

Petter Nafstad

Reflections on the Human

Emotions

Table of Content

<i>Introduction</i>	3
<i>Chapter 1</i>	
<i>The Study of Human Emotions</i>	7
<i>Chapter 2</i>	
<i>Neural and Mental Mapping</i>	30
<i>Chapter 3</i>	
<i>Interlude - The Concept of Rationality as applied to the Emotions</i>	64
<i>Chapter 4</i>	
<i>The Formal Structure of Human Emotions as they relate to Rationality</i>	73
<i>Chapter 5</i>	
<i>Self</i>	113
<i>Chapter 6</i>	
<i>Replaceable and non-replaceable Emotion-objects</i>	145
<i>Summary</i>	157
<i>Bibliography</i>	159

Introduction

The usual way to approach the philosophical study of emotions is to start out from our emotional experiences. The feelings of anger, joy, envy, fear, sadness and so on, that we have all experienced are taken as the given facts, the data, and the research is carried out as an investigation of what is implied and presupposed in such experiences. Such approaches will reveal that emotional states contain cognitive components like beliefs, categorisations and propositional attitudes. I believe that such an approach is indispensable for the understanding of human emotions. However, even given that emotions are conscious phenomena it doesn't follow that they should also be explained exclusively in terms of conscious, or even potentially conscious, states. Such a conclusion would be based on the additional assumption that any true and explanatory relevant sentence describing an emotional state could also, in principle, be recognised by a subject as a true description of her mental condition.

It is certainly true that human emotions are the emotions of beings that also have rational and linguistic abilities. I also take it to be true that these abilities have impact on our emotional development. Understanding the connection between the various human abilities is part of the understanding of human emotions.

On the other hand, emotions might be approached from a perspective that doesn't so much relate them to our conscious and reflective awareness. Emotions might be viewed as parts of our "navigation-system", parts of the way we relate to and map our surroundings, identify our relative positions in these and prepare for adequate responses to changed conditions. To be sure, conscious awareness and propositional knowledge are parts of the human mapping-response system. But by focusing on the general system-features, we might see that

emotions can be studied, not only as a faculty of rational beings, but also as a species of organisms' ways to interact with their environment. The following text is an attempt to unite these two approaches.

One idea underlying the following is therefore that emotions might be regarded as something existing on different levels. There is a basic level where emotions exercise their influence as navigational devices in our orientation in the natural world. This function doesn't necessarily have to be described in mental terms. On this level William James' famous description seems, to me, better suited. According to James "*the bodily changes follow directly the PERCEPTION of the existing fact, and ... our feeling of the same changes as they occur IS the emotion.*" (James, 1884). In short, "we feel sorry because we cry, angry because we strike, afraid because we tremble ... "(Ibid). The mental quale that is usually considered essential to an emotion is, if we are to follow James, a mental response to an organic adaption that has already taken place independently of the conscious reaction, rather than part of the adaption itself.

There is however, I believe, another, a social level, where the conscious feelings play a more primary role. This is a level where our conceptualisations of our own states make them fit into inter-personal co-operation and action-schemes. It will be my claim that these conceptualisations have an underlying structure or grammar that corresponds to the structure or grammar of social situations. Emotions on this level therefore enable us to detect social situations and our own position within them, and the terms used to describe such emotions also serve to describe the corresponding situations themselves. On this level, so is my claim, the phenomenology of the emotions, the conscious experience of them and the experiencing of emotional adjustment, is part of their essential features.

Emotions play several roles in our lives and they may correspondingly be studied from different perspectives. I do not pretend to present a comprehensive theory of this multifaceted subject. For that reason I feel that I should say what this book is *not* about. First of all, it is not about the individual experiences of particular emotions like sorrow, love, happiness etc. Those interested in the phenomenology of these emotions will have an abundance of other literature to consult (for instance F. Alberoni's books on love, friendship and infatuation or Stendahl's writings on love). It is also not about the relationship between emotions and aesthetic experiences. And finally, except for a few remarks, it is not about the ways emotions influence our moral judgements. If it contributes to ethical theory, it does so only by implication.

In the first chapter I will discuss some general problems concerning the study of mental phenomena. In the second chapter I will introduce the concept of "mapping", the process of conscious or unconscious monitoring of the environment that any organism has to perform. The third chapter is an introductory discussion of the relationship between affects and cognitions. In the fourth chapter I discuss the composite nature of human emotions and the consequences of this as relates to rational articulation of emotional states. The key concepts emerging from this discussion will be those of a formal structure or grammar of the emotions and of emotional prototypes or paradigms. In the last two chapters I try to weave together the mapping aspect and the grammar aspect into a theory of the connection between emotions and the conception of "self" and "person". The main thesis in the last three chapters will be that certain emotions, viz. those which I shall call "social", function as a system mapping social situations, and that the same emotions play a major role in the construction of autobiographical selves. The theory of emotions as a system mapping social

situations will be presented in a sketchy and rather tentative way. However; social situations are “social” because we conceive of them or “construe” them as such. So, if we understood the mechanisms behind our construals of certain situations as “social”, I do believe that this would contribute to our understanding of what social situations are.

Useful comments to the manuscript have been given by Atle Måseide and Karl Halvor Teigen, both University of Tromsø. I have also benefited from the seminars of Paul Churchland, University of California SD.

Chapter 1

The Study of human Emotions

On the level of subjective experience, human emotions probably emerge as some sort of unspecified states of awareness whose dominant qualities are vague feelings of comfort or discomfort, not felt to be located in any specific part of the body. As we develop, we learn to think about these states as definite emotional states of various sorts, such as sorrow, happiness, anger etc. and also to think about them as mental, as opposed to bodily states. During our further development, we learn to “analyse” them, that is think and talk about them as having distinct and separate components like objects, motivating aspects, some normative dimensions, descriptive aspects and so on. At this point we may come to think of such states as more or less rational.

But the rationality approach, or rational assessment of our own ways to conceptualise our mental states, will be incomplete if it does not include asking the question whether or not we do actually give correct descriptions of these states as mental states. I take it for granted that when someone says that he is angry because John stole his money, he may in fact give a correct description of what John has actually done. But that does not imply that he is also giving a correct description of his own mental state or even that his conceptual framework provides an adequate tool for understanding this. So, what I am saying is that there are two separate rationality-questions here. One concerns the description of the intentional content of an emotion the other concerns this description as a description of a mental state.

It is possible to approach the study of emotions from two opposite sides that in certain respects relate to the two rationality-questions. One concerns the biological basis and function of emotions, the other one concerns their public expression and propositional articulation. One of the basic ideas upon which this text is founded is that the results of these two approaches need not be contradictory. Any investigation of the emotions would have to include some account of the system of discriminations that they express or realise. The kinds of discriminations I have in mind are those between different emotion-types, different dimensions or structural components of emotions and the degrees in which these are present in particular emotions. Any theory about the relationship between emotions and rationality would have to identify various aspects of emotional expression along which it can be evaluated. It will therefore have to assume that emotions can be analysed. Doing so, it will have to face the question whether the components in this analysis is a result of features of linguistic expression, for instance some basic universal grammar, or if it based on certain features of our biological, or, to be more precise, our neuro-physiological system. Now, we know that the neural networks in our brain map our environment along a number of dimensions, a number that varies with the different kinds of features of the environment that is mapped. We also know that these systems are capable of performing extremely fine grained discriminations, and that damages in different parts of these systems affect our conscious awareness of corresponding differences, including the emotional ones. I find it highly implausible that these facts about the working of our neural system should be irrelevant to the working of our conscious emotional discriminations. For these reasons I assume that descriptions on this level also have some kind of explanatory force as to why our emotional life is the way it is.

