes place in new (speech) acts. This is what Øfsti calls the *vertical indexical system*, referring to the possibility of meta-linguistic statements opened up through this system, with "vertical" pointing to the possibility of *ascending* through "meta-levels" of language.

This is one more sense in which a system of indexical transitions constitutes our grip on the world. The vertical indexical system involves personal pronouns and verb tenses, and renders possible the transitions between (linguistic and non-linguistic) expressions and illocutionary acts ("I hurt", "I promise…"), and the mention of these expressions and acts in other speech acts ("She is in pain", "I promised…"). Such transitions are by no means disregarded by Tugendhat. On the contrary, they are a central topic in his *Selbsbewusstsein und Selbstbestimmung* (1979). However, there are certain ambiguities connected with his treatment of them. Tugendhat's point of departure is Wittgenstein's analysis in *Philosophical Investigations* (1953). Wittgenstein points to the fundamental difference between "I am in pain" and "He is in pain". The first sentence is an expressive utterance, where the speaker expresses his pain, while the second is a description. Nevertheless, it belongs fundamentally

---

[197] Cf. Tugendhat 1976, p. 60ff. "Veritative being" shows itself in that something (i.e. propositions) can be true or false. According to Tugendhat's linguistic-pragmatic approach, what *exists*, what is *real*, is conceivable only in terms of what is *true*. The concept is drawn from the Aristotelian dictum that "to make a true statement is to say of what is, that it is, or to say of what is not, that it is not."

to our linguistic understanding that the "being in pain" predicate means the same in both of the speech acts: Pain is the same whether I ascribe it to others or express my own pain.[198]

Tugendhat describes this relationship between "I am in pain" and "She is in pain" as a "veritative symmetry", which is to say that the first sentence is true whenever the second sentence, when uttered by someone who by saying "she" means me, is true.[199] In a Wittgensteinian spirit, he separates expressions and illocutionary acts from descriptions. He emphasises that only the second sentence (of the form "She φ," where φ designates an expressive or performative predicate) implies the identification of something, while the first sentence ("I φ") is expressive, and does not describe or pick out an object at all. However, Tugendhat sees no other way to explain the unity of the predicates' meanings and the similarity in structure between the expressions/performances and the descriptive sentences than by the phrase "veritative symmetry," i.e. by assimilating performative acts into the prevailing scheme for true and false propositions. In other words, he assembles all sentences under the category of assertoric, i.e. true and false, sentences. This may be diagnosed as a "semantistic prejudice" in Karl-Otto Apel's sense, i.e. a reductionistic view of language whereby

> dasjenige Bedeutungsmoment der performativen Phrase verloren geht, durch das sie, und nur sie, die sprachlich explizite effektive Selbstreflexion der menschlichen Sprechhandlungen ermöglicht. (Apel 1980, p. 47f.)

Tugendhat's semanticist prejudice consists in his levelling of all utterances, including performative and expressive utterances, to the assertoric dimension.

In the *Tractatus*, on the other hand, Wittgenstein paradoxically avoids the semantistic fallacy by altogether banishing the performative dimension from language. However, as Apel points out, this move cannot overcome the problems resulting from "the alternative conception". I have analysed this conception as a kind of linguistic dualism, i.e. as an understanding of language as really consisting of two languages:

> 1) A subject language in the first- and second-person present tense, in which the subject expresses itself, intends and performs acts, and relates to the (rules of the) use of language.
> 2) An object language in the third person, in which the world is described.

[198] Cf. PU §§ 244ff.
[199] Cf. Tugendhat 1979, p. 88.

Although the *Tractatus* forbids the first of these "languages", the dualism charge is not overcome simply by suppressing one of the two realms. The language-silence (saying-showing) duality of the *Tractatus* does not solve the fundamental problems of "the scientistic alternative conception".

Wittgenstein's *later* philosophy, however, counteracts dualism by showing that mental predicates have a two-rooted, but unified meaning: The expressive or performative use (expressions of "my own case": "I φ") and the descriptive use (identification based on outer criteria: "She φ") both attribute to the single meaning of the predicate. In this sense, Wittgenstein hints at a vertical indexical system, but in the anti-systematic spirit of the *Philosophical Investigations* he does not move far in the direction of elucidating this aspect of language. His decisive break with the language theory of the *Tractatus* is couched rather in the break with the semantistic presupposition that sentences always function in the same way:

> Das Paradox verschwindet nur dann, wenn wir radikal mit der Idee brechen, die Sprache funktioniere immer auf *eine* Weise, diene immer dem gleichen Zweck: Gedanken zu übertragen – seien diese nun Gedanken über Häuser, Schmerzen, Gut und Böse, oder was immer. (PU, § 304)

In this context, Wittgenstein emphasises that expressive and performative utterances are not descriptions of something in the world:

> Wenn ich sage 'ich habe Schmerzten' weise ich nicht auf eine Person, die die Schmerzten hat, da ich in gewissem Sinne garnicht weiß, *wer* sie hat. (PU, § 404)

Wittgenstein suggests that the use of "I φ"-phrases is nothing but a sign to other people, in order for *them* to know who is in pain. However, in that case, such phrases seem to become redundant, since other (more or less "natural") pain behaviour would be just as efficient for signalling "where the pain is".[200]

I think Wittgenstein is right in denying that "I φ"-phrases convey knowledge, since in these cases there is no (use for) true-or-false deciding criteria.[201] What is underexposed in the *Philosophical Investigations* is, however, that the syntactic *form* of expressive "I φ"-sentences

[200] Cf. Wittgenstein's own remark in PU § 407.

[201] There may be no harm in *calling* what is expressed in performative or expressive utterances "agent *knowledge*" (Taylor) or "Handlungs*wissen*" (Øfsti, who refers to Fichte's wording of it as "das, wodurch ich etwas weiss, weil ich es thue" – cf. Fichte 1983, bd.I/4, p. 216f. and Øfsti 1997, p. 161). However, the concept of *knowledge* is strongly tied to the concept of (intersubjectively available) *criteria*. Hence I prefer to use the term *consciousness-in-acting*, cf. "Handlungsbewusstsein" (Kuhlmann) or "intention-in-action" (Searle).

link them together with the descriptive forms ("She φ", "I φ-ed"), whereby the unified meaning of the concepts is secured. Indeed, the common form of "I φ" and "She φ" sentences is what led Tugendhat to suggest a "veritative symmetry" between them. I have argued that the first kind of sentence does not belong within the category of true-or-false sentences, but is strictly expressive or performative. However, the syntactic symmetry between the sentences is not a simple coincidence; it ties the *expression* of one's own act consciousness to the *ascription* of act consciousness to other subjects (and to oneself, e.g. in the past tense). This relation is maintained by the *vertical indexical system*. Thanks to this system, it is possible for us to evaluate (other people's and our own) claims and expressions, instead of surrendering it all to a *Tractatus* silence. This is constitutive of the intersubjective dimension of language, and explains how we are able to speak *with* each other, as opposed to a *Tractatus* situation where each individual stares at the world – "parallel" as it were – describing it in a private language which a priori is shared by all.[202] According to (a strong version of) the Private Language Argument, the possibility of intersubjectively shared meaning is a necessary condition for language as such.[203]

6.5 The formal criteria of a complete language

The "anti-systematic spirit" of Wittgenstein's *Philosophical Investigations* reinforces his break with the *Tractatus* theory of language. What is left underexposed, however, is that our access to the world is made possible by certain *systematic* features of a complete language. Wittgenstein's ambiguous use of the key term "language games" is a point in question. Mainly, it seems that the term refers to limited functions within a complete language (like the tools of a toolbox, cf. § 11). Sometimes, however, he seems to imply that a single language game can constitute a language in itself. In PU §§ 18-19, Wittgenstein argues that a language far more limited than our own is easily imaginable:

> Man kann sich leicht eine Sprache vorstellen, die nur aus Befehlen und Meldungen in der Schlacht besteht. (…) Und eine Sprache vorstellen heißt, sich eine Lebensform vorstellen. (PU § 19, cf. also §§ 2 and 6)

[202] Cf. Øfsti 2008, p. 75.
[203] Cf. above, Ch. 5.3.

The question is whether such a meagre language really is imaginable, or if certain features or constituents of a complete language are missing in such a case. Wittgenstein's thesis is that the unlimited amount of "language games" does not necessarily have anything else in common than certain "family resemblances" (PU § 66-67). However, one could argue that there must be something in common for all "language games" in the sense of complete *languages*. A crucial feature seems to be the possibility of *translation* between whole languages, whereas this is certainly not possible between language games in the sense of the "tools in language", e.g. between giving orders and guessing riddles (cf. PU § 23). In the following quote, Apel seem to understand "language games" in the sense of complete languages:

> In der Tat liegt die Gemeinsamkeit aller 'Sprachspiele' m.E. darin, dass mit der Erlernung einer Sprache – und d.h. mit der erfolgreichen Sozialisation im Sinne einer mit dem Sprachgebrauch 'verwobenen' Lebensform – zugleich so etwas wie das Sprachspiel – bzw. die menschliche Lebensform – erlernt wird: es wird nämlich prinzipiell die Kompetenz zur Reflexion der eigenen Sprache bzw. Lebensform und zur Kommunikation mit allen anderen Sprachspielen miterworben. (Apel 1973, p. 347)

Apel indicates here that certain formal criteria must be met in order for something to constitute a language in the usual sense, what we may term a "complete language". But Apel gets into trouble when he attempts to formulate the element common to all languages as a separate, "transcendental language game".[204] While the metaphor of a transcendental language game is unfortunate, it seems to me that Apel's basic point is quite sound, namely that mastering a language implies the constant ability to *reflect upon* and *criticise* the moves of any language game – in short, what Brandom refers to as participating in "the normative game of giving and asking for reasons". Thus, in a certain sense, to learn *one* language is to learn "language as such".

The concept of a "complete language" is more or less excluded by Wittgenstein[205] – he prefers the image of a "juxtaposition" of smaller language games, like houses in a town (cf. § 18). But in doing so, he leaves the common syntactic form of "I φ" and "She φ" sentences more or less unexplained. Øfsti addresses this problem when he asks whether the mentioning of moves in a language game belongs to that very game, or whether – as Wittgenstein seems to indicate – it should be understood as a separate, self-contained language game.[206] Øfsti

[204] Cf. Øfsti 1994, p. 801f.
[205] Cf. PU § 18: "[F]rage dich, ob unsere Sprache vollständig ist …"
[206] Cf. Øfsti 1995, p. 803f.; cf. PU § 23.

argues that the descriptions of moves in a game are typically not moves in the game, and in this sense do not belong to the game:

> Eine Partie Schach zu beschreiben ist nicht, Schach zu spielen. (Øfsti 1994c, p. 803f.)

However, we could argue that the possibility of describing the moves is a necessary part of any game in the sense that it is essential to games that we are able to formulate their rules. In general, Øfsti argues, we could formulate the following formal principle for a complete language:

> Alles, was mittels der Sprache getan werden kann, kann auch in ihr beschrieben (gesagt) werden. (Op.cit., p. 804)

Replacing the "saying-showing doctrine" of the *Tractatus*, Øfsti introduces a "doing-saying doctrine," emphasising the linguistic unity of "use" and "mention". In a more general version, the doctrine says that anything that can be *done* – in the sense that it counts as an intentional action – can also be *said* – in the sense that it can be described (and *a*scribed):

> Es kann zuweilen den Anschein haben (zumal, wenn man an die primitiven Sprachspielen in PU denkt), dass es ausreicht, die relevanten Äußerungen sozusagen 'in der *ersten Person Präsens*' zu beherrschen, um in dem Sprachspiel (des Versprechens, Wettens, Befehlens usw.) agieren oder 'ziehen' zu können. Aber der Schein trügt. Um im vollen Sinne als intentional Handelnder gelten zu können, der in einer Situation ***weiß, was er tut***,[207] genügt es *nicht*, eine gleichsam auf die erste Person Präsens beschränkte Akteursprache zu haben, in deren Termini das aktuelle Handlungsbewusstsein geformt werden kann bzw. Sprechhandlungen ausgeführt werden können. Offenbar muss diese Sprache zugleich auch eine 'Metasprache' umfassen, in der diese Handlungen beschrieben werden können. (Loc.cit.)

This is one way of specifying what it means to be a rational agent: Consciousness-in-acting means in principle always being able to put one's action "under a description" (Davidson).

6.6 The performative-propositional double structure of speech

An attempt to grasp the formal requirements of a complete language is delivered by Apel and Habermas with their respective "transcendental" and "universal" pragmatics. In a 1980 article,

[207] Or rather: is *conscious-in-acting*, cf. above, fn. 23. (My italics and footnote)

Apel displays how the pragmatic turn in twentieth century philosophy of language generated new answers to what distinguishes human language.[208] In the beginning of the article, Apel refers to a piece of advice he once got from Karl Popper. The advice was not to attach so much importance to communication – after all we have that in common with animals. *Propositions*, on the other hand, distinguish man alone as capable of rendering truths about the world. Apel lets this statement represent a philosophical paradigm for language which can be traced all the way back to Aristotle. Traditional philosophy of language is proposition-oriented: The possibility of true and false propositions is decisive, while the pragmatic dimension of language is left for the empirical sciences. The pragmatic turn reverses this: *Speech acts,* not propositions, are now seen as the basic unit of language – since propositions can only occur within the frames of speech acts. This challenges yet another basic assumption of the traditional paradigm: The view that the descriptive and the communicative dimension can be examined independently of each other.[209] A correct understanding of the pragmatic dimension is necessary in order to understand the "logos-distinction" of human language, and the pragmatic dimension can be properly understood only in relation to the way language functions in the rendering of facts.

In spite of certain historic attacks on the proposition-oriented paradigm – Apel mentions a.o. Vico as an early challenger[210] – the first real blow against it was dealt by mid-twentieth century's "ordinary language philosophy", represented by the later works of Wittgenstein, J. L. Austin and John Searle. The traditional view did not award the pragmatic dimension of language a constitutive cognitive or "logical" role. The Private Language Argument of Wittgenstein challenges this view by showing that the possibility of a successful *mediation* of intentions is constitutive for linguistic meaning. Austin's concept of *performative expressions* as the linguistic carriers of *illocutionary acts* is also decisive. By uttering such expressions, Austin argued, we do not (only or primarily) state something about a state of affairs, we moreover actually *perform an action*. John Searle's speech act theory picks up on this, adding the *principle of expressibility*: Everything that can be meant, can be said, i.e. every genuine intention of meaning can in principle be expressed in an explicit performative sentence. The principle of expressibility is a decisive argument for a pragmatic philosophy of language, showing that questions concerning the *meaning of a sentence* and

[208] Cf. K.-O. Apel (1980): "Zwei paradigmatische Antworten auf die Frage nach der Logos-Auszeichnung der menschlichen Sprache".
[209] Apel quotes the Aristotle pupil and successor Theophrast (and the German psychologist Karl Bühler) in order to illustrate this view of language, cf. op.cit. p. 16.
[210] Op.cit., p. 17.

questions concerning the *performance of speech acts* do not belong to two different fields of study. Rather, that uttering a sentence in a given context constitutes the performance of a particular speech act, belongs to our very concept of linguistic meaning.