On the other hand, the fact that we are social and linguistic beings has important impact on our emotional development. More about this later.

Whether or not a theory that explains emotions in neural terms is correct, is an empirical question that cannot be decided on a priori grounds. The reason for this is that a priori decisions in such questions (except of course in cases of self-contradiction) confuse the subject with the interpretation of the semantic content of the name given to the subject.

Our ways of conceptualising our emotions very often have explicit or implicit references to action. Different emotions are often taken to be states that are typically followed by certain forms of action. The same emotions are also regarded as typical responses to situations that precede them. Some philosophers, like Gilbert Ryle, have regarded this as part of the logic of reports of mental states. Such reports would then really signify inclinations to act in certain ways. On the other hand, the same phenomenon could also be explained by the propensity that a perceptual system has to fill in missing typical pieces of information about a situation. Either way, the emotions are conceived as inter-mediate, in the way that they stand between a situation and actions that are regarded as typical consequences of such situations. Conceptualised that way, they come to be mental representations of situations and event/action-sequences as they pass through us.

Newborn babies are believed not to conceptualise their condition in mental terms as states of definite kinds. If the only correct description of a mind would be in terms that the mind uses to know or describe itself, then it would follow that there is no possible description of newborn minds at all. And if the criterion of success for such a description should be to make it vivid to us how it is for the baby to be in the state it is, then there would indeed be

no possible description of infant minds. But the fact that I cannot experience this is just another way of saying that I am not that infant.

On the other hand, we could describe what is going on in the infant's body including the neural system and the brain. And this would perhaps be the best description of that mind that we could possibly give. So, why wouldn't such a description still be the best when it comes to adult minds? One obvious answer to avoid easy a priori conclusions is that it might, and that the burden of argument is on those who say that it is not!

Let's assume that someone asks why an infant behaves in a certain way, for instance why it has a certain motor behaviour. An adequate answer could probably be given in terms of that infant's brain activity. Let's on the other hand assume that someone asks why an adult does certain things, for instance reads the newspaper or runs towards the subway-station. Would an answer in the same terms still be adequate? It might very well be the case that such an answer would be correct in the sense that certain brain circuits initiate and co-ordinate his behaviour into the sequences of movements that we recognise as reading or running. Even so, such an answer would in most cases not answer the question. Even if it were correct that what seems to be caused by a wish to be informed about the latest news or to reach the train really is caused by certain neural events, still the question would be unanswered. If we still were to say that the answer in neural terms would be more scientific, and therefore in a sense more true, we would probably also have to say that the question that would adequately be answered in terms of intentions is in some sense the wrong question, a question that, at least scientifically minded people ought not ask because there are no scientific answers to them.

What I have said here might seem to contradict what I have said earlier about a priori arguments. I said that it shouldn't be decided a priori that emotions, and more generally mental phenomena, are not really neuro-physiological processes. Now I say that answers in terms of such processes wouldn't answer certain questions, which very reasonably may be posed. However, there is a difference between saying that (1) an adequate answer to a certain kind of questions about an object, *o*, should have certain features, *f*, and (2) any account of *o* should have the features *f*. The adequacy of the answer is a function of the question, not only of the object. I believe that John Searle is right in assuming that questions about human actions in daily-life contexts should be answered in intentionalistic terms. The reason is that the meaning of such questions presupposes possible answers in such terms. But if we admit this, we should also recognise the possibility of other contexts where answers in other terms, e.g. neural, would be more adequate. Let's for instance consider conduct or behaviour that is considered as expression of some kind of malfunction like dyslexia or HDAD. In such cases it would often be apt to explain the behaviour in neural terms. Consider now that dyslexia and HDAD are not necessarily sources of malfunction under any condition. In fact the behavioural profile connected with HDAD would under many conditions give the individual an adaptive advantage. Would it, under such conditions still be adequate to explain his behaviour in neural terms? If the answer is no, why would such explanations be adequate on one occasion, but not on another? If yes: What's so special with this kind of behaviour that makes it call for a kind of explanation that is radically different from the explanation of other forms of behaviour? So, answering a certain question about an object is not necessarily the same as giving the best possible account of that object, at least if we are to follow Kant's advice to the scientist and ask like a judge addressing a witness, not like a student addressing a teacher.

How should we describe a certain state as “mental” or as a certain kind of “emotional” state, as opposed to an external state? We might of course indicate a different modality by using mentalistic words like believe, hope, fear or feel, followed by some proposition. Such expressions are taken to describe certain contents, the ones described in the proposition, as mental contents. The mental states thus connected with a proposition are usually called “propositional attitudes”. What we describe here is a “mind” as it is involved in, or part of, observable actions, relating to culturally defined situations, having projects in a social world that is ruled in certain ways. The use of first, second or third person modality indicates that the described interaction system is seen from the perspective of a certain individual partaking in that system. If that is so, what we are doing is to express and interpret some mental state by describing it in terms of external, publicly accessible phenomena, but seen from a certain perspective. But what about the mental mode itself, the way some content is said to be experienced? It can be named, but it evades further description. A description of a mental content is also a description of publicly accessible phenomenon. Therefore the mental seems to be indescribable except for exemplifications as perspectives on such phenomena. All this seems very reasonable. I can see no reason to assume that we have some state, that we in due course recognise as, say “embarrassment”, only to discover that this state fits perfectly into certain situations, those where we are caught doing something we shouldn’t do. But even so, features of externally observable situations are also represented as mental content. And as such they seem to be parts of the explanations; that is adequate answers to questions about why we act the ways we do.

The explanations in question would be intentional explanations. The question about why an adult is reading the paper or is running towards the subway-station is supposed to be answered in intentional terms, explicable as an intentional explanation. Intentional

explanations are of course based on the assumption that the ways we conceive of our own mental states have explanatory force as to how we in fact act. This in turn is based on the assumption that descriptions of mental states in terms of propositional attitudes might be true description. For instance; If x is running towards the subway-station and we explain this behaviour by saying that she does so because she wants to reach the train, then the following statements are supposed to give true descriptions of her mental state:

1. x desires to reach the train.
2. x believes that the train is leaving soon.

One more condition applies: It should be possible to identify the references of 1 and 2 independent of the fact that it claims to explain. However, we should consider it an open question whether or not this is possible. How should we for instance identify “a desire”? Two possible answers immediately present themselves. A desire is a tendency or disposition to bring about a state described in a proposition, namely the proposition that together with the attitude mentioned (desire) makes the propositional attitude in question. In that case our explanation of x’s behaviour would go like this: x runs towards the subway-station because x is in a mental state that tends to make her run towards the subway-station. Does this *explain* anything? Hardly! The other possibility is to define a desire in terms of antecedent conditions. A desire is something that people get under certain conditions. Given that someone depends on the subway to get to work, then this someone would desire to reach the subway. The explanation would then go like this: x runs toward the subway-station because x depends on the subway to get to work (and x is late). This seems to be a perfectly reasonable explanation. However, it works perfectly well without the use of any mentalistic term (desire). If the proposition expressing the explanation is true and if we accept it as an explanation, then x’s behaviour can be explained without reference to her mental state at all. But then again, this would hardly be an *intentional* explanation. Consider how easy it is in

daily life encounters to predict, or rather anticipate, the behaviour of other people without any reference to their mental states or propositional attitudes. You walk into a grocery store and ask for milk. You anticipate that the man behind the counter is going to get you some milk, which he does. Why? Well, because you asked for milk! You enter a bus and anticipate that the driver is going to drive to the destination where the sign in the front says that the bus is going. Surely he does just this. Why? Well, because that is where his schedule tells him to go! You stop your car at an intersection and behave towards the intersecting driver in a way showing that you are going to wait. He enters the intersection. Why? Well, because you showed him that you were going to wait! These explanations, or rather explanatory fragments, work perfectly well, and they make no use of mentalistic terms.

If anyone would try, from these examples, to infer that we could, in general explain human action without the use of mentalistic terms, the obvious objection would be that such terms (or rather the phenomena to which they refer) are, in some way, presupposed in interaction of the sort mentioned, and that the phenomena thus presupposed should be part of the explanation of the action. And basically I believe that this is a correct and adequate objection. We should however ask ourselves what this objection does and does not imply. *Being presupposed* does not imply *being object of conscious awareness* at the moment of action. At most it implies the possibility of being the object of conscious focus. But if so, then mentalistic terms are not necessarily parts of the self-understanding of the agents *as they act*. If we still insist that they should be part of the explanation of action, we would therefore also have to say that such explanations might go beyond the conscious intentions of the agents. Now, this seems very reasonable; but would such explanations still be intentional explanations? If so, we would also have to say that “intentions” include more than what is part of the conscious mental content. This wouldn’t necessarily be very

problematic. We might say that human action is performed on the background of some intersubjectively shared convictions about what is the case, that convictions about mental states are among these and that such convictions might, if the demand is made, be mobilised as reason for the actions even though they are not at the moment in the focus of awareness. This distinction between smooth, unproblematic, linguistically mediated interaction and, on the other hand, the mobilising of reasons corresponds to Habermas' distinction between communicative action and discourse (Cf. Habermas, 1981). If I have understood Habermas correctly, he is partly formulating a norm for rational interaction and cooperation and partly describing a communicative practice and its foundation in a modern differentiated life-world. But does the possibility of discourse also *explain* the pre-discursive interaction? I believe that it might do so in cases where this possibility is part of the conscious awareness when the action is initiated, like when you know that you can throw in more reasons if necessary without spending time figuring out in advance what those reasons more precisely would be. But in many instances this is simply not the case. In many cases the discourses are actions distinctively different from the communicative actions (or other kinds of actions) that precede them. In such cases the mentalistic terms might be part of the explanation of the temporally last action, but not of the first. This doesn't, I believe, conflict with Habermas' theory, as this is set forth as a theory of the preconditions for certain kinds of *rational* action with a certain degree of consensus-orientation, not action in general.

One more possibility must be considered. Perhaps one would say that intentional explanations do not presuppose the possibility of independent identification of motives or intentions. The motive, one might say, is considered to be part of the action of any actor as far as this actor is considered rational and capable of being ascribed an action at all. The reason would be that the concept of action in some way incorporates the concept of an actor,

which, in turn, presupposes the concept of an I who has taken some decision. It seems to me that this concerns the way other people are able to relate to someone's actions and only indirectly how the actor herself is able to do this. It is, in my opinion, indeed true that partakers in human communication and interaction relate to their own actions through some kind of anticipation of the likely response of other people. But that also entails that the *explanandum* of an intentional explanation is an action as it seen from at least one more perspective than that of the actor i.e. an action at the level of social mediation. I should add that I take this to be nothing more than a special case of the relation that Donald Davidson has called *triangulation* and that he takes to be basic in any knowledge or interpretation of any object in the world (Cf. Davidson 1991, 1996 & 1997). This concept is part of a theory that takes a subject's knowledge to be the result of a double relation, this subject's relation to the object of knowledge and her relation to some other subject's knowledge of the same object and response to the first persons conceptualisation of this object. The general significance of this theory is that "(T)he identification of the objects of thought rests, then, on a social basis. Without one creature to observe another, the triangulation that locates the relevant objects in a public space could not take place." (Davidson, 1991 p. 8). Davidson's theory is developed in relation to external objects. Nevertheless I find his arguments to apply equally well to conceptualisations of any object, external or internal.

Anyhow, the main outcome of this is that the presupposition of proper application of certain mentalistic terms applies to the possibility of explanation or interpretation of the action. If this is taken to imply that it also applies to the action itself, an additional assumption would be necessary: The self-understanding and self-interpretation are themselves parts, or in some way essential to, the explanandum. For many kinds of actions I take this assumption to be

perfectly legitimate. It does however leave the part played by human nature and the organic response not mediated by determinate conscious states unexplained.

I said above that the adequacy of an answer is a function, not only of the object in question, but also of the question itself. When something is done several conditions must be fulfilled, but that does not imply that these things should be part of any answer to the question about what is done and why it is done. Let's say that Peter takes his car and drives to his office. What is he doing? A good answer might be that he is driving to work. Why? Maybe because there is a bus strike. Surely he couldn't have done this if there were no fuel on his tank or air in the tires. A description or explanation of his action shouldn't for that reason include any account of these facts. There might however be given other descriptions of his action that would make such facts relevant parts of such description and explanation. These considerations might suggest that we should ask ourselves what kind of questions is it that makes answers in mentalistic terms adequate. Under what kind of perspective is a situation conceived when it seems meaningful to ask such questions?

One outcome of these considerations might be that it is in rather complex interaction-schemes, whose basic framework is one of deliberation of alternatives, actions that are performed under the conscious awareness of the possibility of some kind of failure and that are being performed under the conscious awareness that the action is part of a larger sequence of actions, that mentalistic terms enter into consideration. But if that is so, then such terms work in interpretations and explanations of actions as these are involved in more or less developed and culturally structured forms of action-schemes.

The problems relating to the use of mentalistic terms in explanation of behaviour can, therefore, be met with another kind of considerations, considerations taking common-life mental terms and their embeddedness in common-life human interaction as their basis. So far the remarks concerning the problems related to the use of mental terms has mainly referred to the single isolated agent and his mind. But there is also a world of human action-systems that we grow into and in which our emotions are developed and elaborated. It is possible that this world has a logic of its own, like the one described by Shakespeare or Machiavelli. This logic is based on mentalistic words like “ambition”, “greed”, “motives”, “hidden goals”, “pride”, “vanity”, “revenge”, “shame” and so on. Many writers, from Hobbes to Searle, has suggested that such terms lead us into a world different from the natural one, call it “artificial” or “intentional”. Maybe it is the way Thomas Hobbes assumes, that such words denote nothing but subjective experiences of some deeper process going on, the attempt to keep up the “vital motion” inside the body. Anyhow, in this world people act and other people respond in anticipated or not anticipated ways. As a result of these responses, new situations are created, situations that diminish the manoeuvring-space of the initial agent or create new, unexpected forms of interaction and institutions, and so on. To understand this world seems to be a question of understanding people’s actions in the terms of their own self-interpretations. But is this necessarily so? A closer look at such situations seems to reveal that what they really show is how people may reduce their own options (and perhaps ascribe this to destiny or fate). On the other hand it also shows that it is possible to study the relationship between, on the one hand people’s ways to conceive of themselves, and, on the other, their actions and construction of their action-space. Some such spaces are constructed as “political”, others as “social”, still others as “private”. There are emotions that fit into one space, but do not fit into the others.