Searle distinguishes between the illocutionary act and its propositional content. He argues that the propositional content cannot occur independently, but only within the frames of questions, assertions, promises, or other speech acts. Grammatically, the propositional content can be isolated in subordinate "that" clauses, whereas the illocutionary act can always be made explicit in a dominating performative sentence ("I wonder…," "I claim…," "I promise…"). This points to a unified structure of all speech acts,[211] which Habermas later terms the *performative-propositional double structure* of speech:

> In der Standardform zeigt sich die charakteristische Doppelstruktur jeder Sprachhandlung. Darin treten zwei Sätze auf: der mit einem performativen Verb in der ersten Person Präsens gebildete Satz und ein abhängiger Satz propositionalen Gehalts. (Habermas 1976a, p. 334)

Habermas's "universal pragmatics" aims at a further development and formalisation of the pragmatic turn, by constructing a comprehensive system of rules for the standard form of speech acts. His point of departure is Searle's principle of expressibility, which implies that any possible speech act can be uniquely determined by one (or several) sentence(s), if the speaker expresses her intention accurately, explicitly and literally.

Following Austin, Habermas assumes that every speech act has a certain illocutionary force determining the communicative function of the utterance by establishing a connection between speaker and hearer:

> Mit dem gelungenen Vollzug eines Sprechaktes wird eine interpersonelle Beziehung zwischen mindestens zwei sprach- und handlungsfähigen Subjekten zugleich hergestellt und dargestellt. (Habermas 1976a, p. 333)

A speech act succeeds when the intended connection is established, i.e. when the hearer understands and accepts the speaker's utterance *as* e.g. a question, a promise, or a claim. This depends on the right context, as well as on whether the speaker is ready to guarantee that he will fulfil certain commitments, e.g. withdraw the claim in case it turns out to be false.[212]

[211] The exceptions are "pure expressions" such as "Ouch!" or "Yuk!" which lack propositional content.
[212] We see here that Habermas's universal pragmatics points in the direction of an inferentialist theory in Brandom's sense, cf. above, Ch. 4.3.

The partition of the speech act into an illocutionary and a propositional component – explicated in a performative sentence in the first person present tense and a dependent proposition – is consistent with the division between two levels of communication:

> a) der *Ebene der Intersubjektivität*, auf der Sprecher und Hörer durch illokutive Akte die Beziehungen herstellen, die ihnen erlauben, sich *miteinander* zu verständigen, und b) der *Ebene der Gegenstände in der Welt*, über die sie sich in der durch (a) festgelegten kommunikativen Funktion verständigen möchten. Ein Sprechakt kann nur gelingen, wenn die Beteiligten die Doppeltstruktur der Rede ausfüllen und ihre Kommunikation *auf beide Ebenen gleichzeitig* führen: sie müssen die Kommunikation eines Inhalts mit der Metakommunikation über den Verwendungssinn des kommunizierten Inhalts *vereinigen*. (Op.cit. p. 334)[213]

Habermas's theory seems to involve the understanding that every speech act has a performative-expressive element. Obviously, though, we constantly use – and understand – sentences such as "A right angle is 90°" and "World War II broke out in 1939", which are third-person sentences with no apparent trace of the act performance or the self-consciousness of the expressing subject. This, according to Habermas, is due to the fact that the performative part of the speech act is not always verbalised:

> Auch wenn die performativen Bestandteile nicht ausdrücklich verbalisiert werden, sind sie im Sprechvorgang stets impliziert: sie müssen daher in der Tiefenstruktur eines jeden Satzes auftreten. (Habermas 1971, p. 104)

The "deep structure" of the above sentence is thus something like "I claim that a right angle is 90°." The often non-verbalised performative part of a speech act tends to be made explicit in *reports* of the act of the type "She claims that World War II broke out in 1939".[214] In this way, the universal pragmatics of Habermas sheds light on the relation between use and mention, what Øfsti refers to as the *vertical indexical system*: That a complete language necessarily involves the possibility of ascending through meta-levels of language.

This is one way of emphasising the pragmatic aspect of meaning. Another way is Tugendhat's analysis of the role of situation-relative indexical expressions in securing a situation-independent meaning of propositions. This corresponds to what Øfsti calls the

[213] The speaker may, however, focus on one or the other of the levels. In *cognitive* use of language the focus is on the content of the utterance as a proposition about something being the case, whereas in *interactive* use of language, the focus is put on the connection established between speaker and hearer through warnings, promises etc.

[214] The "deep structure" of *this* sentence would then e.g. be "I claim that she claims that World War II broke out in 1939°."

horizontal indexical system. With his distinction of two (interrelated) indexical systems of a complete language, Øfsti points out two aspects of the performative-propositional double structure of speech: Through the horizontal indexical system, the possibility of referring objectively to the world in true statements is constituted, whereas the vertical indexical system makes it possible to express one's own subjectivity in a language shared with other subjects. Thus, Øfsti speaks about a "double double structure":

> Die eine Hälfte dieser Doppeltstruktur wäre dann die Searle-Apel-Habermassche, die besagt, dass Propositionen – die nach Tugendhat durch das 'horizontale' deiktische System ermöglicht werden – immer in Sprechakte eingebettet sind (…). Die *zweite* Hälfte wäre nun wieder eine Art performativ-propositionale Doppeltstruktur, aber diesmal nicht im Sinne der *Einbettung* von Propositionen in mit performativen oder expressiven Verben geäußerten Sprechakten, sondern im Sinne der möglichen propositionalen *Abbildung* (Beschreibung) jener Sprechhandlungen – die durch die genannten Verbalphrasen in der ersten Person Präsens *konstituiert* sind. (Øfsti 1997a, p. 164f.)

This structure is a key to a meaningful distinction between subjective, intersubjective and objective, a distinction that can only be made sense of in a language that both (through Searle-Habermas's double structure) separates and (through indexical transitions) connects the performative and the propositional. Wellmer comments on the uniqueness of this system:

> The peculiar relationship between subjectivity and intersubjectivity, between the 'transcendental' and the 'empirical,' which for transcendental idealism remains an insurmountable obstacle, becomes comprehensible once we recognize it in the unique structure of ordinary language. (Wellmer 1976, p. 251)

6.7 The worlds of Habermas

To Habermas, language constitutes a system of "I-demarcation" by which a subject delimits herself, from the outer world, from society, and from her own subjectivity. This is a process of "decentralising" (in the sense of Piaget and Mead) whereby the subject achieves a differentiated world-view:

> [L]anguage can be conceived as the medium of interrelating three worlds; for every successful communicative action there exists a threefold relation between the utterance and (1) 'the external world' as the totality of existing states of affairs, (2) 'our social world' as the totality of all normatively

> regulated interpersonal relations that count as legitimate in a given society, and (3) 'a particular inner world' (of the speaker) as the totality of his intentional experiences. (Habermas 1976c, p. 128f.)

This view of language is a basic element of Habermas's theory of communicative action:

> Allein das kommunikative Handlungsmodell setzt Sprache als ein Medium unverkürtzter Verständigung voraus, wobei sich Sprecher und Hörer aus dem Horizont ihrer vorinterpretierten Lebenswelt gleichzeitig auf etwas in der objektiven, sozialen und subjektiven Welt beziehen, um gemeinsame Situationsdefinitionen auszuhandeln. (Habermas 1981, p. 142)

In his *Theory of Communicative Action* (1981), Habermas explicates the demarcation process of the subject, as part of his continuation and correction of Max Weber's theory of the rationalisation process of society. A differentiated world-view is a distinctive mark of modernity. In the first place, our world-view is distinguishable from the world itself, rendering possible a constant revision of our world-view. The world is further differentiated into the objective world (what *is*) and the social world (what *counts*). These two together constitute the *outer world*, which can be distinguished from the *inner world* of subjective experience. Habermas admits that the concept of different worlds is easily misread as spheres that can exist independently of each other, and stresses the complementarity of the worlds:

> Der Bereich der Subjektivität verhält sich komplementär zu der Außenwelt, die dadurch definiert ist, dass sie mit anderen geteilt wird. Die objektive Welt wird gemeinsam als die Gesamtheit der Tatsachen unterstellt (…) Und eine soziale Welt wird gemeinsam als die Gesamtheit aller interpersonalen Beziehungen unterstellt, die von den Angehörigen als legitim anerkannt werden. Demgegenüber gilt die subjektive Welt als die Gesamtheit der Erlebnisse, zu denen jeweils nur ein Individuum einen privilegierten Zugang hat. (Op.cit., p. 84)

The different worlds are related to different ways of claiming the *validity* of a statement. Habermas here shows himself as a kind of inferentialist in Brandom's sense: He emphasises that using language means entering into a web of commitments and entitlements, and uses the term *validity claims* in order to specify the nature of these commitments:

> [J]eder kommunikativ Handelnde [muss] im Vollzug einer beliebigen Sprechhandlung universale Geltungsansprüche erheben und ihre Einlösbarkeit unterstellen (…). Sofern er überhaupt an einem Verständigungsprozess teilnehmen will, kann er nicht umhin, die folgenden, und zwar genau diese universalen Ansprüche zu erheben:

- sich verständlich *auszudrücken*,
- *etwas* zu verstehen zu geben,
- *sich* dabei verständlich zu machen,
- und sich *miteinander* zu verständigen.
(Habermas 1976b, p. 176)

The speaker makes four validity claims: Trivially, she claims the *intelligibility* of her utterance; furthermore she commits herself to the *truth* of the propositional content; to the *truthfulness* of how she expresses her intentions; and to the *rightness* (or appropriateness) for the norms governing the interpersonal relations that are established through the speech act.

All validity claims are in play in all language use, but one of the claims is typically accentuated, all according to which world the speaker primarily relates to. If this is the objective world, *truth* is at stake: "(It is true that) all swans are white". If it is the social world, prominence is given to the normative *rightness* of the utterance: "(Our relationship is such that I am entitled to direct you to) close the door!" When the speaker relates to her subjective world, her *truthfulness* is the main concern: "(I honestly tell you that) I am sad." Which validity claim the speaker emphasises is decisive for what kind of relation between speaker and hearer the illocutionary act establishes. A negative reply from the hearer will thus respectively mean "No (that is not true)!", "No (you are not entitled to direct me)!" or "No (you are not being sincere)!"[215]

Habermas's theory of communicative action and the differentiation of worlds amounts to a critique of a scientistic absolutation of a value-neutral, causal-nomological explanation. Communicative action is a kind of action that is not *result*-oriented, but oriented towards understanding and agreement [*Verständigung*]. Habermas argues that communicative action is a basic concept needed to explain the rationalisation process of society – which cannot be made sense of from the perspective of instrumental and strategic action alone. His approach has, in his own words,

> the aim of clarifying the presuppositions of the rationality of processes of reaching understanding, which may be presumed to be universal because they are unavoidable. (...) The most important achievement of such an approach is the possibility of clarifying a concept of communicative rationality that escapes the snares of Western logocentrism. Instead of following Nietzsche's path of a totalizing and self-referential critique of reason, whether it be via Heidegger to Derrida, or via Bataille to Foucault, and throwing the baby out with the bathwater, it is more promising to seek this end through

[215] We may thereby distinguish between *assertoric*, *regulative* and *representative* use of language, cf. e.g. Habermas 1976a, p. 334ff.

> the analysis of the already operative potential for rationality contained in the everyday practices of communication. Here the validity dimensions of propositional truth, normative rightness, and subjective truthfulness or authenticity are intermeshed with each other. From this network of a bodily and interactively shaped, historically situated reason, our philosophical tradition selected out only the single thread of propositional truth and theoretical reason and stylized it into the monopoly of humanity. The common ground that unites both von Humboldt and pragmatism with the later Wittgenstein and Austin is the opposition to the ontological privileging of the world of beings, the epistemological privileging of contact with objects or the existing state of affairs, and the semantic privileging of assertoric sentences and propositional truth. Logocentrism means neglecting the complexity of reason effectively operating in the life-world, and restricting reason to its cognitive-instrumental dimension (a dimension, we might add, that has been noticeably privileged and selectively utilized in processes of capitalist modernization). (Habermas 1985, p. l96f.)

Like Apel, Habermas seeks to point out a dimension "between" objectivated nature and the subjective and intersubjective conditions for the objectivation of nature. The aim of Apel's criticism of "the alternative conception" is to show the possibility – and necessity – of a hermeneutic understanding of other persons, which cannot be reduced, neither to direct "I-You"-communication nor to the description of others in scientific explanation. A genuine hermeneutic understanding means recognising others as "virtual" communication partners or "possible" co-subjects. Describing intentional actions does not involve an objectivation in the same way that a description of natural phenomena does. Apel speaks about "secondary objectivation", as opposed to the primary objectivation of nature.[216] Habermas's differentiation of worlds can be regarded as an attempt to make room for such a secondary objectivation and hermeneutic understanding of virtual communication partners. However, as I will attempt to show, his solution is in a certain sense problematic.

In *A Theory of Communicative Action*, Habermas presents a table in order to display the different possible "formal-pragmatic relations" between the different worlds and the basic attitudes a speaker can take towards the respective worlds:[217]

[216] Cf. Apel 1979, p. 173; cf. above, Ch. 6.2.
[217] Habermas 1981, p. 324 (my italics in table).

Worlds **Basic attitudes**	**1. objective**	**2. social**	**3. subjective**
1. objectivating	cognitive-instrumental relation	***cognitive-strategic relation***	***objectivistic self-relation***
2. performative	Moral-aesthetic relation To non-objectivated Surroundings	obligational relation	judging self-relation
3. expressive		self-engineering	perceptive-spontaneous self-relation

The combination of the speaker's attitudes and the worlds towards which she directs them decides what sort of rational relation is established through the statement. This suggests a differentiated view of rationality; however, some of these relations have a rather dubious status.

Habermas assumes that it is possible to identify different worlds independently of the speaker's attitudes. When we speak about the objective world, it seems intuitively unproblematic to assume that the world is somehow "there", independent of speakers' attitudes. However, this is not so easy when we are talking about the subjective and the social worlds, since they in a certain sense are established *through* our performative and expressive attitudes. The table shows that Habermas thinks it possible to relate in an *objectivating* manner to the *social* world (column 2.1, "cognitive-strategic relation"), as well as to the *subjective* world (column 3.1, "objectivistic self-relation").[218] As I will attempt to show in the following section (6.8), Habermas's definitions of the different worlds are ambiguous and unfit to attain the diversification Habermas seeks.[219]

[218] Although he does indicate the limits of – and warns against – this kind of objectivation of the subjective and the social, cf. e.g. Habermas 1981, p. 137.
[219] Cf. also Granum (Carson) 2000, p. 244ff.

6.8 One world

In the second paragraph of the *Tractatus*, Wittgenstein writes:

> The world is the totality of facts, not of things. (TLP 1.1)

He thereby chooses sides in a modern version of the medieval debate between realists and nominalists. The basic, conceptual decision to conceive of the world as having *objects* or *facts* as its component parts has considerable philosophical consequences, not least for the conception of truth.[220] Habermas, however, refuses to commit himself to either side in this matter. Instead, he argues that the two views are complementary:

> [D]ie objektive Welt [muss] als Korrelat zur Gesamtheit wahrer Aussagen verstanden werden; nur dieser Begriff behält die im strengen Sinne ontologische Bedeutung einer Gesamtheit von Entitäten. (Habermas 1981, p. 125f.)