Again; do the way we conceptualise our own mental states and actions really explain anything about why we think and why we act the ways we do? If the answer is going to be yes, there would have to be situations were someone is justified in saying the following: *x* gave a true description of the way she conceived of her mental state when she said that she desired that *d* and believed that *b*. This conception of hers caused her to perform the action *a*. So the fact that she desired *d* and believed *b* explains why she did *a*. But then again; how could we possibly know that such a description is true? We have, in the conceptual framework of intentions, after all, no independent access to the state of affairs described.

In spite of the problems mentioned; my answer to this question is going to be that the ways we conceive of our emotional states have explanatory force as to how we really think and act. One reason, given in the context of human co-operative schemes, is that the way we conceive of ourselves is a sort of anticipation of the corrections, sanctions, i.e. responses we are likely to be met with if we act in certain ways (under certain condition). The co-operative scheme as such is built into our conception of our own action and of the anticipated responses. We therefore adjust our actions in accordance with our way of conceiving them as parts of an inter-personal interaction-system. The reason why I believe that there is a legitimate way to study action that goes through the way in which the agents themselves conceive of their own actions is not that people's way of conceiving of their own actions necessarily also provide the only possible, or even a correct description of the same actions, taken as an existing entity waiting to be described. It is rather that it would be impossible to include some notions of purposes, aims, values and beliefs in the description of their actions if their own conceptualisations were taken to be irrelevant to their behaviour. And in that case the language games that are played within existing action-spaces like the political, forensic, private or public would lack the adequate pawns. The assumption of some adequate

relation between self-conception and action seems to be constitutive for a certain aspect of human cooperation- and interaction-schemes. This goes even when people misunderstand themselves. Let's say that Hobbes is right in assuming that when someone performs an act that he conceives of as an act of friendship, what he really does is to attempt to extend his own power. He cannot see this because he has no direct experience of "the vital motion" as such in his own body. So why should we pay attention to the agent's own description instead of going directly to the core? I believe that Hobbes himself provides us with the answer. Such a conception of action would be in conflict with an essential human need, the need to establish a space or an arena for human co-operation (in Hobbes' case, a political space, i.e. a space based on legitimate power). Put in a more general way; there must be some possible description of human actions that make them conceivable as part of a scheme of human actions. For this to be possible, there must also be some way to conceptualise this scheme (or these schemes). And there must be some points where the conceptualisation of the schemes and those of the individuals' selves, must converge. The individual's conceptualisation of herself is caused by the way cooperating individuals conceptualise their cooperative scheme. To follow the Hobbes-case, one such point of convergence is his concepts of the Rights of Nature, defining some essential property of persons as well as some constitutive rules of a certain sort of interaction-scheme. It is then, if I am right, the need for peaceful legitimate cooperation that explains why the individuals can understand their own actions as exercise of individual interests (as a source of problems) as well as natural rights (as a source to the solution of problems).

I will try to generalise the assumptions that are at the bottom of the foregoing considerations. I invite the reader to accept, at least for the sake of argument, that it is a fact that people's conceptualisations of their own beliefs and desires have explanatory force as to how they

really act. Having done this we should however ask under what conditions it is at all possible to form conceptions of one's own beliefs and desires (as well as those of others). Forming conceptions involve linguistic acts. Such acts should not be understood as involving a concept-giving subject and a conceptualised phenomenon only. Such a model would amount to a belief in the possibility of a private language. Concept-formation is part of interpersonal cooperative practices involving more than one perspective on the segment of the world that is the focal area of cooperation. Having beliefs and desires in the form of propositional attitudes relevant to action depends on the possibility that others can recognise that we have such. Our correct use of terms needs confirmation from those with whom we are involved in some kind of cooperation. And this in turn presupposes a common involvement in some action-scheme. Once this is the case, the language of beliefs and desires is the way we can make our own participation, and the ones of others, intelligible to others and to ourselves. And it is within the same framework that we can conceptualise, and hence comprehend, the various kinds of actions that are being performed. The bottom line of this is that it might very well be true that the language of mental qualia, intentions and propositional attitudes doesn't explain anything about our actions. But that holds only as long as these are considered from a perspective from which they are seen as not already involved in human interaction and cooperation.

A relevant and highly interesting example of mentalistic terms getting their significance in the context of cooperative schemes is to be found in the legal sphere. In American law the term "knowingly and willingly" plays a central role. You will probably find corresponding terms in all modern legal systems. In Norwegian law the terms "forsett", "hensikt" and "overlegg", meaning various degrees of premeditation and deliberation, denotes something that must be ascribed to anyone who is to be held responsible for an action in the field

regulated by criminal law. In spite of the obvious mentalistic character of these terms, they do not designate anything that could be identified independently of their meaning within the relevant action-space. In short, these terms get their significance through the way a certain cooperative scheme functions. The real foundation for their meaning is specific features of this system. There is of course the possibility that they refer to some independent mental entities, but anyhow; the alleged presence of such entities can only be demonstrated through the presence of some actions or utterances relevant to the specific case at hand (“He bought the gun the week before”, “He gathered information about the victim’s whereabouts” etc). So, I find it reasonable to hold that the kind of terms in question gets their significance through their role in the interpretation of the relevance of the law in relation to certain actions and the ascription of the relevant qualities. Apart from this, it would hardly be possible to ascribe to them any reference at all. An alternative strategy, which I consider to fall under the “hardly possible”, is, of course, to give some metaphysical account of such terms. This would be the strategy employed by St. Augustine or Kant¹. But even such strategies necessitate some kind of conceptual edifice, built up around concepts like “immortal souls” or “transcendental egos” to make the relevant terms intelligible.

The question behind these considerations was whether or not our conception of our mental states has some explanatory force as to how we in fact act. My answer has been that it has to the extent that explanations in these terms give adequate answers to questions about human behaviour that we can reasonably ask and because our actions are parts of different cooperative schemes. Such schemes must be understood by the participants to be of certain kinds, certain pragmatic modes. There are political actions because people understand certain acts to be of this kind. Had they not done so, there would be no such actions. Similar applies to other kinds of action. To be sure, this is not to say that people cannot misconceive

¹ The relevant sources are Augustine: *De libero Arbitrio* and Kant: *Grundlegung zur Metaphysic der Sitten*.