Clearly, both ways of defining the world have advantages. Defining the world as a "totality of objects", or as nature in Kant's sense,[221] has the immediate advantage that it represents a clear-cut alternative to relativism or contextualism. The problem is how this concept of an objective world can encompass cultural, societal, or political entities, such as acts or statements. Obviously, such entities can be made objects of (true) statements, descriptions, and explanations. But the "totality of objects" definition seems to imply that the only way to objectivate subjective and social entities is by *scientistically* reducing them to objects among others within a "Kantian nature". In that case, however, Habermas is unsuccessful in establishing a hermeneutic dimension as an alternative to scientism.

Understanding the objective world as a correlate to the totality of true statements is equivalent to the "totality of facts" definition. This one seems more promising when it comes to encompassing cultural, societal, and political entities. But the definition implies that as soon as I make a (true) statement, I am already referring to the objective world. That would, however, overturn Habermas's attempt to establish a differentiated world-view through the matching of worlds and attitudes in the table above. According to the correlation definition,

[220] Cf. e.g. the debate between Habermas and Brandom: Habermas 1999 p. 138ff. and Brandom 2000b.
[221] Cf. Habermas 1981, p. 83: "…die äußere Natur oder die objektive Welt…"

an objectivating attitude (i.e. the trying to say something true about something) per definition means relating to the objective world of describable facts. In this case an objectivating attitude to the social and subjective world is impossible: I can only relate to the social and subjective world in the first-person present tense, in performative or expressive (parts of) speech acts.[222]

In other words, each definition of the objective world seems to invalidate the diversification Habermas was striving for with his table (cf. above, 6.7). The result is in either case "the alternative conception" all over again: no mediation between a performative-expressive attitude in the first-person present tense and an objectivating attitude seems possible. The social and subjective world thus becomes, in the words of *Tractatus*, something that cannot be "said," in the sense of being *depicted* in statements, but can only "show itself" in the *actualisation* of acts. Of course, as opposed to the early Wittgenstein, Habermas certainly does not leave the expressive and performing subject "speechless"; only "third person-less". However, the correct account should explain the possibility of transitions within the *same* language between objectivated and non-objectivated, propositional and performative. The question is whether such an account can be achieved by explicating how the two definitions of the objective world relate to each other.

In *Wahrheit und Rechtfertigung* (1999), Habermas attempts to combine his universal pragmatics with a pragmatic theory of action and learning. In this context, he further specifies the need to view the world *both* as a "totality of things" and as a "totality of facts". The background is an attempt to amend certain weaknesses connected to his earlier proposed theory of truth. His 1973 article "Wahrheitstheorien" criticises both coherence and correspondence theories of truth, leading him to introduce the so-called *consensus* theory, later developed into the *discursive* theory of truth. In *Wahrheit und Rechtfertigung* he calls himself to account, and argues that the discursive theory creates the false impression that a sentence is true if the involved parties are able to reach an agreement about it. What must be

[222] The first-person present tense is of course not *enough* in order to establish a relation to the social or subjective worlds. Even if verbs for emotions, attitudes etc. are used (in contradistinction to sentences such as "I am 180 cm tall", "I live in Trondheim", etc.), we must further evaluate the speech situation and the mode of the speech act to find out to what world the speaker relates. If I refer to my condition in a propositional manner – "I hope he will come" as a true report about my state of mind – this may be seen as a reference to the objective world. Cf. Wittgenstein, PU § 585: "Wenn Einer sagt 'Ich hoffe, er wird kommen' – ist das ein Bericht über seinen Seelenzustand, oder eine Äußerung seiner Hoffnung? – Ich kann es z.B. zu mir selbst sagen. Und mir mache ich doch keinen Bericht. Es kann ein Seufzer sein; aber muss kein Seufzer sein. Sage ich jemandem 'Ich kann heute meine Gedanken nicht bei der Arbeit halten; ich denke immer an sein Kommen' – so wird man das eine Beschreibung meines Seelenzustands nennen". In this sense, as soon as I *assert* something, I am already referring to the objective world. Correspondingly, "(It is the case that) q is forbidden" – rendered as a social *fact* – is an assertion about the objective world.

emphasised is that we should come to agreement about a sentence because it is true – not the other way around. According to Habermas, the discursive truth theory suffers from

> einer Überverallgemeinerung des speziellen Falls der Geltung moralischer Urteile und Normen. (Habermas 1999, p. 16)

Habermas here implies that the validity of moral norms rests on the ideal assumption that we can reach an agreement about them in a discourse where all parties have been heard – and this assumption cannot simply be applied to the case of scientific knowledge of nature. Although all parties have spoken and agreement is reached, we can nevertheless not be sure that the theory about the Sun's orbit around the Earth agrees with reality. In order to take care of our realistic intuitions in epistemology (the objective world exists independently of our ability to know it), without having to pay the price of moral realism (moral norms exist independently of our ability to acknowledge them), Habermas now distinguishes more clearly between truth and rightness, and emphasises that a sentence being true is not the same as it being rationally grounded.

Habermas now accentuates *fallibilism*: Epistemological realism implies that the objects we refer to may not satisfy our descriptions. Even given close to ideal communicative conditions, the truth may be something other than what we – with the best of reasons – believe it to be. Habermas does not, however, ditch the idea of nominalism all together. He sticks with his previous view of realism and nominalism as complementary, suggesting a "division of labour" between them:

> In den Grundbegrifflichkeiten von Realismus und Nominalismus spiegelt sich der methodische Unterschied zwischen dem hermeneutischen Zugang des Teilnehmers zur intersubjektiv geteilten Lebenswelt einerseits und der objektivierenden Einstellung des Hypothesen prüfenden Beobachters in der Interaktion mit dem, was ihm in der Welt begegnet, anderseits. Der grammatische Begriffsrealismus ist auf eine Lebenswelt zugeschnitten, an deren Praktiken wir teilnehmen und aus deren Horizont wir nicht heraustreten können. Demgegenüber trägt die nominalistische Begriffsfassung der objektiven Welt der Einsicht Rechnung, dass wir die Struktur der Aussage, mit der wir etwas in der Welt beschreiben, nicht zur Struktur des Seienden selbst reifizieren dürfen. Zugleich erklärt die Konzeptualisierung der Welt als 'Gesamtheit der Dinge, nicht der Tatsachen', wie die Sprache mit der Welt in Kontakt tritt. Der Begriff der 'Referenz' muss erklären, wie der ontologische Vorrang einer nominalistisch begriffenen objektiven Welt mit dem epistemischen Vorrang der sprachlich artikulierten Lebenswelt in Einklang zu bringen ist. (Op.cit., p. 44)

Habermas gives *ontological priority* to a nominalistically explicated objective world, and *epistemological priority* to a linguistically articulated lifeworld. Through this distinction, Habermas wants to have his cake and eat it too, in the sense of tending to both his realist and his nominalist intuitions. His point is that a concept of objective reality must be secured after the linguistic turn, while retaining the thesis that intersubjective language is our access to the world in the sense of "veritative being". The concept of an objective world is thus put forward as defence against the contextualism of Rorty and others. However, the concept of an intersubjectively shared life world is the basic concept of world in the sense that it is the fundament of the objective world of things, the social world of norms, and an inner world of subjective experience.

In Chapter 3.5, I suggested a distinction between epistemological and ontological priority as a way of explicating the complementarity of intentional agency and causality. However, I would not follow Habermas in using this as a method of distinguishing worlds. The reason is that I cannot see how Habermas can resolve the ambiguity of his concept of the objective world in order to make it strong enough to carry the weight of his arguments. In a review of *Wahrheit und Rechtfertigung*, Albrecht Wellmer calls attention to similar problems concerning Habermas's concepts of world and reality, and especially to the ambiguous concept of the objective world:

> *Entweder* […] bezeichnet die objektive Welt den Bereich des wissenschaftlich Objektivierbaren – dann wird er unbrauchbar, um den ganzen Bereich der wahrheits- und diskursfähigen Aussagen abzustecken; *oder* dieser umfasst auch noch die geschichtlich-kulturelle Wirklichkeit – dann ist das Wort 'objektiv' nicht viel mehr als eine bloße Wortschleuder gegen den Kontextualismus. (Wellmer 2007, p. 211f.)

The concept of an objective world brings Habermas "one step too far back to Kant", according to Wellmer.[223]

The sound point behind Habermas's differentiation of worlds is his attempt to show that the relations between subject and world cannot be just a matter of the subject's shifting attitudes – in that case we would have no safeguard against contextualism and pure relativism. There is "something" *being* objectivated, or *resisting* objectivation, viz. only allowing for "secondary objectivation" (Apel), or being a *condition* for the possibility of objectivation – diverse "somethings" that Habermas classifies in different concepts of world. However, the classification does not turn out in a good way. The concept of an objective world is in my

[223] Cf. Wellmer 2007, p. 212.

view systematically ambiguous because I cannot see how it can be defined independently of the concept of truth. This leaves Habermas with two equally unsatisfying alternatives: *Either* truth value is reserved for the area of natural science, in which case we cannot ascribe truth value to e.g. "historic" sentences ("A said/thought/promised that p"); *or* the objective world includes e.g. historical and cultural facts, in which case it cannot carry the burden that Habermas lays on it as a safeguard against contextualism.

If the multiple world concepts are unfit to make the distinctions Habermas is aiming for, the concepts seem to be redundant. If social and subjective facts enter into the objective world, we have no longer any use for the concepts of social and subjective *worlds*. The first-person present tense is all that remains, and this should be adequately covered by performative and expressive attitudes. Without the concepts of social and subjective world, the contrasting concept of objective world loses its raison d'être as well. Thus we are left with a single concept: The concept of world.[224]

6.9 Veritative and performative being

Habermas seeks to develop a differentiated view of the world as containing nature (in a Kantian sense) as well as society and subjective experience. In my view, this differentiation must be explicated grammatically, in order to avoid the ontological ambiguities of the "multiple world" allegory. Habermas's own theory about the performative-propositional double structure of speech points us in the right direction. By indicating a universal structure for all intentional acts, Habermas sheds light on the relation between the performative-expressive aspect of language and the propositional aspect. Both aspects can be further elucidated by differentiating the indexical systems of ordinary language:

1) The propositional aspect can be explored by looking at how the *horizontal indexical system* of language renders possible the maintenance of the same propositional content through different speaker's positions, thus opening up the world to us in the sense of "veritative being" (cf. Tugendhat).

[224] Which would maybe be equivalent to Habermas's undifferentiated (or *pre*differentiated) concept of lifeworld. However, whereas the concept of world is methologically "neutral", the concept of lifeworld emphasises the primacy of our "living" experience with the world (the world as "erlebt") over the world-view resulting from objectivation or theoretical analysis.

2) The performative aspect can be explored by examining the *vertical indexical system*, the transitions between use and mention in language, opening up the world to us in the sense of "performative being" – i.e. being in the sense of being conscious-in-acting.

Instead of differentiating worlds, we may differentiate between primary and secondary objectivation, as Apel does, or between different versions of the third-person mode, as Øfsti does. A sentence in the "absolute third person" deals with entities that are not language users, i.e. that can only be referred to in the third person, not be *addressees* (in the second person) or *speakers* (in the first person)[225], while a sentence in the "relative third person" is about persons and actions. In line with Apel's "secondary objectivation", Øfsti's concept of "relative third person" is employed in order to point out the hermeneutic dimension "between" objectified nature and I-you-communication, a dimension that is disregarded within "the scientistic alternative conception":

> The failure to assess correctly this category of the relative third person is, I believe, closely related to what K.-O. Apel deplores as the 'scientistic fallacy' of assuming that the transition from speech in the first and second person to speech in the third person mode must necessarily be a transition to the standpoint of pure observation and theoretical explanation in the Hempelian sense. (Øfsti 1985, p. 47, note 16)

In the Hempelian approach, the relative third person is, one could say, assigned the dubious function of a "turntable" in the scientistic objectivation of persons and actions:

> Die 'dritte Person' kann sozusagen innerhalb von zwei 'Sprachspielen' (…) gehandhabt werden, als 'relativ' dritte Person im 'Person-Handlung-Verantwortung'-Spiel, und als 'absolut' dritte Person im 'naturalistischen' Sprachspiel. Die alltagssprachlichen Formulierungen in der dritten Person (über Personen und Handlungen) funktionieren beim szientistischen Fehlschluss als eine Art Drehscheibe. Personen und ihre Handlungen werden durch diese Form in das naturalistische Sprachspiel eingeschleust und den dort üblichen Bedingungen der Begriffs- und Theoriebildung unterworfen. Was übersehen wird ist, dass in ihrer normalen Funktion im Person-Handlung-Verantwortung-Spiel (das essentiell umgangssprachlich ist) der Sinn der Handlungsverben essentiell von ihrer handlungs*konstitutiven* Verwendung in der ersten Person Präsens, überhaupt von ihrer Rolle in der unmittelbaren Kommunikation und Interaktion, *lebt.* (Øfsti 1994a, p. 182)

[225] Cf. Øfsti 1985, p. 26f.

The possibility of describing persons and actions in the third person may lead to ignoring the decisive distinction between the objectivated "absolute third person" and the secondary objectivated "relative third person". The relative 3p turns unnoticeably into absolute 3p, reaching towards a scientistic ideal (à la Hempel) of replacing all "colloquial", "prescientific" phrases with well-defined, precise, experimentally testable, "scientific" predicates. The problem is that the new, "scientific" phrases cannot be conjugated *back* to the first-person present tense without absurdities: The performative consciousness-in-acting is lost on the way to a scientific language. If we remain within colloquial language, the transition from the consciousness-in-acting expressed in I-you-communication to a report in the third person is not a transition to a theoretical perspective. The verbs that constitute or express consciousness-in-acting in the first person are adequate in the third person as well in order to achieve the desired understanding of (and virtually *with*) the agent mentioned in the sentence. This is because – as Wittgenstein demonstrates in PU – the meaning of expressive-performative predicates necessarily has two roots: the act-/consciousness-constitutive use ("I promise you that…") and the descriptive use ("I promised/she promises that…").

Habermas's theory of communicative action is a good point of departure from which to look at the relationship between agency, language, and world. Habermas sheds light on the connection between linguistic meaning and action by connecting the sense of utterances to our interest in reaching understanding for the sake of orienting action.[226] His approach has the advantage of showing how the idea of a world existing independently of us actually evolves through our practical dealings with the world, where we encounter objects offering resistance to our actions. This may be viewed as a generalisation of the idea behind an interventionist theory of causality. Making explicit the practical dimension that underlies our ability to have representational knowledge forms the rational basis of the double structure of speech, where true or false propositions are placed within the frame of intentional agency. Habermas's universal pragmatics and the theory of communicative action carry on the pragmatic-linguistic turn of Heidegger, Wittgenstein, Austin, and others. His concepts of validity claims, classes of speech acts, speakers' attitudes and so forth contribute to a systematic theory of language based on a view of reason as communicative competence. Further, by combining this normative pragmatics with a pragmatic theory of action and learning, he generates strong arguments against relativism and contextualism. However, in my view the differentiation of

[226] An aspect that may be regarded as underplayed in Brandom's theory, cf. above, Ch. 5.4.

worlds contains the remainders of an "alternative conception", however contrary this is to what Habermas aims at.