of their own actions. Actions that should properly be understood as political ones are, by the agents, often understood to belong to the private sphere. But were there no conceptions of a field of action as political, neither would there be such a field at all. Each field contains certain rules that action within this field should be (in the sense of “ought to be”) in accordance with. What I am saying is that people’s understanding of such rules has impact on how they in fact act. It could be the case that our actions and mental states could be given descriptions in quite different terms. Anger, revenge, greed etc can be described in, say terms of synaptic connections and brain circuits. In fact it is not an open question whether or not this can in principle be done. To deny it would amount to the implausible opinion that mental processes are not sustained by neural and other physiological (hormonal, muscular) processes that are specific to each mental modification. I believe that it is important to be aware of this and to recognise the explanatory force of descriptions in such terms, that is their explanatory force in relation to *certain kinds of explananda* involved in human action. If not, we would be in danger of trying to explain mental phenomena and consciousness under the presupposition that mental phenomena and consciousness already exist. We would also be unable to appreciate the enormous amount of discriminations and connections that are performed by our neural system beyond our conscious awareness, and the significance of these processes for our conscious life. But the kind of explananda that would be explained would not be a person’s actions considered as political, moral or social actions, but his actions considered as automated responses to his antecedent synaptic configurations. It would give a description of a person’s behaviour at another level than the one where we find politics or morals or even jealousy, envy, greed or happiness. Should a description in terms of brain physiology altogether replace this, it would also have to grasp the political, social and other kinds of modes of human action with the moral implications of these. But on the other hand, maybe it will, even if I can’t see how. After all I don’t believe that we should

place too much trust in apriori or categorial arguments when it comes to considering a possible future. And, as a matter of fact, I don't think we have to. After all there is, to my knowledge, no empirical evidence suggesting that the different modalities (political, moral, social, private) of human action could adequately be described in terms of brain psychology. But again, maybe the future will be different, and perhaps this will result in new and better forms of human interaction. Maybe such a change would alter the way people conceive of their own mental states. Maybe future beings will be able to be directly aware of their own neural processes, and conceive of these as the reason for their actions. If it were true that we need politics because we are greedy, or even that we are greedy because we recognise politics as a potential field of interaction, it wouldn't be a great loss to lose the need for politics. But even so, such a possible future state cannot explain why people interact the way they in fact do today.

Paul Churchland has, however, made the claim that the terms we use to conceptualise the mental states of ourselves and others are based on a theory of human action (folk psychology) that is in fact wrong (Churchland: 1988). He envisages a development where this theory might be eliminated and replaced by theories formulated within a different kind of paradigm, that of neuroscience. An alternative might be that it would be reduced to a description given on a deeper level. In that case we could still consider its explanations as true, even if there might be some more basic description of the same phenomena. But if it is correct that the theory is in fact untrue, we can, as Churchland points out, hardly expect to find some smooth one to one reduction to a basic level.

Again; whether or not Churchland is right here can only be decided empirically and with reference to common standards of evaluating scientific theories. One should for instance

evaluate folk psychology, provided that one accepts that it is a theory at all, as to its explanatory force in relation to phenomena that it should be expected to explain. I will not try to assess Churchland's theory here. It seems however to me possible that even if people understand their actions in terms of a wrong theory, this theory might still be part of the explanation why they in fact act the way they do. The alternative seems to be that beliefs play no role in actions at all. We now know that there are no witches, and most of us also assume that there has never been any. Even so, people have burned other people on the assumption that they were witches. And such an assumption might very well be part of the explanation why they did so.

On the other hand; we should, as the example shows, correct our wrong theories and stop acting in accordance with them. When we consider this it might be that there is a certain paradox connected to Churchland's position. Let's, as he does, assume that there is a possible correct theory of human action. Let's also assume that the terms of this theory would give us a true conception about what is really going on in the human mind. Let's also assume that people in general hold a wrong conception of their own mental states and actions. Now, do we at all act according to a theory? If yes, does this theory explain our actions? Is this theory the correct one or is it the one of folk psychology? Assume now that the answers to these questions are: yes, yes, and the correct one. My point is not that this involves some logical contradiction. It does not! There is rather a moral problem related to the relationship between scientific progress and civilisation. If our wrong theories do not affect what we really do (and if they do affect them we would have to say that these theories are part of the explanation why we act the way we do), what would be the moral benefit of scientific progress?

Besides: Wrong theories certainly affect what we say. And what we say certainly affect what other people do to us. So, in some way or other the existence of wrong theories determine or influence the course of events.

Alternatively we might answer that we act according to a wrong theory, but that this does not explain our actions in the sense that we would have acted likewise even if didn't have that theory. As we know, any event, or rather any description of an event, can be coherent with a number of wrong theories about the cause of that event. Coherence therefore does not ensure correct causal explanation. We think we do what we do because we have certain wishes, desires and beliefs, but that is simply wrong! We do it for some other, for us unknown reason, and this really explains why we do it. There was for instance a time when some people had some urge to burn other people, and so they developed a (wrong) theory about the nature of those people and of their own motives. Now, this is perfectly possible and even probably often the case. But should we say that it is always the case when common life vocabulary is involved in explanations of human action? My answer will, as the reasoning through this book will show, be no!

Whatever the answer to this question might be, Churchland's contribution is important in relation to the distinction between a mental state and the content of this state indicated above. His position seems to be that the explanatory power of the mental in general is not primarily to be found on the content-level. The reason is that this level is filled with false opinions about what is going on. This does not imply that he takes the semantics of the mind to be unimportant or without explanatory force as to how we act. However, he advocates a different theory of the origin of semantics, a theory based on the structure of neural system and its power to make distinctions.

Brains and experiences

The problem I have been discussing is whether emotions should be taken to be references of subject-related or absolute terms. A few words to clarify this distinction: Some terms refer to things whose existence depends on conscious activity of human beings, others to things that have an existence quite independent of any human perspective on them or any involvement in some human project. Among the first ones are nations, contracts, friendship, and war, among the second are oceans, pine trees, mountains and bears.

Not everyone would agree that there are things existing independent of the human perspective. Some will emphasise the fact that our perceptual and conceptual make-up determines every aspect of the world as it is presented to us. I will not argue against this conception, but merely point to the fact that it is also part of the human perspective on the world to presume and act in accordance with the validity of the distinction I have just made.

On which side of this division should we then place the human emotions? It would at first sight seem obvious that they are among the first class of things, and I believe this to be correct with regard to some aspects of human emotions. But that relates to human emotions that have already undergone a considerable development, i.e. the emotions of beings that have already learned to conceptualise their own mental states. It is certainly possible to formulate sensible statements about emotions that do not presuppose the development of such abilities. Such an approach will be based on the assumption that emotions do not have to be studied from the point where there is conscious awareness of some definite emotion.

It will regard such consciousness as a later stage in a longer sequence, a state that in a sense occurs after some of the most important things have already been put in place.

As a tentative approach to the question whether or not our emotions are some sort of “constructs”, I will assume that there are naturally given emotions, that these are “reconstructed” by us into certain conceptual complexes, and that such “constructions” come to have real impact on the way we actually feel and behave. The reason why I assume that there are naturally given emotions is that some basic emotional awareness of certain features of our environment have obvious evolutionary advantages.