The problem with Habermas's concept of the objective world is that it cannot *both* function as a defence against Rortian contextualism while *at the same time* encompass the historic and cultural reality. The problem with his concept of a social world is that it remains ambiguous whether objectivated action still belongs to this world, or whether action at the moment of objectivation is "sluiced" out of it and into the objective world. The same problem arises in the case of the subjective world: If subjective experience is objectivated, i.e. subjected to the objectivating attitude, is it not thereby already transferred to the objective world?

A further problem with the concept of an inner world of subjective experience is that it inevitably suggests a "something" that can be (truthfully or not) expressed, and to which I myself have a "privileged access".[227] Undoubtedly it is important to point out the special case of the first person present tense as "the limit of experience", but when doing this we must be careful not to postulate a "something" to which only I have access, in which case we end up with a paralogism in Kant's sense (cf. KrV A 348ff./B 376ff.) or a Wittgensteinian "beetle in a box" (cf. PU § 293). Characterising the subjective as "inner" (vs. what is "outer") conceals the fact that the subjective essentially is *expressible*, i.e. has outer criteria (cf. PU § 580).[228]

A general problem with Habermas's worlds is that the concepts suggest a certain constancy regarding what kinds of entities belong within the different categories. However, as the systems of indexicality display, what is from one perspective an expression of the subjective experience of the speaker, is from another perspective the object of assessment or description. The concept of worlds lacks the necessary flexibility to address such transitions. A grammatical division seems in that sense more promising than the ontologically charged concept of world. However, we should maintain Habermas's anti-contextualist point: Objectivity, truth, validity, etc. are not *only* a question of the subject's attitudes – the world exists independently of us, and offers resistance to our attempts. This can be complemented

[227] Cf. e.g. Habermas 1981 p. 84. Against this formulation however, consider p. 137, where Habermas stresses that subjective experience must not be perceived as mental episodes or inner episodes; "damit würden wir sie an Entitäten, an Bestandteile der objektiven Welt angleichen". The ambiguities of Habermas's position is, I suggest, due to the difficult balance of his position, as well as to an inherent problem in the use of the inner-outer metaphors, cf. next fn.

[228] Of course, Habermas's intention is to avoid such implication, but as I have argued above, I find the use of the inner-outer metaphor to be a continuous source of misconceptions in the philosophical debate on free will/agency, cf. above, Ch. 1.3. Cf. also my critique of Sebastian Rödl, Ch. 4.4.

with an anti-naturalist[229] point: Not everything that exists can be objectivated in the scientistic sense that we can turn it into an object for natural science. On the contrary, "something" must be considered as epistemologically prior to the objectivated world, and thus as a condition for us having knowledge at all. This is a Kantian motive that is constantly reinvented in philosophy, and that Wellmer – commenting on Adorno's theory of the sublime – gives a striking version of:

> In gewissem Sinne (…) hatte Adorno recht, wenn er für das Absolute, wenn er für das intelligible Ich einen Ort zwischen Sein und nicht-Sein suchte. Die Subjekt-Objekt-Dialektik aber ließ hier als ein Drittes nur die Idee eines künftiges Seins zu. Schon Kant aber hatte, und zwar vor jeder kritischen Metaphysik, jenen Ort zwischen Sein und Nicht-Sein überzeugender als den eines praktischen Seins bestimmt: Es 'gibt' Freiheit in der Welt, sofern wir nur unter der Idee der Freiheit handeln können. Dies Sein der Freiheit bezeichnet keinen Zustand der Versöhnung, es bezeichnet vielmehr einen Seinsmodus der Welt sprachlich erschlossenen Sinns, durch welchen diese objektivierender Erkenntnis im strikten Sinne unzugänglich bleiben muss, ihr als ein Nicht-Sein erscheinen muss.[230] Bei Kant bleibt dieser fruchtbare Gedanke freilich noch eingehüllt in ein Gewebe bewusstseinsphilosophischer Voraussetzungen; erst die neuere Philosophie – ich denke vor allem an Heidegger, Wittgenstein und die amerikanischen Pragmatisten – hat die Voraussetzungen dafür geschaffen, den Kantischen Gedanken sprachphilosophisch zu reformulieren und hierdurch zu verallgemeinern. Der Gedanke besagt dann – so findet er sich in besonders klarer Form bei Habermas –, dass das Sein der sprachlichen Sinns, der Freiheit, der Wahrheit, der Vernunft ein performatives Sein ist, ein Sein, das sich erst in der performativen Einstellung sprachlich kommunizierender Subjekte konstituiert und nur in ihr sich erhält. (Wellmer 1991, p. 183f.)

In Wellmer's wording, "being" has more than one sense. Wellmer points to Habermas's universal pragmatics and the double structure of speech as a way of explicating this point. In this spirit, I suggest distinguishing between:

1) *Veritative being*: Being in the sense that something (i.e. propositions) can be true or false. This mode of being is opened up to us by help of the horizontal indexical system, enabling us to maintain a propositional content throughout different speech

[229] I.e. naturalism in the sense discussed above, Ch. 5.2; not in the sense of Habermas's own concept of "weak naturalism".
[230] [fn SGC] I understand Wellmer here as pointing to objectivating knowledge in the sense of *scientistic*objectivation (by which the link to the 1pp-case is cut), not the secondary objectivation in Apel's sense, by which we are able to *hermeneutically* describe actions. Otherwise his formulation would be leaning towards an "alternative conception" in the sense that free agency can only be recognised in the first-person present tense, and must emerge as a "non-being" from the perspective of an observer.

situations, thus gaining access to a context-transcending truth and a "logical space of reasons".

2) *Performative being*: Being in the sense of consciousness-in-acting. This mode of being is opened up to us by help of the vertical indexical system, making the transitions between use and mention possible, thus distinguishing *and* connecting "the limit of the world" (the self-consciousness of the subject) and the world (what can be described/said).

In accordance with the performative-propositional double structure of speech, performative being is the superior mode in the sense that the veritative being that is expressed in propositions can occur only within the limits of performances, i.e. speech acts. If we look at the different grammatical modes of tenses and persons, we see that these categories have the necessary flexibility missing in Habermas's account of different worlds. First-person present tense *typically* relates to the world in sense 2.[231] "Absolute third person" typically refers to the world in sense 1.[232] "Relative third person" essentially involves relating to the world in *both* senses: A sentence in the relative third person proposes to say something *true*, e.g. about a performed action, but at the same time, in dealing with the irreducibly normative, it relates to the performative being expressed in the corresponding first-person present form. In short: Veritative being is *asserted* in propositions, while performative being is *expressed* in the performative parts of speech acts, or (normatively) *ascribed* in propositions in the relative third person. These are not mutually exclusive, but *overlapping* modes of being. This account emphasises that free agency does not belong in a separate world, realm, kingdom, or language game. Free actions *exist*, and take place in the very same world as everything else.

This distinction is based on the performative-propositional double structure, and aims to shows the *complementarity* between the performative and the propositional dimension of speech and reason, as well as between consciousness-in-acting and consciousness of objects. Øfsti gives an account of this positive connection in *Abwandlungen*:

> Alle meine Handlungen bedürfen einer Art des Selbstbewusstseins, das eben nicht die Form eines Bewusstseins von mir als Objekt hat (…), sondern eher mit dem 'Ich denke' der transzendentalen

[231] It is, of course, possible to *describe* oneself in 1pp, thus relating to oneself in the sense of "veritative being", e.g. in sentences such as "(It is true that) I have blue eyes", or even "(It is true that) I am in pain" when this sentence is used to inform my doctor, cf. above. The mode of an utterance is, in other words, not given by grammatical criteria alone.

[232] Exceptions to this could be what Habermas calls "moral-aesthetic relation to non-objectivated surroundings", cf. his table above, Ch. 6.7.

> Apperzeption zu vergleichen ist. (…) Das Handlungsbewusstsein des Akteurs ist (…) konstitutiver Bestandteil der Handlung selbst und insofern Bedingung der Möglichkeit ihrer späteren (oder aktuellen) Beschreibung/Zuscreibung zum Akteur als eine bestimmte intentionale Handlung. (Øfsti 1994a, p. 135)

But at the same time:

> Es kann das performative Handlungswissen in der ersten Person Präsens des Aktes *auch* nicht geben ohne die *mögliche* Abwandlung in die anderen Formen (Präteritum, dritte und zweite Person), die ja *tatsächlich* Beschreibungen der Handlung (bzw. ihre Zuschreibung zu einer Substanz/Person) sind. Dieses Moment hat z.B. Kant vernachlässigt, als er sich nicht klar machte, dass auch das 'Ich denke' des aktualen Erkenntnisaktes bodenlos wird, wenn die Verbindung mit einer möglichen späteren Selbstbeschreibung oder Selbstzuschreibung ('Ich habe gedacht', 'Ich habe behauptet' usw.) beziehungsweise mit der möglichen Einnahme der Perspektive der anderen (…) fehlt. (Loc.cit.)

This complementarity cannot be recognised within a mentalistic philosophy of mind, but is explicable within a pragmatic philosophy of language. This, however, demands that we manage to avoid getting trapped in an "alternative conception". In the next and final chapter of my dissertation, I look at how this decisive point is taken care of within the philosophical debate on free will/agency, specifically with regards to the debate between compatibilists and libertarians.

Chapter 7: Freedom and first-person priority

[D]as Bewusstsein meines eigenen Daseins ist zugleich ein
unmittelbares Bewusstsein des Daseins anderer Dinge außer mir.
(Kant, KrV B276)

Beobachten ist selbst ein Fall der Teilnahme.
(Seel 2005, p. 145)

7.1. Introduction: The limits to self-objectivation

Methodological naturalism rests on the conviction that all knowledge – including the knowledge we have of ourselves – is revisable in light of new scientific progress. In one sense this is nothing more than sound, epistemological prudence. Any attempt to say otherwise, i.e by pointing out epistemological conditions for the process of gaining knowledge itself – must take care not to recede into infallabilistic apriorism. Habermas warns against this:

> Die erkenntnistheoretische Wendung darf nicht den starken transzendentalen Sinn haben, die intersubjektiven Bedingungen des Zugangs zur objektiven Welt gegen weiter empirisch informierte Nachforschung zu immunisieren. (Habermas 2006, p. 695)

At the same time, any attempt to objectivate self-consciousness threatens to cross a "performative" limit in the sense of draining our everyday and scientific practices of their meaning. Thomas Nagel talks about an opposition, not between two views about how matters stand in the world (e.g. one according to which it is up to me what I do and one according to which my actions are the results of strict laws of nature), but

> between a theory about how things are and a practice that would be impossible if this was how things were. (Nagel 1997, p. 116f.)

Suppose I am taught, and become convinced, that all my actions happen as the result of psychological mechanisms, neuro-processes, or external manipulation. I will have no rational way of reacting to such information. I would have to reckon my belief in it among the results of the causal forces influencing me, thus it cannot count as a reason on the basis of which I could make a decision or draw a conclusion. Furthermore, even to think *this* would be a rational argument that I would no longer be in a position to draw a conclusion from:

> Doubt about your own rationality is unstable; it leaves you really with nothing to think. (Nagel 1997, p. 116)[233]

Thus, it is *practically* impossible for me not to conceive of myself as a rationally thinking being.

According to Habermas, the continuing expansion of the area of knowledge about ourselves – e.g. through Darwin's theory of evolution or even Freud's theory of the unconscious – may be fruitful and may function as liberating right up to the point where *I* attempt to view *myself* as subjected to strict laws of nature. Paradoxically, once the limit of self-objectivation is crossed, there is no longer anyone whose self-knowledge can expand, or who may be liberated. The subject of ever-expanding knowledge and continuous self-liberation dissolves:

> Die Grenze naturalistischer Selbstobjektivierung wird mit Beschreibungen überschritten, unter denen sich Personen nicht mehr als Personen wiedererkennen können. (Habermas 2006, p. 681)

The practical impossibility of going beyond the view of myself as a freely acting subject is what Kant refers to as a "Fact of Reason".[234] The fact of reason is not a theoretically thought-out concept, but something that reveals itself in my capacity to make decisions in the sense of acting according to reasons, thereby answering the question "What should I do?" To the extent that scientific discovery is a human activity, it seems that the "fact of reason" must be primary to the principles of science:

---

[233] In *The Last Word* (1997), Nagel carries on his critique of "the external standpoint" from *The View from Nowhere* (1986). In general I agree with the critical side of his endeavour, but will avoid at this point going into a discussion of Nagel's own approach to the problem of self-objectivation, which is in my view problematic (problematic, as Habermas comments as well, because of his mentalist understanding of the distinction between first- and third-person perspectives, cf. below, fn 25 and above, Ch. 1.3).

[234] Cf. KpV, e.g. 9; 56; 72.

> The Scientific World View is a description of the world which serves the purposes of explanation and prediction. When its concepts are applied correctly it tells us that things are true. But it is not a substitute for human life. And nothing in human life is more real than the fact that we must make our decisions and choices 'under the idea of freedom'. (Korsgaard 1996, p. 97)

The performer of scientific activity cannot – any more than any other agent – objectivate her own perspective without at the same time suspending the claims she is making. Thus, the carrying out of a self-objectivating scheme paradoxically dissolves itself. This is because the attitudes we take up against states of affairs or other persons are not *experiences* that the subject can have or not have, but *acts* that the subject *performs* or *carries out*. In other words they have a *normative character*; they constitute rule-governed behaviour and may as such fail or succeed.[235] This normative character is inaccessible from the perspective of pure scientific objectivation:

> Unter einer objektivierenden Beschreibung, die aus begrifflichen Gründen die Differenz zwischen dem Gelingen- und Misslingenkönnen einer Operation zu Gunsten eines faktischen Geschehens – das ist, wie es ist – einziehen muss, kann sich eine Person nicht als solche wiedererkennen. (Habermas 2006, p. 683)

In ways similar to this, many have argued the practical impossibility – or pragmatic inconsistency – of arguing against a human capacity for free agency. Given an interventionist theory of the interdependence between causality and agency, it seems moreover possible to argue that it is not only *practically*, but also *epistemologically* speaking impossible to get beyond the presumption of free agency. If causal statements, e.g. about the functions of the brain, conceptually point back to an idea of intentional agency, then this sets a limit as to what can be objectivated. In other words, it is not only impossible to *act* except under the idea of freedom, but furthermore also to *think* if not for a presupposition of free agency.

In Chapter 3.5 above, I argued that the concept of intentional action may be seen as *epistemologically primary* and the concept of causality as *ontologically primary*. From this it is understandable that the naturalistic branch of philosophy focuses on ontological questions, as Habermas comments:

> Für den Naturalisten empfiehlt es sich, die epistemische, auf das Bewusstsein von Akteuren bezogene Betrachtungsweise aufzugeben, um stattdessen ontologisch zu untersuchen, welchen Platz geistige

[235] Cf. above, Ch. 4.

> Zustände in einer materialistisch begriffenen und kausal geschlossenen Welt einnehmen. (Habermas 2006, p. 688)

Kantian and/or pragmatically minded philosophers, on the other hand, argue that since there is no such thing as a "pure" ontology, since there can be no "view from nowhere" from which such a "pure" insight into the world is attainable, even our most objective knowledge of the world is rendered possible by, thus subjected to, certain epistemological conditions of possibility. One such condition, already suggested by Kant, and developed further by among others Peirce and von Wright – is our ability to actively intervene with the ways of the world.

A one-sided focus on the epistemological conditions of our access to the world, however, threatens to end in relativism or contextualism, reducing all truth to "true for me" or "true for us". Hence, we must rush to add that the *ontologically primary* idea of a world existing independently of – and offering resistance towards – the acting and knowing subject (or community) is an equally decisive condition for the possibility of knowledge as the *epistemologically primary* conception of ourselves as agents actively engaged with the world.

Further in this chapter I will look at the debate between compatibilistic and libertarian defences of human freedom, regarding the question of whether action can meaningfully be called free independently of how matters stand in nature, i.e. whether natural events are the result of deterministic laws of nature or not (7.2). In 7.3 I critically analyse compatibilism by setting this view in connection with scientism and the above analysed "alternative conception".[236] I discuss whether a compatibilistic position is defensible given a rejection of the scientific fallacy (which is Habermas' position), or whether a rejection of the scientific fallacy equals a defeat of compatibilism. In the final section, 7.4, I give a few concluding remarks, both to this chapter and to the thesis as such.

7.2 Doing otherwise: The debate between compatibilists and libertarians

Throughout this thesis I have with a few exceptions stayed away from the traditional philosophical debate on free will. Instead I have directed my efforts towards establishing arguments against what I consider to be an underlying and misleading assumption in great parts of the debate. This assumption is that our "observational" view of the world – a world of

[236] Cf. above, Ch. 6.2.

scientifically predictable events – is a primary and relatively unproblematic view, into which we must try to fit a conception of free agency. Against this I have argued that our primary access to the world is formed by our condition of being free, in the sense of being rational, communicating agents.

At this point I turn briefly towards the debate on free will in order to look at how my approach fits into some central issues there. Specifically, I consider some versions of compatibilistic or libertarian positions that I consider to be closest to my own position. Compatibilists and libertarians share an "optimistic" view on the possibility of freedom,[237] i.e. both view free agency as possible. But while compatibilists claim that the idea that we are free is consistent with the possibility of determinism, libertarians claim that the falsity of determinism is presupposed in the way we speak about free actions, choices, and decisions.

Both sides of the debate between compatibilism and libertarianism plead for intuitions supporting their view. And to me it seems that there truly are strong intuitions that support both sides. At face value, common sense backs a libertarian view: Our actions cannot reasonably be *both* free *and* governed by deterministic laws of nature.[238] Trying to combine these two seems like an attempt to have one's cake and eat it too, or at least like a position for which one would have to be a professional philosopher to even try to uphold. Furthermore, at first glance, compatibilism would seem to require some form of dualism.

Nevertheless, an "agnostic compatibilism",[239] according to which the topic of determinism is *irrelevant* for the question of freedom, has an intuitively sound point as well: It seems somewhat strange that the matter of free agency should depend upon the outcome of a controversy within the theory of science over whether or not the laws of nature are exceptionless. In Gary Watson's words:

> [T]he difference between free and unfree actions – as we normally discern it – has nothing at all to do with the truth or falsity of determinism. (Watson 2003, p. 338)

After all, we are not speaking about the possibility of non-governed, accidental behaviour (in which case we would not even apply the term *agency*), but of free agency in the Kantian sense

[237] On freedom "optimism" vs. "pessimism", cf. e.g. P.F Strawson 1974 and G. Strawson 1989/2004. The vocabulary points out the fact that we somehow *want* it to be true that we are free, since we otherwise apparently base our life on a giant self-deception.
[238] J.M. Fisher's "Principle of the Transfer of Powerlessness" might be seen as a formal preparation of this intuition (Cf. Fisher 1994, p. 8).
[239] The alternative to agnostic compatibilism is "deterministic compatibilism", according to which freedom is not only *compatible with* determinism, but that it even *demands* the truth of determinism, cf. Keil 2007b, e.g. p. 8. I do not discuss this position further here.

that we act according to *reasons* and are *responsible* for our actions. From this perspective it hardly seems implausible to accept our capacity to act freely prior to even taking a stand towards the status of laws of nature.

From a non-philosophical view, the debate between compatibilists and libertarians may seem hopelessly academic. A strict, Laplacean determinism is a rare position, and at that a purely theoretical one that cannot be empirically tested. And even for weaker forms of determinism (e.g. psychological or logical determinism) there is (and can be) no empirical evidence.[240] The compatibilistic claim is that *even if* determinism is true, this would not rule out our actions being free. It may seem futile to discuss such a hypothetical question, since compatibilists and libertarians agree on the important matter of recognising the possibility of free agency. The two ways of arguing reveals, however, substantial differences in the understanding of the concept of free agency. These differences may turn out to be decisive when it comes to putting up a defence against "freedom pessimists", i.e. against opposing views denying the possibility of freedom.[241]

A famous defence for compatibilism is to be found in P.F. Strawson's article "Freedom and Resentment". Strawson's move is to view moral attitudes such as approval and condemnation as generalised forms of emotional reactions such as resentment, gratitude, and love. These attitudes are "suspended" under certain conditions, namely when the person who is acting in a way that we normally would resent or praise, is someone who we do not reckon as responsible (notably a child or a mentally ill person). Normal emotional reactions and moral attitudes are those we have towards persons with whom we reciprocally interact – our co-subjects. In certain extra-ordinary cases, however, we take on an "objective attitude".[242] Decisive for Strawson's compatibilistic argument is that for conceptual reasons we must view such cases of objective attitudes as *exceptions*. As opposed to these exceptional cases, the determinist's claim that her being late to the appointment was an unavoidable result of preceding causes would contribute *nothing* to our moral judgement of her lateness. In other words, whether determinism is true or not is simply irrelevant to the question of whether we ought to consider each other as free and responsible agents.

Marcus Willaschek sees the advantage of Strawson's approach in that by understanding approval and condemnation as generalised forms of natural emotional attitudes,

[240] Although empirical evidence may support or weaken a deterministic thesis, cf. Keil 2007b, p. 38.
[241] Although "freedom pessimism" is a relatively rare view within philosophical circles, it is more widespread in other academic areas. The debate on free will is very much alive and has not least been fuelled as a result of progress within the neurosciences, cf. below, Ch. 7.4.
[242] Cf. Strawson 1974, p. 81.

he avoids having to reinterpret them instrumentally the way many previous compatibilists, e.g. Hume and Moore, did:

> Er [Strawson] kann daher anerkennen, dass der Grund dafür, dass wir Menschen loben und tadeln, in erster Linie in dem liegt, was sie *getan* haben, und nicht darin, dass wir ihr *zukünftiges* Verhaltens beeinflussen wollen. (Willaschek 2003, p. 200)

Willaschek does, however, agree with certain critical objections to Strawson's argument:[243] Strawson argues that *morally speaking*, the incompatibilist's claim is irrelevant. However, incompatibilists make a *moral claim* when they say that it is unfair to hold people responsible for their actions if determinism is true. We cannot simply brush aside this claim by saying that it is irrelevant in a moral discussion.

Willaschek argues that it is possible to *reconstruct* Strawson's argument in a way that meets this criticism. He attempts to accomplish this by liberating the argument from the discussion of whether a compatibilistic interpretation of punishment and reward is fair or not, and to see it as a conclusion based on two premises:

(P1) That an action is not free counts as an extenuation.
(P2) The (real or presumed) fact that all our actions are causally determined does not count as an extenuation.
(K) Freedom and determinism are compatible.[244]

The first premise is relatively unproblematic, and may be viewed as a conceptual truth: Freedom is a necessary condition for responsibility. The second premise is the decisive one here. Willaschek defends it by showing that neither our colloquial nor our legal use of the concept of responsibility implies a strong, incompatibilistic concept of freedom, but only a weaker one that is compatible with determinism. An incompatibilist would have to argue that our general use of the concept is wrong, an argument that according to Willaschek would be futile:

[243] He refers to U. Pothast and J. Wallace regarding this criticism cf. Loc.cit.
[244] Cf. Willaschek 2003, p. 201. In fn. 2 he formalises the argument thus:
(P1) Non-freedom => Non-responsibility
(P2) Determinism ≠> Non-responsibility
(K) Determinism ≠> Non-freedom

> Eine Begriffsanalyse, nach der sich herausstellt, dass der analysierte Begriff bisher immer falsch verwendet wurde, kann einfach nicht stimmen. (Op.cit., p. 204)

In other words, the incompatibilists in fact discuss a different concept of responsibility than the regular one, one of no practical concern. Thus, our *normal* concept of freedom as responsibility is shown to be compatible with determinism, since determinism has nothing to say regarding our essentially "local" references to cases of unfreedom. Willaschek admits, however, that nothing is thereby said as to *how* causally determined actions can be free.

To Geert Keil, this solution is unsatisfactory. What Strawson and Willaschek rightly argue is that we in everyday and legal practice hold each other responsible with no thought of determinism. But between this sound argument and their conclusion, that such a practice would be justified even in a deterministic world, there is a loophole. What Strawson and Willaschek fail to consider, is how to make sense of a concept of human agency and decision making which does not include the existence of alternative possibilities. According to Keil, a robust concept of freedom presupposes a *space* of freedom that the existence of deterministic laws of nature would exclude. If the outcome of the-event-which-under-some-description-is-my-intentional-action is predetermined by strict laws of nature, then – however unaware I may be of the outcome – there is really literally nothing left for me to decide, and nothing left for me to reason about:[245]

> Unterlassbarkeit [ist] eine analytische Komponente des Handlungsbegriffs und libertarische Willensfreiheit ein integraler Bestandteil der Handlungsfreiheit (…) Unsere gewöhnliche Rede über Handlungen, Überlegungen und Entscheidungen ist im Rahmen der selbstverständlichen vortheoretichen Annahme entstanden, dass die Zukunft offen und beeinflussbar ist, und dass wir im Handeln eine dieser offenen Möglichkeiten ergreifen. Wer diese Annahme zurückzieht, weil er den Weltlauf für alternativlos fixiert hält, sollte besser von *Quasi*-Entscheidungen, *Quasi*-Handlungen, *Quasi*-Überlegungen, *Quasi*-Fähigheiten und *Quasi*-Freiheiten sprechen. (Keil 2007b, p. 79)

Keil argues that compatibilists underestimate the degree to which determinism undermines our ordinary conception of agency. If we remove the actual possibility of "doing otherwise" from the concept of free agency, there is no freedom left that deserves the name.

[245] For a similar critique of compatibilism, cf. Searle 2001, p. 277ff. Keil argues that, although this reductio argument does not count as a *proof* against determinism, it shows that it cannot be reasonably claimed ("dass man ihn nicht *begründet vertreten* kann", cf. Keil 2007b, p. 77), since the determinist must view the result of any reasoning of his own – including the one leading to his advocacy of the determinist thesis – as determined.

This critique constitutes an argument against determinism, but not yet, it seems, a positive argument for the existence of a "robust" freedom of agency. Keil defends a "natural libertarianism", i.e. he does not refer to any supernatural entity in order to account for free will. Instead he assumes that the universe must be indeterministic to make room for free decision making. A common objection to this view is that it seems powerless when it comes to accounting for why an agent makes the decisions she makes. David Hume argued that a choice that is not determined is simply a random event, in other words something that cannot count as a rational choice. Natural libertarianism often points to quantum mechanics as a support for the claim that determinism is false, thus that freedom is possible.[246] However, as Roy C. Weatherford writes in the *Oxford Companion to Philosophy*:

> [T]he random behavior of atoms certainly does not by itself make for the freedom and moral responsibility asserted by libertarians. (Weatherford 1995; "Libertarianism")

The quantum mechanical argument in a sense "proves too much", since the indeterminism it suggests applies everywhere in the universe, and thus amounts to nothing more than "the freedom of a roasting jack".[247] In other words, "natural libertarians" must follow up the rejection of determinism with a positive account of how rational choice is possible in an indeterministic world. If my decisions are not determined, they seem to be a result of coincidence. In what sense can such "sheer accidents" support the assumption that I am a free and responsible agent?

Indeterminism might be said to constitute the *negative* or defensive side of libertarianism, namely the one that guarantees that nothing stands in the way of freedom. It must, however, be supplemented with a *positive* account of what constitutes freedom. We could say that the concept of freedom consists both of a "freedom from" and a "freedom to":

> Zum einen muss Freiheit positiv als Vermögen erläutert werden, beispielsweise als das Vermögen, praktische Überlegungen anzustellen und diese Überlegungen handlungswirksam werden zu lassen (…) Zum anderen muss dieses Vermögen in die Welt passen. (Keil 2007b, p. 106)

[246] It is worth noticing that quantum mechanics does not surrender the category of causality, which seems to suggest that the principle of causality is independent of determinism, cf. Keil 2007b, p. 42: "Der Determinismus gilt in der modernen Physik als weithin diskreditiert, während die Kausalität quicklebendig ist".
[247] Cf. Kant, KpV 174; cf. above, Ch. 2.4.

A negative side of libertarianism that is common to its different versions is indeterminism. This secures the "space" needed in order to establish a strong conception of freedom on the positive side: In Keil's version an "ability to do otherwise under given circumstances".[248]

I have largely left out the "negative side" of this discussion and focused on giving a positive account in support of the concept of free agency. As I see it, this implies an "ability to do otherwise under given circumstances". What I have not done in my thesis, however, is to move into a discussion of how the world would have to be so that this ability could "fit in", over and above a rejection of a dualist view of the world. Given a monistic world-view, a strict, Laplacean determinism cannot in my view be combined with free agency. In the next section, I will take a closer look at the problems connected with a compatibilistic concept of action. However, I remain agnostic as to what exact status laws of nature could have or would have to have in order for them to answer to the demands of a robust conception of free agency.

7.3 Compatibilism and Scientism

In contrast to the mainstream philosophical debate, I have hardly mentioned the concept of free will, and have instead investigated the concept of free agency. As I argued in Chapter 1, I see no strict lines between the two concepts. The reason I have focused on agency is that I disagree with the view that free will is a stronger concept than free agency; rather, the reverse is the case. In my view, actions are what can be referred to as free, fundamentally speaking, in the sense that they are what we are held responsible for. Other uses of the predicate "free" are derivative from this: Free (building of the) will is a part of *or* itself an instance of freedom of agency, while a person may be called free to the degree that she has an ability to act freely. I take free will to mean the freedom to "build one's will" in the sense of forming an intention and/or making a decision. An intention-in-action is part of the action itself, whereas the process of forming an intention prior to an action is a separate act, viz. a mental action:

> '[E]ntscheiden können' ist ja selbst ein Fall von 'handeln können': Plausiblerweise sind Entscheidungen selbst zurechenbare Handlungen, wenn auch mentale. (Keil 2007b, p. 58)

[248] "Anderskönnen unter gegebene Bedingungen", cf. e.g. Keil 2007b, p. 87ff.