In recent years neuroscience have made substantial progress in understanding the neuro-physiological processes that “underpin” our emotional life, at least with regard to some of our emotions. Joseph LeDoux (Cf. LeDoux, 1996), who has been particularly interested in the emotion *fear*, assumes that emotions to a large extent are determined by processes that take place outside our consciousness, i.e. without our conscious awareness of them. He also shows that some of the typical emotional-physiological reactions take place independent of conscious involvement. Antonio Damasio, who has developed a comprehensive theory of the biological basis of human consciousness (Damasio, 1999), starts from the same assumption. And it would indeed be odd to suppose that consciousness originally evolves from phenomena that are themselves conscious, or, even odder, that it pops up *ex nihilo*. LeDoux’s and Damasio’s approaches are thus very different from what is often termed phenomenological. By a “phenomenological approach” I mean an approach that starts from some consciously experienced mental content and proceeds by analysing the experience of this content and asking what is involved in such experience. Even though these two approaches are very different, I do not think that they contradict

each other in the sense that the results of the one would exclude the results of the other. I believe that they are about different levels of the evolutionary history of our emotions. Put in a synchronic fashion, they represent different levels of description. A short comment on the concept of “levels of description”: An object could be described from the perspective of the physical elements and laws that determine its existence. Take for instance a building like a 13th century cathedral. The building is a system of physical elements put together in such a way as to carry a certain weight and construction. The cathedral could be given a purely physical description that would be a description of materials and construction-principles. It could however also be given several other types of descriptions. Some would be in terms of the meaning and significance of the different elements and of the edifice as a whole. Some Italian renaissance architects in fact described their works in moral terms and conceived of the different architectural elements as moral expressions. But they were still masters of physical construction-principles. Had they not been, their architectural moral and religious expressions would have been impossible. What I am saying is that something might be given different, but true, descriptions on different levels. This is the case with mental phenomena. They can be described as neural events and sequences, and they can be described as conscious experiences. And I take it to be a plain fact that there could not be any conscious experiences, had there not been something that could be given a true description of the first kind. But this also implies that emotions are more than the conscious experience we have of them. It may hence be consistent to hold a theory of the evolutionary origin of consciousness and emotions on the one hand, and a theory of the structure of the conscious experience of emotions on the other. In fact, this could be stated stronger : If we take for granted that our emotions have a pre-conscious evolutionary history and we also take for granted that we have conscious emotional experiences which are accessible to phenomenological analysis, then it must be some possible consistent

theory including both levels. More about this later. At the moment it is enough just to state that we as human beings are biological beings, that we are among those biological beings that have consciousness, that we are conscious biological beings that have language, and that all these features are results of the development of human biology. All these things are important in the formation of what we know as human emotions, but it is not necessarily the case that everything that is part of the formation of these emotions is also part of their consciously experienced mode of being.

One reason to emphasise the importance of the phenomenology of the emotions is, as already mentioned, that this could open up an approach to a quality of emotions that is important for a certain kind of emotion-theory, namely the role played by our conceptualisations of our own mental states as they relate to describable action and interpersonal action-schemes. One thesis of this book is that this makes conceptualisations essential to *some* emotions, which means that they wouldn't exist as the kinds of emotions they are, were they not conceptualised. Another is that such emotions play an important role in the making of autobiographical selves and of an important class of social phenomena. So, I assume that it is an essential property of at least some human emotions that they are experienced as such, that they exist in a certain mode, as subjective experiences of the situations we are in.

All this implies that human emotions may be explored from two perspectives, bottom-up and top-down. The bottom-up approach will show us how our biological, especially our neural, devices have evolved so as to make possible and support conscious phenomena, among them emotions. This line of investigation has provided us with a theory of mind that works in terms of systems and networks. Now, one might give some standard anti-

reductionism arguments against this approach, saying that it does not describe consciousness as we experience it. One might say that it reduces the phenomenon to be studied in such a way as to change the very object. Descriptions of consciousness in terms of neural networks cannot be recognised as descriptions of consciousness at all. If such objections are valid, the bottom-up approach could be said to start out by raising one problem and, at best, end up by solving a different problem. But in fact I do not believe this to be the case. What we recognise as terms adequate to describe conscious states, is a question of acquaintance with cultural codes. The Greeks, and most noteworthy Plato, taught us to recognise descriptions in visual terms as descriptions of different modes of knowledge. There are descriptions in terms of systems and network that are at least as recognisable as descriptions of consciousness as are descriptions in terms of, say “reflection”, “insight”, “enlightenment” and other visual metaphors, or in terms of logical structures or mental qualia. Besides, the argument that the “how conscious phenomena are felt” cannot be recognised in neural descriptions and that such descriptions therefore cannot be descriptions of mental phenomena, misses an important point: The “how-feeling” cannot be described at all. A description does not picture a feeling. At the most, some descriptions presupposes them, for instance ostensive definitions. Perhaps the objection mentioned is based upon the feeling that the network approach diffuses the distinction between the subjective and objective side of consciousness, the first and the third person description. If so, it might be answered that the borderline between these two sides is a changing one. Descriptions that are at one time taken to be descriptions of objects are often later felt as descriptions of subjective experiences. Now and then you hear complaints that there is a trend to use the language of the computer-world to describe the mental. What such complains often ignore is the possibility that this linguistic development might correspond to a change in how mental processes are actually felt. After all, people

have probably always used analogues from technology that they considered to be advanced to conceptualise their mental states, and such metaphors may very well shape their own feelings of what is going on while they think or feel. Certainly the computer-language applied to mind is metaphorical. But we shouldn't therefore a priori assume that it is just another change of metaphors that is just as good or bad as another. One should even be open to the possibility that it in fact comes closer to a true description of what is in fact going on.

I believe however that there is one feature of human consciousness that perhaps will be inadequately described if dealt with in network-terms only. This doesn't have so much to do with the first-person aspect; it rather concerns the semantics of the mind. To put it plainly, it has to do with the meaning and content of mental phenomena. I must confess that I am not too certain about this, but I do suspect that a theory about the semantic content of the mind will have to contain some top-down approach. It is plain enough what "semantics" is all about. It has to do with the meaning of linguistic units. Applied to mental phenomena, I take it to deal with the so-called "intentionality". The adjective "intentional" is connected with a description of a mental content, what a mental state is about.

There is an argument against the claim that semantics can be explained in terms of the neural system, that goes like this: At some point in time Jimmy Carter decided to campaign for the Presidency of the United States. At that moment there was a certain synaptic configuration in his brain. Assume now that 5000 years ago, there was a man who at a certain moment in time had the exact same synaptic configuration. Did this stone-age man then decide to try to be the next president of the United States? However decisive this argument may appear, I believe it is a kind of sophism. Consider the following. July 12th

2000 I was sitting at my computer in my office in the philosophy department at the University of Tromsø trying to write a book on the human emotions. Let's now assume that 5000 years ago there was a man who, at a certain moment was in the exact same situation as I was on that date. Was he then sitting at his computer in his office in the philosophy department at the University of Tromsø, trying to write a book on the human emotions? If my situation was the one described and he was in the exact same situation, then, yes of course he was! But on the other hand, the conditional isn't very likely to describe any actual situation. Just as unlikely as the conditional saying that there was a stone-age man having the same synaptic configuration as Jimmy Carter had at a certain time. This cave dweller had no concepts of presidency, of states (united or not), of political parties, of election campaigns and so on. So how could it be that his neural system was mapping a situation identical with the one facing Jimmy Carter? After all there are virtually infinitely many possible synaptic configurations. And any realised possibility maps a situation in which a subject finds himself. In fact the argument presupposes that it is anachronistic (which it indeed is) to assume that the stone-age man should aspire to become president of the United States. On the other hand it presupposes that it is not equally anachronistic to assume that his brain-state was identical with Carter's the moment the former president decided to do so. In short, the argument presupposes the truth of its own conclusion. Admittedly, so does my counterexample. Of course, if A and B are in identical situations, then they are in identical situations. My point is that the Carter-argument does the same. It assumes a neural situation causing and supporting someone's decision to be the president of USA and then pretends that the existence of presidency and USA are not part of that situation.