Although Geert Keil speaks of *Willensfreiheit* in his book by this name, it is worth noticing that he defines freedom, not as a property, an attribute or a condition ("the condition of being free"), but as a *capability*, viz. the ability to *do* otherwise under given circumstances. In support of this definition, Keil refers to Kant and von Wright, as well as to Aristotle's breaking down of the concept of agency to "that which is in our power to do or not to do"[249]:

> Aristoteles' Feststellung, wo das Tun unserer Gewalt sei, sei es auch das Unterlassen, verschiebt das Gehalt des 'starken' Freiheitsbegriffs auf den Begriff des Tuns, Vollziehens oder Handelns. Wenn das Anderskönnen analytisch zum Handlungsbegriff gehört, implizieren schon unsere gewöhnlichen Handlungsbeschreibungen eine massive Freiheitsmetaphysik. Es gäbe keinen Grund, das Vokabular des Handelns oder Vollziehens in Anschlag zu bringen, wenn in meinem Körper nur etwas Prädeterminiertes, Unausweichliches geschähe. (Keil 2007b, p. 89)

As Keil emphasises, the "massive metaphysics of freedom" involved in our ordinary language of agency can hardly count as a *proof* of freedom. Further, the existence of a possibility to "do otherwise under given circumstances" cannot be empirically tested, since our world is such that an identical situation of decision making can never reoccur.[250] However, this argument seems to shift the burden of evidence to those who claim that free agency does not involve alternative possibilities. A common critique against compatibilism is that it does *not* allow for alternative possibilities, thus that it ultimately conceives of the course of events describable by reference to laws of nature as the superior mode. Many compatibilistic strategies seem (explicitly or implicitly) to adjust to the demands of naturalism by subordinating "the perspective of participation" under "the perspective of observation". However, some versions of compatibilism call such underlying scientistic prejudices to account. The question is whether these positions steer clear of Keil's criticism.

The goal of compatibilistic arguments is to show that the idea of a causally closed universe is simply irrelevant to us to the degree that we regard ourselves as free agents. In *Freedom Evolves*, Daniel Dennett pushes this argument to its limits:

> There are those who don't believe in free will and *thereby* don't have free will, and there are those who believe in free will and *thereby* actually have free will. (Dennett 2004, p. 13)

This view, however, severely limits the range and meaning of the concept of free agency. Against it, other compatibilists might argue that being a free agent is not about assuming a

[249] Cf. *Nicomachean Ethics* Book III. 1, 1110a10.
[250] Cf. Keil 2007b, p. 90.

certain belief *in particular*, but is implied in the act of believing (or disbelieving) *in general*, as well as in any other act. In this sense, mistrusting that one is a free agent is to already be one.

Compatibilistic arguments generally rest on the possibility of changing your perspective: What an acting person, as a participant, conceives of as something she can do or refrain from doing, she may simultaneously – qua scientifically minded observer – conceive of as a determinated state of the world.[251] Thus, Peter Bieri writes:

> Solange ich überlege und mir verschiedene Möglichkeiten vorstelle, ist die Willensbildung nicht abgeschlossen, und es ist wahr, wenn ich denke: Jetzt, während ich an die Alternativen denke, ist noch nicht alles festgelegt. Doch das Nachdenken über die Alternativen ist insgesamt ein Geschehen, das mich, zusammen mit meiner Geschichte, am Ende auf einen ganz bestimmten Willen festlegen wird. Das weiß ich, und es stört mich nicht, im Gegenteil: Genau darin besteht die Freiheit der Entscheidung. (Bieri 2003, p. 287.f)

That Bieri is "not bothered by" the double thought of his own act as simultaneously open for further consideration *and* determined as to a certain outcome, may simply be viewed as reflecting his own psychological make-up. Habermas, on the other hand, admits that he would easily be bothered if his decision was represented to him as a neuronal event:

> Es wäre *nicht mehr meine* Entscheidung. Nur der *unbemerkte* Wechsel von der Teilnehmer- zur Beobachterperspektive kann den Eindruck hervorrufen, dass die Handlungsmotivation durch verständliche Gründe eine Brücke zur Handlungsdetermination durch beobachtbare Ursachen baut. (Habermas 2004, p. 876)

Aspect dualism, according to which reasons and causes are two aspects of the same thing, is not in itself sufficient in order to ease the tension between the two perspectives. Even if we only presume a token-token identity between reason and cause, we would still have to specify further the relationship between the levels or aspects. A "crude" epiphenomenalism is blind to the special characteristics of the participating perspective and to the impossibility of getting beyond it:

[251] A common way to make sense of this relation is by assuming Davidson's theory of a token-token identity between mental and physical events: Mental events (tokens) are identical with physical events (tokens). This is a non-reductive theory, and he denies the possibility of "translating" between the mental and the physical. Type-type identity theories, on the other hand, rely on the existence of psychophysical laws, and say that for every mental event of type a, there is a physical event of type b to which it is identical, cf. above, Ch. 3.6.

> Um den Widerspruch zwischen dem naturalistischen Weltbild und dem Selbstverständnis handelnder Personen auszuräumen, genügt es daher nicht, *aus der Vogelperspektive des Wissenschaftlers* festzustellen, dass solche Alternativen und Freiheitsgrade nur *aus der Perspektive von Beteiligten* bestehen. Denn die handelnden Personen sind trotz der Endlichkeit eines situierten Verstandes, der den Weltlauf nur in engen Grenzen vorauszusagen vermag, mit den Personen, die gleichsam von außen oder oben auf die Welt im Ganzen schauen, identisch. Was die eine weiß, kann die andere nicht einfach ignorieren. (Habermas 2006, p. 688)

Martin Seel supports Habermas on the implausibility of seeing the participation and the observing perspective as mutually exclusive:

> Zwishen Teilnahme und Beobachtung besteht (…) keine strikte Alternative. Auch Beobachter sind Teilnehmer – potensielle Teilnehmer an einer Praxis der Rechtfertigung oder Vergegenwärtigung des Beobachteten. Beobachten ist selbst ein Fall der Teilnahme. (Seel 2005, p. 145)

What Habermas and Seel call attention to is that mainstream compatibilistic theories are based on a fundamentally scientistic assumption; namely, that a superior, scientifically objectivating observer's perspective *outranges* the more limited participant's perspective. This implies the danger of an "alternative conception", where the language game of scientifically predictable events is seen as independent from and superior to the language game of action, freedom and responsibility.

Martin Seel argues that many compatibilists see the participating perspective as subjected to the authority of a naturalistic observer's perspective. Thus freedom is degraded to an "epiphenomenon" in the sense of a "necessary illusion". However, it is rather the assumption that this perspective can be viewed as independent and superior that should be seen as illusory, since it implies a fictive "view from nowhere" in the sense of a "pure" observation beyond participation:

> Es ist gerade die Idee einer absoluten, letztgültigen, ultimaten Beschreibung 'des Universums', die inkonsistent ist. *Sie* ist die Fiktion, aus der – und aus der allein – sich ergibt, dass unser Freiheitsbewusstsein von außen betrachtet fiktiv ist. Sobald man sieht, dass die Fähigkeit zur Teilnahme an Praktiken der Rechtfertigung für alles Erkennen – und jeden verständlichen *Begriff* des Erkennens – grundlegend ist, bricht diese Konstruktion zusammen. (Seel 2005, p. 151)

In the pragmatic philosophy of language, the order of elements is reversed in such a way that the participating perspective is seen as superior. In this sense, to borrow a pun from Audun

Øfsti, the "view from now, here" is seen as a necessary, and in a sense primary, supplement to a "view from nowhere" in order to gain understanding.[252] That the participating perspective is superior does not mean that it is independent or self-sufficient, but that it is the necessary "framing" of the subordinate perspective. Even scientific observation is – at least virtually – participation in an intersubjective practice. If we didn't have the possibility of *objectivating* experience, though, we would remain within an *undifferentiated* "view from now, here" (like a bat, to adhere to the vocabulary of Thomas Nagel). It is the possibility of *perspectival change* that constitutes the multiplicity necessary for gaining access to the "space of reasons". As Tugendhat shows, a perspectival basis is necessary for our possibility to formulate non-perspectival truth in "eternal" propositions.[253]

The formal-pragmatic theory of speech acts points to a mutual dependency between the participating and the observational perspectives, each of them conditioning the other. Habermas elaborates on Hans Skjervheim's "simple opposition" between participator and observer,[254] turning it into a dynamic relation. It is this relation that Habermas refers to as a "complementary restriction of the perspectives of knowledge",[255] while Martin Seel speaks about an "asymmetric interdependence" between participation and observation:

> *Teilnehmer sind potentielle Beobachter, Beobachter sind virtuelle Teilnehmer.* (Seel 2005, p. 145)

The *asymmetry* of the relationship allows us to speak of a first-person priority – or (to adhere to Skjervheim/Habermas/Seel's vocabulary) the superiority of the participant's over the observer's perspective – without losing sight of the mutual dependency between the perspectives.

This line of argument traditionally makes little impression on methodological naturalists. From a naturalistic point of view, the priority of the observational perspective is justified in:

1) The unquestionable progress and prosperity of the methods of modern science. No other method of approach has had an even remotely comparable success when it comes to gaining new knowledge, as among others Hilary Kornblith argues:

[252] Cf. Øfsti 2000, p. 116.
[253] Cf. above, Ch. 5.4.
[254] Cf. Skjervheim 1959 and 1996, cf. also Habermas 1981, p.163f.
[255] Habermas 2006, p. 700.

> What does have priority over both metaphysics and epistemology, from the naturalistic perspective, is successful scientific theory, and not because there is some a priori reason to trust science over philosophy, but rather because there is a body of scientific theory which has proven its value in prediction, explanation, and technological application. This gives scientific work a kind of grounding which no philosophical theory has thus far enjoyed. (Kornblith 1994, p. 49)

Related to this, a further justification arises:

2) Sound scientific prudence and the rejection of apriorism. From this perspective it is always possible to continue research, thus gaining an ever-increasing overview over human consciousness and agency. Nothing is principally beyond reach for this scientific endeavour, not even the participatory, first-person perspective of the scientist herself.

As argued above, in 7.1., naturalism tends to give predominance to *ontological* questions about the existence of mental states as integral parts of the world over and above *epistemological* questions about what kind of access acting subjects have to the world. That these two methods of approach lead to different results should not surprise us, Habermas says. He argues that although an ontological dualism is out of the question, a "methodological" or "epistemic" dualism is inevitable:[256]

> Der Widerstand des personalen Selbstverständnisses gegen eine naturalistische Selbstbescreibung erklärt sich aus der Nichthintergehbarkeit eines Dualismus von Wissenperspektiven, die sich miteinander verschränken müssen, um dem in der Welt situierten Geist einen orientierenden Überblick über seine Situation zu ermöglichen (Habermas 2006, p. 688).

Habermas traces the two "complementary restricting" perspectives of *knowledge* back to two, simultaneously originating perspectives of the *world*:

> Die pragmatischen Universalien der Umgangssprache stiften für Sprecher und Hörer, die sich ihrer bedienen, um miteinander über etwas zu kommunizieren, einen doppelte Weltbezug: Indem sich die Teilnehmer im Horizont einer gemeinsamen Lebenswelt als Erste und Zweite Personen aufeinander beziehen, nehmen sie zugleich in der objektivierenden Einstellung einer Dritten Person auf Gegenstände in der Welt Bezug, von denen etwas ausgesagt werden kann. Die Teilnehmer an einer solchen

[256] Cf. Habermas 2004, p. 878: "Der methodologische Dualismus der Erklärungsperspektiven von Teilnehmern und Beobachtern darf nicht zu einem Dualismus von Geist und Natur ontologisiert werden". Cf. also the subtitle of Habermas 2006: "Wie lässt sich der epistemische Dualismus mit einem ontologischen Monismus versöhnen?"

> Verständigungspraxis verstehen sich als Personen, die einander für ihre Äußerungen Gründe schulden. (Op.cit, p. 693)

Habermas's methodological dualism shows itself in his suggestion for a "division of labour" between an *epistemologically primary* lifeworld and an *ontologically primary* objective world. According to him, these two dimensions of our involvement with the world are both expressions of the same rationalisation process, and do not come into conflict until the demand arises that scientifically objectivating self-description *replace* our "manifest" understanding of ourselves from within the lifeworld.[257]

"The alternative conception", as Apel defines it, is a variant of methodological dualism.[258] However, Habermas seeks to avoid the negative consequences of this conception by rejecting the strict opposition between the participant's and the observer's perspectives that is presupposed by Dennett and Bieri, and by suggesting a dynamic relation between the perspectives. By introducing a historic dimension into the relation, Habermas is able to argue in favour of a relative priority of the participant's perspective. Based on G.M. Mead, among others, Habermas introduces an evolutionary view of the non-circumventability of a "complementary restriction of the perspectives of knowledge":

> Aus einer pragmatischen Sicht, die Kant mit Darwin versöhnen möchte, spricht die These der Nicht-Hintergehbarkeit dafür, dass die komplementäre Verschränkung anthropologisch tief sitzender Wissensperspektiven gleichseitig mit der kulturellen Lebensform selbst entstanden ist. Die Hilfsbedürftigkeit des organisch 'unfertigen' Neugeborenen und eine entsprechend lange Aufzuchtsperiode machen den Menschen vom ersten Augenblick an von sozialen Interaktionen abhängig, die bei ihm tiefer in die Organisation und Ausprägung der kognitiven Fähigkeiten eingreifen als bei irgendeiner anderen Spezies. (Habermas 2004, p. 884)[259]

Habermas argues that the evolutionary approach shared by many compatibilists – e.g. Daniel Dennett – seems itself to undermine the scientistic assumption of the priority of a third-person perspective understood as a self-sufficient "view from nowhere"[260]:

---

[257] Cf. Habermas 2006, p. 694.
[258] Cf. above, Ch. 6.2.
[259] Cf. Habermas 2006, p. 700: "Offensichtlich markieren die Fähigkeit zu gegenseitiger Perspektivenübernahme und die Beherrschung einer propositional ausdifferenzierten Sprache einen tiefen evolutionären Einschnitt."
[260] Habermas quotes Thomas Nagel's *The view from nowhere* (1986) in a footnote in the English version of this essay: "If we push the claims of objective detachment to their logical conclusion, and survey the world from a standpoint completely detached from all interests, we discover that there is *nothing* – no values left of any kind: things can be said to matter at all only to individuals within the world" (Nagel 1986, 146). However, Habermas adds: "Since Nagel sticks to the mentalist opposition of first- and third-person perspectives, I won't go into his otherwise quite compelling critique of objectivism" (cf. Habermas 2007, p. 46).