On the other hand, the argument has a certain appeal because it reminds us of something important. The content of mental states must be described in terms of the phenomena they are about. If John loves Mary, then “Mary” is part of the content of this mental state of his. If he envies Paul because Paul is richer than he, then Paul and his wealth is part of this mental state. The case is more complicated than this, but for now it is enough to say that the intentionality of a mental state can be described in terms that seem to have certain meanings because they also refer to things that are outside the mind.

One way to put this is to say that syntax, or structure, alone is insufficient for the generation of semantics. One reason would be that syntax is conceived as a formal system that is indifferent to the semantic content of the units filling the spaces in the formal syntactic scheme. Any given space in any syntactic scheme can be filled with different words expressing different semantic content. The scheme in itself therefore cannot determine distinctions of meaning. Even if the conclusion is correct, I believe that one premise is false, or at least inaccurate. A neural system can be described as a formal (network) system. And in fact it is capable of making far more distinctions than is our language. I can distinguish between far more colours, tastes, sounds, shapes, and moods, etc than I can name and ascribe semantic meaning. In fact, our semantic system probably blurs far more distinctions than it clarifies (which probably is the intuition at the bottom of philosophical nominalism). So, if we were to hang our defence in favour of a theory that semantics require more than a formal system on semantics’ capacity to express distinctions, we would really have a bad case.

However; could we find a neural pattern in John’s brain, that we could possibly recognise as “Mary”, “Paul”, “money” or “theft”? The common sense answer, and the one I will

tentatively stick to, seems to be no! Now, I do realise that this is not so obviously true as it might seem at first sight. In fact a neural network-map could very well represent John's mind as containing a space of possible persons, a space where Mary and Paul hold exclusive positions. The same could be done with regard to possible actions. I find it highly plausible that this would actually explain how it comes to be that John can reserve a unique place in his mind for Mary and another for Paul. But the example at hand includes more than John's ability to individuate them among potentially infinitely many persons. It also picks them out as having some sort of special significant importance for John. My assumption is that the semantics of the terms, is a result, partly of distinctions/exclusions and partly of the significance attached to them. Such significance is influenced by the way John figures that these distinct entities have impact on him, and how he ought to organise his own behaviour towards them. All this makes up a kind of configuration including persons and their significance with regard to oneself and one's actions. Such configurations could be regarded as prototypes of various sorts. The tentative thesis is here that the component parts of such prototypes get their semantic meaning in light of the prototypical configuration as a whole and their general significance in people's lives. The word "theft" gets its meaning partly from the significance and importance that situations thus referred to have for our lives and our conception of our person (our own and other's) and of collaborate and other inter-personal schemes.

Let's, at least for the sake of argument, accept that semantics, in addition to being sustained by a system of discriminations, is based on the significance of the phenomena to which we refer. If so; are some information of a certain significant importance because it is processed in certain parts of the brain, or is it processed in these parts because of its special kind of importance? It is for instance assumed that information about imminent danger is

processed in the amygdala. It seems implausible that the danger represented by an approaching tiger depends on the fact that the information about it is being processed in the amygdala. The fear that is the result of such processing may depend on this. But the fear merely informs us that something of special importance is going on, it doesn't (normally) create the danger. This line of reasoning seems to indicate that a mental state refers to something external that determines the content, the semantic dimension of that state.

Conscious phenomena are experienced as subjective. But what does that mean? Especially: What would it mean that at least some emotions are distinguished by the conscious subjective mode of their existence? After all, any appearance in our consciousness is a subjective state. I can see a tree outside the window and the tree is my subjective perception. But there is a difference. I perceive the tree *as* something existing independent of my perception. This mode of perceiving is in fact distinctive of the way this and similar perceptions are experienced. Pride and shame, on the other hand, are not sensed in this way. However we conceive of these emotions, they have a subjective, first person phenomenology (some would even say ontology). Having a possible phenomenological description is part of what it is to be, at least some, human emotions. Such a description would cover conceptions of values, of ourselves, i.e. conceptions of who we are and of our future projects. Thomas Nagel has discussed the relationship between an objective description of consciousness and the subjective experience of conscious mental states. According to Nagel (Nagel: 1974), a being having consciousness, means that there is something for this being to be the kind of being it is. That does not imply that it would be impossible to give a neuro-physiological description of the perception-system of this being, but it does imply that such a description would not tell us what it is like to be this being. We may describe how bats fly around at night, navigating with a sonar-like system, how

they sleep upside down and so on. But this does not give us any knowledge about what it is like to be a bat. As Nagel points out; I might try to imagine how I would feel if I developed long fingers connected with skin, if I lost my sight and flew around at night hunting insects and slept upside down in caves. But still, what I would do even if I succeeded in imagining this, would at most be to imagine what it would be like for me to be a bat, not what it is like for a bat to be a bat. Correspondingly for us, and our emotions: we can imagine creatures quite different from ourselves being capable of a complete and accurate neuro-physiological description of how the human emotions work. But that would not give them an understanding of what it is like for us to be us. If this is correct, it must also be correct that such a description would not give us such an understanding of what it is like to be us, even if given by one of us. This certainly does not mean that neuro-science cannot increase our understanding of ourselves, far less that neuro-scientists cannot understand what it is like for humans to be humans. The main point is that we, neuro-scientists included, already have this knowledge from first-person experience. Everyone would agree that the third-person system-description can, and indeed does, increase our understanding of the way our emotional life works, and almost everyone would agree that this understanding can even contribute to change our attitude towards our own emotions. To deny this would amount to claiming that our understanding of our mental life cannot be increased by extension of the field of investigation beyond what can be directly accessed by our immediate subjective awareness. But still, without the underlying first-person experience of the significance of emotions and emotional nuances, we would miss some of the essence of the role played by emotions in our lives.

There are however a couple of problems with Nagel's position. First: if it is the case that a third-person description of our mental system (given by ourselves) really contributes to

changing our mental life, then it cannot be completely true that such a description doesn't say anything about how it is for us to be us. People have, for instance, probably always been confused. After sociology became a social force, some of us began to feel, not only confused, but being trapped in "role-conflicts", a term that originally was introduced as part of a third-person description. So, part of what it is for me to be me, is to be in role-conflicts.

Second: Nagel may be right in claiming that for a being to have consciousness is that there is something for this being to be this being. But the analogy between bats and humans underscores an important difference between bats and humans. Bats probably do not know that they are bats. But part of what it is like for humans to be human is to *know* that we are humans. And knowledge, as opposed to feeling, already implies perceiving from a third-person view. Knowledge goes beyond feeling. And, after all, neuro-science is not something conducted by Extra Terrestrials on humans, it is part of a larger project namely human's endeavours to understand themselves. Who could know what it is like to be a bat, and what does it at all mean to have such knowledge? To answer the first question there can be only one possible candidate, humans, since presumably only humans have propositional knowledge. But since they are not bats, they don't know. Knowing what it is like to be a bat excludes that you know what it is like to be a bat since knowledge is not part of what it is like to be a bat. Therefore no one can possibly know what it is like to be a bat. As to the second question, the answer has already been given; there is no such possible knowledge. More precisely, the closest possible answer would be that it is for a bat to feel "batness". But if that is so, saying that we can't know what it is like to be a bat, amounts to saying that "only bats are bats". But it is hard to see how this could teach us anything about

the relationship between human biological make-up and the knowledge about how it is for humans to be humans.