> Die Endlichkeit eines aus der natürlichen Evolution hervorgegangenen und in der Welt situieren Geistes spricht vielmehr gegen die fraglose Unterordnung der teilnehmenden Perspektive, aus der uns die objektive Welt zunächst im praktischen Umgang mit unbeherrschten Kontingenzen begegnet, unter einen transzendenten Standpunkt jenseits der Welt, der kein bloßer 'Standpunkt' mehr sein darf. (Habermas 2006, p. 688)

The problem, however, with the idea of methodological dualism, is that it to a certain degree blurs the (necessary) possibility of transitions between different perspectives of knowledge. There is a hermeneutic inflexibility to Habermas's system, which seems to be reflected in formulations such as this one:

> Willensfreiheit ist eine zum Sprachspiel verantwortlicher Urheberschaft gehörende Voraussetzung. Der Inhalt dieser Präsupposition erschließt sich nur *Teilnehmern*, die als Hörer oder Sprecher eine performative Einstellung gegenüber Zweiten Personen einnehmen, während er *für Beobachter*, also aus der Sicht einer unbeteiligten Dritten Person, unzugänglich bleibt. (Habermas 2006, p. 671)

The presupposition of freedom is *inaccessible* from an observer's perspective; hence, the question of whether a person "could have done otherwise" has no application from this perspective:

> Für den Beobachter ist die Frage, ob die Person auch anders hätte handeln können, kein Thema. (Op.cit. p. 684)

In these sequences, Habermas seems to suggest – "alternative-conceptually" so to speak – that we can only recognise or acknowledge each other as free agents in a direct "I-you"-communication, not in e.g. "historic" sentences in the relative third person. This seems to undermine a sound, hermeneutic point that Habermas otherwise seeks to defend, as in the following quote:

> [Die] aus performativen Zusammenhängen bekannten Eigenschäften werden Personen auch dann zugeschrieben, wenn sie selber, zusammen mit ihren Praktiken und lebensweltlichen Kontexten, als 'etwas in der Welt Vorkommendes' beobachtet und beschrieben werden. (Op.cit. p. 693)

Here, the "secondary objectivated" (Apel), or "relative third person" (Øfsti) seems to be well taken care of.

To me, the question arises whether the ambiguity of Habermas's position is inherent in *all* versions of compatibilism. In that case, the same scientistic presupposition that Habermas diagnoses as underlying in many compatibilistic positions, is bound to pop up in his own version as well. In a footnote to his article on "the language game of responsible agency", Habermas accounts for the difference between his own position and that of mainstream compatibilism:

> Auch meine Position zielt auf die Einbeziehung des menschlichen Geistes und seiner komplementär verschränkten Wissensperspektiven in das wissenschaftlich erforschbare Universum der Natur; zugleich unterscheidet sie sich jedoch von dem 'kompatabilistisch' genannten Mainstream durch die Ablehnung der szientistischen These, dass dieses Universum als Gegenstandsbereich nomologisch verfahrender Naturwissenschaften (nach dem Normalvorbild der heutigen Physik) hinreichend bestimmt ist. (Op.cit. p. 704, fn 40)

This is an attempt to chisel out a modified version of compatibilism, a version free of scientific prejudice. However, I think there is an inherent problem in compatibilism in that the presupposition of freedom in the full sense – viz. in the sense of assuming that the agent could have done otherwise – only reveals itself as necessary and non-circumventable from within a participant's perspective. This seems to rule out what Apel calls secondary objectivation, i.e. the possibility of recognising the persons we are not directly involved with, but whom we speak about in the third person, as full-fledged, free agents.[261]

Compatibilism – no matter how carefully formulated – allows for viewing the physical substrate of actions as determined. Arguably, though, a robust conception of free agency should preclude the possibility to relativise actions as the results of determined laws of nature *from any perspective*, or under any description. Otherwise, it seems unable to support the strong interpretation of free agency, namely that the agent – as observed from a (relative) third-person perspective – *could have done otherwise*.

To be sure, if we assume only a token-token identity between the action and its physical substrate, this excludes the possibility of formulating a strict law by which the action can be foreseen. Even given full knowledge of every psycho-physical detail preceding it, we could still not draw a conclusion as to what type of action would follow. Nevertheless, given

[261] Obviously, a possible "Letztbegründung" of freedom – in Apel's sense – is only attainable from within a first-person perspective, from which a pragmatic inconsistency reveals itself when I try to deny that my acts are free. However, my point is that the right approach to free agency should support our ability to acknowledge other persons – who we speak *about*, not *with* – as conscious-in-action, viz. as subjects acting freely. Cf. above, Ch. 6.2.

the presupposition of strict causal laws at the physical level on one hand and the supervenience thesis[262] on the other, the outcome of (the substrate of) my decision making process would still *from a certain perspective* be determined. This might be compatible with my performative statement (in the first person present tense) that I can do otherwise, as well as with my appeal to you that you should do otherwise – after all this participating perspective is the superior one from which to view agency. I am still not convinced, however, that it is also compatible with statements concerning the ability of other agents – the people I talk *about* – to do otherwise (or even concerning my own actions in the past tense).

Not even the most carefully formulated versions of compatibilism steer clear of this dilemma, one that Habermas himself is aware of, of course, and which leads him to conclude that there are limits to what we can hope to achieve when it comes to understanding the interplay of freedom and determinism. To a certain degree it will remain an enigma:

> Rätselhaft bleibt einerseits die 'mentale Verursachung' von neurologisch erklärbaren Körperbewegungen durch verstehbare Intentionen (…). Aber in der umgekehrten Blickrichtung ist der Preis nicht geringer. Der Determinismus muss das Selbstverständnis rational Stellung nehmender Subjekte zur Selbsttäuschung erklären. (Habermas 2004, p. 886)

I agree with Habermas to the degree that our "double view" of ourselves as simultaneously free agents and as part of a natural course of events inevitably is an area of philosophical tension, as among others Kant, Nagel and Wellmer have pointed out.[263] However, if Habermas is right when he assumes that determinism forces the agent to view her own self-consciousness as an illusion – even if only from a certain perspective – then a scientistic prejudice seems to be built into the compatibilistic position itself.

7.4 Concluding remarks: Freedom as consciousness-in-acting

My main argument in Chapter 7 is that a robust concept of free agency rests on an argument of the priority of the first person, in other words the relative superiority of a participating perspective. A fundament for such a concept is the interventionist account of the relationship

[262] "[T]here cannot be two events alike in all physical respects but differing in some mental respect" (Davidson 1980, p. 214). In other words: If there is a change at the mental level (an action is carried out), there is necessarily a change at the physical level (an event takes place in space-time).
[263] Cf. above, Ch. 1.2 and 1.3.

between agency and causality. The problem with many compatibilistic theories of freedom is an underlying assumption that the scientific, observational perspective is superior. Habermas points out this scientistic fallacy; however, his position is marked by similar problems. Although Habermas explicitly dissociates himself from the scientistic way of giving prominence to the observational perspective, his modified compatibilism still threatens to collapse into an "alternative conception". By this I mean that his position does not seem to fully support a concept of free agency as a part of the world. Although (free) agency comes into being from a participant's perspective, it should be clear that it can also be recognised and ascribed from an observer's perspective *as* rational, free action, and not as behaviour determined by the laws of nature. Distinguishing the participating and the observational perspective is necessary in order to uncover the priority of the first-person perspective. However, such a distinction cannot justify a limitation of the scope of free agency, in the sense that it can only be recognised from certain perspectives.

Geert Keil argues that a robust conception of free action involves "the ability to do otherwise under given circumstances". This seems to exclude a strict, Laplacean determinism, according to which the total course of events in the world is fixed once and for all. The question is if it also excludes *any* form of determinism in the sense of nominalistic conceptions of causality, i.e. whether it excludes all versions of compatibilism. I do not try to give a final answer to this question, but restrict myself to pointing at certain inherent tensions and ambiguities, even in Habermas's attempted "non-scientistic" compatibilism.

The core of this thesis is not a negative limitation of the "metaphysical space" within which freedom is possible, however, or a theory of what the world would have to be like in order for us to be able to act freely. Rather, I have aimed for a positive account of what the concept of free agency implies, and of its central position in our relation to the world. In this last chapter I have argued an "asymmetric interdependence" between the participating and the observing perspective. This is based on the argument that freedom of action is a "non-circumventable" condition for rational argumentation in the sense that no proof against freedom is thinkable without running into pragmatic self-contradictions.

I have tried to avoid the implication that freedom is a purely "practical idea", viz. something that has a reference "only" to the degree to which we must act under this idea. Rather, I have argued that, given a linguistic-pragmatic view of knowledge, a concept of free agency is an *epistemological condition*, in the sense that it constitutes our access to objective knowledge about the world. According to the view I defend, free actions *exist in the world*. This primarily means that they have a *performative being* in Wellmer's sense. Furthermore,

however, it means that the free actions I perform – via indexical transitions – can be "conjugated into" the "logical space of reasons", i.e. into a space of true-or-false statements, and in this sense belong to *veritative being* in Tugendhat's sense.[264]

As I warned in the first chapter of my thesis, I am not conducting a systematic or historic tour through different positions within the philosophical debate on freedom. More specifically, I have made no excursions into the current debate which has risen as a consequence of the continuing progress within the neurosciences. I fully agree with Geert Keil's demarcation of the freedom problem as a "typical philosophical question", and a particularly tricky one at that:

> Das Problem der Willensfreiheit ist wie das Geist/Körper-Problem vielsichtig und tückish. Denkfehler, Verwechslungen, Kurzschlüsse und Kategorienfehler lauern an jeder Ecke. Ist man dem einen Fallstrick entgangen, droht der nächste, und allen zugleich zu entgehen erfordert enorme Umsicht. Kurz: Das Freiheitsproblem ist ein typisches philosophisches Problem. (Keil 2007b, p. 190)

Keil argues that the "neuro-debate" says little or nothing about what is relevant for this problem, namely whether we have an ability to do otherwise in *normal cases of agency*. He stresses that this is not to say that the findings of neuro-science cannot contribute to questions about freedom of agency. Specifically, these findings may be useful in practical matters, e.g. regarding the legal delimitation of pathological cases.[265]

What this thesis *does* attempt, is to clarify the concept of free agency through an alternative route, differing from both the traditional and the current debate. I have tried to display how deep-rooted the concept is, and thus, by implication, the cost of having to do without it. The route of my argumentation, chapter by chapter, has been such:

> 1) The introduction is an attempt to accentuate some basics about the thesis, among other things the fact that I discuss the concept of *free agency*, not free will. Another fundamental issue raised by the introduction is that, while retaining the sense of a *tension* between the different perspectives from which we may view agency, I aim to avoid *dualism* in my approach to the problem.

[264] Cf. above, Ch. 6.9.
[265] Cf. Strawson 1974: Concerns about the possibility that an action was not performed freely make sense regarding the *exceptions*, not the normal cases of agency.

2) In the second chapter, I look into *Kant's theory of freedom*. I approve of Kant's view that the relevant kind of freedom (the one worth having) is not a "Pinocchio freedom", viz. having "no strings attached", but rather the ability to be bound by reasons or norms. However, I criticise Kantian dualism, i.e. the divide between the realms of nature and freedom, making it in principle impossible to acknowledge the actions of other subjects, or even my own concrete acts, as free.

3) In the third chapter I defend an interventionist account of the relation between *causality* and *action*. This theory is a decisive element of an argument going through the thesis: That free agency is not only a *practically* "non-circumventable" idea, but an *epistemological* one, in the sense that it constitutes our access to the world in terms of *knowing* as well as of acting.

4) The forth chapter contains the outline of a *normative theory of action*, regarded as a generalisation of a rational-pragmatic theory of language.

5) In the fifth chapter I argue that normativity is based in intersubjective language. I also include the outline of an argument that this normativity ultimately must be explicated by reference to a categorical imperative, thus that it is based in practical reason.

6) Chapter 6 constitutes the core argument of the thesis: That given the right (unified) view of *language* and *world* respectively, a non-dualist account of human freedom is within reach. Thus, free agency is to be regarded not as an unknown x for which we must find a place within the existing world, but as *the necessary starting point* from which we can get to know the world.

7) Wrapping up my case in Chapter 7, I argue that a robust concept of free agency must rest on a non-scientistic assumption of first-person priority, but at the same time on the possibility to ascribe freedom to other agents. We cannot in my view settle this matter by means of a "duck-rabbit" freedom ("now you see it, now you don't").

I have argued that freedom cannot be analysed independently, but must be seen as an analytical component of the concept of action. To act is to act freely, i.e. we are not ascribing anything new to an action when we say that it is free. Free action is a *pleonasm*, since the concept action already contains a "massive metaphysics of freedom".[266] This begs for misunderstandings, since it is perfectly possible to say, e.g. "I was forced to do it". This

[266] Cf. above, Ch. 3.7.

would count as an extenuation, i.e. it would reduce or lift my moral responsibility. If I can be (politically, violently, or psychologically) forced to act, then some acts are clearly unfree – but then how can "free action" be a pleonasm? This is because freedom of action *under normal circumstances* is a transcendental-pragmatic condition in the cases of exception and extenuation. The idea of forced action derives its meaning only on the basis that actions normally are free. This is in line with Wittgenstein's argument in *Philosophische Untersuchungen*:

> 'What sometimes happens might always happen.' – What kind of proposition is that? It is like the following: If '*F* (a)' makes sense '(x) *F*(x)' makes sense.
>
> 'If it is possible for someone to make a false move in some game, then it might be possible for everybody to make nothing but false moves in every game.' – Thus we are under a temptation to misunderstand the logic of our expressions here, to give an incorrect account of the use of our words.
>
> Orders are sometimes not obeyed. But what would it be like if no orders were ever obeyed? The concept 'order' would have lost its purpose. (PU § 345)

Strawson makes a similar move against the attempt to represent actions as determined; it is only meaningful to assume this in the case of exceptions. If actions in general are seen as determined, then the meaning of our concept of agency and everything that goes with it dissolves. In this thesis I have tried to give good (practical *and* epistemological) reasons for maintaining a robust concept of (free) agency.

However, as I have argued, the reality of freedom is not something we can prove theoretically. In a sense, we cannot even reasonably view it as an intelligible belief among the other things we believe about the world and about ourselves. Rather, it must be recognised as a condition for the possibility of reasoning, thus as something that "shows itself" in the performance of acts. This is the idea behind Kant's characterisation of freedom as a "fact of reason". At the same time this "fact" should not be viewed as a pure practical idea or simply as "the limit of experience". Vital to a non-dualistic epistemology is that the actions we conceive of as free from the first-person perspective are the very same actions as the ones we experience as entering into the natural course of events. One way to avoid the dualistic implications of many attempts at clarifying this relation is, I have argued, to exchange the inner-outer distinction with a dynamic relation between a performative (participating) and an objectivating (observing) perspective. This way, the misleading mentalistic connotations of the inner-outer metaphor are avoided, and we can accentuate the necessary possibility of transitions between the perspectives.

What is then the fundamental meaning of characterising actions as free? In one sentence, it means that a consciousness-in-acting that is *in principle always explicable* in a performative 1pp-sentence (hence always declinable/ascribable to "relative third persons") is constitutive of agency.

Bibliography

Allison, H. (1990): *Kant's Theory of Freedom*, Cambridge.

(2004): *Kant's Transcendental Idealism*, New Haven & London.

Apel, K.-O. (1973): *Transformation der Philosophie, Band II*, Frankfurt am Main.
(1979): *Die Erklären:Verstehen-Kontroverse in transzendental-pragmatischer Sicht*, Frankfurt am Main.

(1980): "Zwei Paradigmatische Antworten auf die Frage nach der Logos-Auszeichnung der menschlichen Sprache", p. 13-68 in *Kulturwissenschaften*, H. Lützeler (ed.), Bonn.

(1988): *Diskurs und Verantwortung*, Frankfurt am Main.

Aristotle: *Nichomachean Ethics*, London 1975.

Austin, J. L. (1975): *How to do things with words*, Oxford.

Baker, L. R. (2000):"Die Perspektive der ersten Person: Ein Test für den Naturalismus", p. 250-272 in G. Keil, H. Schnädelbach (eds.), *Naturalismus*, Frankfurt am Main.

(2007): "Social externalism and First-Person Authority", p. 287-300 in *Erkenntnis* 2, Springer Netherlands.

Becker, W. (1987): "Zum Handlungsbegriff in Kants theoretischer Philosophie", p. 41-59 in *Allgemeine Zeitschrift für Philosophie 12.3*, Stuttgart.

Bennett, J. (1984): "Kant's Theory of Freedom", p. 102-112 in A. Wood (ed.), *Self and Nature in Kant's Philosophy*, New York.

Bieri, P. (2003): *Das Handwerk der Freiheit*, Frankfurt am Main.

Brandom, R. (1994): *Making it explicit*, Cambridge.

(2000a): *Articulating Reasons*, Cambridge.

(2000b): "Facts, Norms, and Normative Facts: A reply to Habermas", p. 356-374 in *European Journal of Philosophy*, Oxford.

(2002): *Tales of the Mighty Dead*, Cambridge.

Broad, C. (1952): *Ethics and the History of Philosophy*, London.

Carson, S. G. (1999): *Språk og Intersubjektivitet – En kritikk av den metodiske*

solipsismen, Trondheim. (Siri Granum)

(2000): "Handling og beskrivelse", p. 237-250 in *Norsk Filosofisk Tidsskrift* 35, Oslo.

(2002): "The Dependency of Non-Indexical Languages", p. 55-64 in A. Øfsti/ P. Ulrich/ T. Wyller (eds.): *Indexicality and Idealism II*, Paderborn.

Davidson, D. (1980): *Essays on Actions and Events*, Oxford.

(2000): "Die zweite Person"

(2001a): *Subjective, Intersubjective, Objective*, Oxford.

(2001b): *Inquiries into Truth and Interpretation*, Oxford.

Dennett, D. (2004): *Freedom Evolves*, London

Dray, W. (1957): *Laws and Explanations in History*, Oxford.

Fichte, J.G. (1983): *Gesamtausgabe der Bayerischen Akademie der Wissenschaften*, R. Lauth, H. Gliwitzky (eds.) Stuttgart-Bad Cannstatt.

Fischer, J. M. (1994): *The Metaphysics of Free Will*, Cambridge.

Fodor, J. A. (1990): *A Theory of Content and Other Essays*, Cambridge.

Guyer, P. (1993): *Kant and the Experience of Freedom*, Cambridge.

(2005): *Kant's System of Nature and Freedom*, Oxford.

Habermas, J. (1968): *Erkenntnis und Interesse*, Frankfurt am Main.

(1971): "Vorbereitende Bemerkungen zu einer Theorie der kommunikativen Kompetenz", p. 101-141 in J. Habermas/N. Luhmann, *Theorie der Gesellschaft oder Sozialtechnologie*, Frankfurt am Main.

(1973): "Wahrheitstheorien", p. 127-183 in *Vorstudien und Ergänzungen zur Theorie des kommunikativen Handelns*, Frankfurt 1984.

(1976a): "Universalpragmatische Hinweise auf das System der Ich-Abgrenzungen", p. 332-347 in M. Auwärter, E. Kirsch, K. Schröter (eds.), *Seminar Kommunikation, Interaktion, Identität,* Frankfurt am Main.

(1976b) "Was heißt Universalpragmatik?" (1976), p. 353–440 in *Vorstudien und Ergänzungen zur Theorie des kommunikativen Handelns*, 1984

(1976c): "What is Universal Pragmatics?", p. 118-131 in W. Outhwaite, *The Habermas Reader*, Cambridge 1996.

(1981): *Theorie des kommunikativen Handelns 1*, Frankfurt am Main.

(1985) “Questions and counterquestions”, in: *Habermas On Modernity*, Richard Bernstein (ed.), Cambridge, MA: The MIT Press.

(1999): *Wahrheit und Rechtfertigung*, Frankfurt am Main.

(2003): *Truth and Justification*, Cambridge.

(2004): “Freiheit und Determinismus”, p. 871-890 in *Deutsche Zeitschrift für Philosophie 6*, Berlin.

(2006): “Das Sprachspiel verantwortlicher Urheberschaft und das Problem der Willensfreiheit”, p. 660-707 in *Deutsche Zeitschrift für Philosophie 5*, Berlin.

(2007): “The language game of responsible agency and the problem of free will” and “Reply to Schroeder, Clarke, Searle, and Quante”, p. 13-50 and 85-93 in *Philosophical Explorations* 1, London.

(2008): *Between Naturalism and Religion: Philosophical Essays*, Cambridge.

Haga, Å. (1997): "Språk, intersubjektivitet og transcendental solipsisme", p. 221-252 in *Norsk Filosofisk Tidsskrift 32*, Oslo.

Haugeland, J. (1998): *Having Thought: Essays in the Metaphysics of Mind*, Cambridge.

Hegel, G.W.F. (1959): *Enzyklopädie der philosophischen Wissenschaften*, Hamburg.

Heidegger, M. (1927): *Sein und Zeit*, Tübingen 2001.

Hempel, C. G. (1942): “The Function of General Laws in History”, p. 35-48 in *Journal of Philosophy* 39, New York.

Hume, D. (1739): *A Treatise of Human Nature*, Oxford 1978.

(1748): An *Enquiry Concerning Human Understanding*, Oxford 1999.

Kant, I. (1902ff.), *Kants gesammelte Schriften*, Akademienausgabe (AA), Berlin.

GMS: *Grundlegung zur Metaphysik der Sitten*, Hamburg 1994.

KpV: *Kritik der praktischen Vernunft* , Hamburg 2003.

KrV: *Kritik der reinen Vernunft*, Hamburg 1990.

KU: *Kritik der Urteilskraft*, Hamburg 2001.

Rel: *Die Religion innerhalb der Grenzen der bloßen Vernunft*, Hamburg 2004.

Keil, G. (2000a) *Handeln und Verursachen*, Frankfurt am Main.

(2000b): “Naturalismus und Intentionalität”, p. 187-204 in G. Keil/H. Schnädelbach, *Naturalismus*, Frankfurt am Main.

(2004): “Anthropologischer und ethischer Naturalismus”, p. 65-100 in B. Goebel/ A. M. Hauk/ G. Kruip (eds.), *Probleme des Naturalismus*, Paderborn

(2007a): “Making something happen – where causation and agency meet”, p. 19-35 in F. Castellani/J. Quitterer (eds.): *Agency and Causation in the Human Sciences*, Paderborn.

(2007b): *Willensfreiheit*, Berlin – New York.

Kornblith, H. (1994): “Naturalism: Both Metaphysical and Epistemological”, p. 39-52 in P. A. French, T. E. Uehling, H. K. Wettstein (Eds.), *Midwest Studies in Philosophy, Vol. XIX: Philosophical Naturalism*, Notre Dame, IN.

Korsgaard, C (1996): *The Sources of Normativity*, Cambridge

(2002): *The Locke Lectures: Self-Constitution: Agency, Identity, and Integrity*, forthcoming in 2009, Oxford.

Kripke, S. (1982): *Wittgenstein on Rules and Private Language*, Cambridge.

Landemann, C. (1965-66): “The New Dualism in Philosophy of Mind”, p. 324-349 in *Review of Metaphysics* 19.

Lewis, D. (1986): *Philosophical Papers II*, New York – Oxford.

Locke, J. (1690): *An Essay Concerning Human Understanding*, London/New York 1910.

McDowell, J. (1994): *Mind and World*, Cambridge.

Mill, J. S. (1843): *A System of Logic*, in *Collected Works VII*, Toronto 1973.
(1863): *Utilitarianism*, Second Edition, Indianapolis 2001

Moore, G. (1903): *Principia Ethica*, Cambridge 1993.

Nagel, T. (1986): *The View from Nowhere*, New York.

(1987): *What does it all mean?*, Oxford.

(1997): *The Last Word*, Oxford.

Øfsti, A. (1985): “Act Performance and Description”, p. 9-49 in H. Høibraaten, I. Gullvåg (eds.), *Essays in Pragmatic Philosophy*, Oslo.

(1994a): *Abwandlungen*, Würzburg.

(1994b): “Searle, Leibniz and the First Person”, p. 682-688 in G. Meggle, U. Wessel (ed.), *Analyomen 1 Perspectives in Analytical Philosophy*, Berlin – New York.

(1994c): “Das Metasprachenproblem”, p.801-818 in *Deutsche Zeitschrift für Philosophie* 5, Berlin.

(1997a): “Das ’vertikale’ deiktische System der Umgangssprache”, p. 153-181 in *Kommunikationsversuche: Theorien der Kommunikation*, G.-L. Lueken (ed.), Leipzig.

(1997b): “Gesellige Ungeselligkeit”, p. 62-95 in J.-P. Harpes/W. Kuhlmann (ed.), *Zur Relevanz der Diskursethik*, Münster.

(1998): “Methodischer Solipsismus, Metasprachen(problem) und Deixis”, p. 24-72 in W. Kellerwessel, T. Peuker (eds.), *Wittgensteins Spätphilosophie*, Würzburg.

(2000): “Fregean thoughts and two dimensions of Kantian ‘thinking’ of intuitions”, p. 101-125 in A. Øfsti, P. Ulrich, T. Wyller (eds.), *Indexicality and idealism : the self in philosophical perspective*, Paderborn.

(2002): “Apriori der idealen Kommunikationsgemeinschaft: metaphysische oder dialogpragmatische Implikationen?”, p.78-129 in G. Skirbekk (ed.) *On pragmatics : contributions to current debates*, Bergen.

(2008): “Natur und Geist als zwei nichthintergehbare Rahmen”, p. 58-87 in W.-J. Cramm, G. Keil (eds.), *Der Ort der Vernunft in einer natürlichen Welt*, Weilerswist.

Quine, W. v. O (1974): *The Roots of Reference*, La Salle.

Riceur, P. (1977): *The Rule of Metaphor*, Toronto.

Rödl, S. (1998): *Selbstbezug und Normativität*, Paderborn.

(2000): “Normativität des Geistes versus Philosophie als Erklärung”, p. 762-779 in *Deutsche Zeitschrift für Philosophie* 5, Berlin

Rohs, P. (2003): “Libertarianistische Freiheit”, p. 39-60 in S. Mischer, M. Quante, C. Suhm (eds.), *Auf Freigang*, Münster.

Russell B. (1913). “On the Notion of Cause”, p. 171–196 in *Proceedings of the Aristotelean Society*, London.

Searle, J. (1969): *Speech Acts*, London.

(1979): *Expression and Meaning: Studies in the Theory of Speech Acts*, New York.

(1983): *Intentionality*, New York.

(1990): “Is the Brain's Mind a Computer Program”, p. 25-31 in *Scientific American* 262, New York.

(1998): *Mind, Language, and Society: Philosophy in the Real World*, New York.

(2001): *Rationality in Action*, Cambridge.

(2007): "Neuroscience, Intentionality and Free Will", p. 69-76 in *Philosophical Explorations* 1, London.

Seel, M. (2005): "Teilnahme und Beobachtung", p. 141-153 in *Neue Rundschau* 4, Frankfurt am Main.

Sellars, W. (1956): *Empiricism and the Philosophy of Mind*, Cambridge 1997.

Schopenhauer, A. (1839): "Über die Freiheit des Willens", in *Sämtliche Werke III*, Stuttgart 1962.

Skjervheim, H. (1959): *Objectivism and the Study of Man*, Oslo.

(1996): *Deltakar og tilskodar og andre essays*, Oslo.

Strawson, P.F. (1959): *Individuals*, London.

(1966): *The Bounds of Sense*, London.

(1974) "Freedom and Resentment", p. 72-93 in G. Watson (ed.): *Free Will*, Oxford 2003.

Strawson, G. (1998, 2004). "Free will", in E. Craig (Ed.), *Routledge Encyclopedia of Philosophy*. London. (Retrieved June 24, 2008, from http://www.rep.routledge.com/article/V014)

Taylor, C. (1985): *Human Agency and Language – Philosophical Papers 1*, Cambridge.

(1989): *Sources of the Self: The Making of Modern Identity*, Cambridge.

Tugendhat, E. (1976): *Einführung in die sprachanalytische Philosophie*, Frankfurt am Main.

(1979): *Selbstbewusstsein und Selbstbestimmung*, Frankfurt am M.

Von Wright, G. H. (1971): Explanation and Understanding, London.

(1980): "Freedom and Determination", p. 5-88 in *Acta Philosophia Fennica* 31, Amsterdam.

Watson, G. (2003): "Free Agency", p. 337-351 in G. Watson (ed.), *Free Will*, Oxford.

Wegner, D. (2002): *The Illusion of Conscious Will*, Cambridge.

Wellmer, A. (1967): *Methodologie der Erkenntnistheorie. Zur Wissenschaftslehre Karl R. Poppers*, Frankfurt am Main.

(1976): “Communications and Emancipation: Reflexions on the linguistic turn in critical theory”, p. 231-263 in O’Neill (ed.), *On Critical Theory*, New York.

(1991): “Adorno, die Moderne und das Erhabene”, p. 165-190 in F. Koppe (ed.), *Perspektiven der Kunstphilosophie*, Frankfurt am Main.

(1999): “Die Wahrheit über die Wahrheit”, p. 35 in *Die Zeit* no. 50.

(2003): “Der Streit um die Wahrheit”, p. 143-170 in D.Böhler, M.Kettner, G.Skirbekk (eds.): *Reflexion und Verantwortung*, Frankfurt am Main.

(2004): *Sprachphilosophie*, Frankfurt am Main.

(2007): *Wie Worte Sinn Machen*, Frankfurt am Main.

Weatherford, R. (1995): “Libertarianism” in T. Honderich (ed.), *Oxford Companion to Philosophy*, Oxford 1995.

Willachek, M. (1992): “‘Inneres Handeln’. Handlungstheoretische Überlegungen zu einem Grundbegriff des Perspektivismus”, p. 131-160 in V. Gerhard, N. Herold (eds.), *Perspektiven des Perspektivismus*, Würzburg.

(1998): “Agency, Autonomy and Moral Obligation”, p. 176-203 in C. Fehige, U. Wessels (eds.), *Preferences*, Berlin – New York.

(2003): “Freiheit als Bedingung für Verantwortung. Ein kurzes Argument für den Kompatibilismus”, p. 199-206 in S. Mischer, M. Quante, C. Suhm (eds.), *Auf Freigang*, Münster.

Wittgenstein, L. (1958*): The Blue and Brown Books*, Oxford.

TLP: *Tractatus Logico-Philosophicus*, London 1922.

PI: *Philosophische Untersuchungen*, Oxford 1953.

Wood, A. W. (1984): “Kant’s Compatabilism”, p. 73-101 in A. Wood (ed.), *Self and Nature in Kant's Philosophy*, New York.

Woodward, J. (2003): *Making Things Happen*, Oxford.