Maybe these comments don't hit Nagel's argument very good. After all he doesn't believe that bat-consciousness, being the "something" that it is for bats to be bats, consists in flying around trying to find propositional descriptions of its activities. But then again, perhaps his point would be better stated as an answer to the question "What does it feel like to be a bat?" rather than the ambiguous; "What is it like to be a bat?" Imagine a conversation like this:

A: I know what it is like to burn my arm on boiling oil.

B: Oh do you! I've never done that, so please tell me.

A: Well, it hurts!

B: Sure, but please describe the specific feeling of having that experience, so that I can recognise it and appreciate your pain!

A: It's like an extremely intense biting feeling that almost paralyses every other perception.

B: Those analogies and consequences still do not make me know what it is like to have such an extremely intense biting feeling.

A: Of course I can't describe it in a way that would make you know that. You will have to experience it, or something similar, yourself.

B: But you said you knew. And surely you can describe what you know.

A: Well, I meant that I know how it feels.

Bottom line: We can *know that* some events are taking place in A's body, but that does not imply that we *know how* that feels. We can also *know that* bats perform certain operations,

but that doesn't imply that we also *know how* that feels.² *Knowing how* is not a species of propositional knowledge.

Third: As an expansion of the preceding considerations there is a certain moral problem with Nagel's argument, at least if we are to make it relevant, not only to human-bat relations, but also to interpersonal relations. It is obviously correct that I am not a bat, therefore I cannot know the feeling of batness. But I am not you either, so I cannot feel what it is like for you to be you. But does that imply that I cannot know what it is like to be you in any sense? And what does it mean to know what it is like to be you? What would be a morally relevant meaning of such knowledge? After all I am expected to abide to certain rules in my actions towards you. Many of these rules can be applied only on the condition that I know something about you and the situation you are in. What kind of knowledge is required here? Is it for instance the first-person knowledge that I do not have about what it is like to be a bat? Now, I do not have that knowledge about you either, so I miss the conditions required if I were to follow moral rules in my behaviour towards you. Would things be better if I had such knowledge? I will never be very successful in pretending to be a bat. But I might be quite successful in pretending to be you. I might take over decisions concerning your life, I might tell you what to mean and do and I might pretend, or even imagine that your sorrows and joys are actually my sorrows and joys. In fact I believe that people often do that sort of things. If we give a certain phenomenon a name, we will more easily detect it when it is present. I propose we call the described

² This is actually a reformulation of an argument given in lecture by Paul Churchland. His version is an argument against the assumption that consciousness is not a feature of an organic system. The argument for the assumption mentioned goes like this: *Mary is a neuroscientist specialised in vision. In fact she is the world's leading scientist in this field. She knows everything that can possibly be known about our visual system. But Mary is born blind. She therefore doesn't know what it is like to see red. Therefore this knowledge cannot be inferred from knowledge about our visual system.* The argument can be formalised like this: P 1. Mary knows everything about our neural visual system. P 2. Mary does not know how it is to see red. Therefore: How it is to see red is not a feature of our neural visual system.

phenomenon *The Talented Mr. Ripley Syndrome* (TRS), the essence of which is a lacking ability to see the significance of the difference between oneself and another person. The reason why I do not consider this to be morally desirable, is that I believe it to conflict with a central ethical concept, that of “a person”. Since “person” is, as John Locke says, “a forensic term” it is perfectly possible to act as another’s person. I will treat this concept below. Here I will just say that it is used to ascribe certain rights, actions and experiences to human individuals and that it is used to define the borders between human beings. Pretending to be the other person is to violate these borders. In fact I believe that TRS in some form is an essential ingredient in all abuse. What I am saying is that it is desirable to know the situation of other human beings, but it is not desirable to confuse such knowledge with an attempt to experience another’s situation as he himself experiences it. My ethical point here is simply that we may represent other people in various respects, and such representation usually requires knowledge about them. But *representation* should not be confused with *substitution*.

Iris Marion Young (Young: 1997) has elaborated a conception of the moral significance of the points just stated. In her reconstruction of a moral conversation or relationship the asymmetry of the interpersonal relation is a constituent of its moral significance. The key term is “asymmetrical reciprocity” meaning that mutual respect is built on the assumption that each party involved has a perspective that is genuinely hers in a non-substitutional way. Any attempt to “take over” the other person’s experience, identify with the “what it is like” to hold her position, should be considered a kind of violation of integrity. Young’s basic message seems to be that the moral significance of human experience is inseparable from the fact that they are made from the perspective of individual persons.

This inference is however invalid since the meaning of “know” is changed from “know that” to “know how” through the inference.

Emotions and Rationality

We will now leave this for a while and start at another end. Many emotion theorists have asked questions about the relationship between emotions and rationality. In fact, explicit or implicit questions about this are probably among the oldest in the history of emotion theory. If the question is conceived to concern what we could call *the rationality of emotions*, it could be taken to mean two different, but interrelated things. It could either be taken as a question about the cognitive³ content of emotions, or it could be taken as a more general question about our possibility of some sort of reflective and corrective attitude towards our emotions. By “reflective and corrective attitude towards our emotions”, I mean our ability and willingness to adjust our emotional attitudes to reasonable norms, interpretations of facts, social expectations and personal goals. Interpreted either way, the question about the relationship between emotions and rationality starts with the fact that we are conscious biological beings with language. I consider this to be a legitimate starting point for an investigation of human emotions in spite of the fact that it puts aside a lot of facts that determine our emotional life. One reason for this is that language makes a difference because it makes it possible for us to relate ourselves to some symbolic aspect of objects, to ascribe some sort of significance to this aspect and to let this determine our actions. The rationality question simply starts from the fact of this difference. Language may not be the point where reflective consciousness starts, but it surely is the point where symbolic representations start.

Further; it is possible that emotions of different kinds play an important role in creating circumstances for human discursive practice that we usually conceive of as reasonable. When we, for instance, accept other peoples opinions and the reasons they give to hold them, there is often an element of credibility involved, an element without which such acceptance would have been impossible or at any rate difficult. The feeling we have towards someone whose credibility we accept can be called “confidence” or “trust”. In adult humans confidence/trust is a meeting-place between emotional desires, perhaps grounded in the need for safety and stability, and the consideration of reasons. This implies that there is an aspect of human life that should be understood as determined by certain features of emotion-holders who are also rational. All this touches a classic philosophical question: What is the relationship between accepting some description as rational and, on the other hand, being motivated by this description to perform some action? One might of course take the Kantian stand and say that part of the acceptance of something as rational is an implicit acceptance of its claim to motivate action. But that seems to presuppose that you are already emotionally attached to some social, cultural or institutional context that makes it meaningful to you to accept this.

The same state of affairs seems to be confirmed by Damasio (Damasio, 1994), whose case-studies show a connection between, on the one hand certain brain damages causing a loss or narrowing of the emotional repertoire and, on the other hand, a loss of ability to organise life and conduct on the basis of rational deliberation and planning. Damasio’s general conclusion on the basis of such evidences is that emotional engagement in one’s own life is essential even to instrumental because without it, there would be no significant goals.

³ I will use “cognitive content” in a restricted sense, meaning an object of which some property can be predicated.

Why should we be aware of our biological nature and evolutionary history as long as we start from the point where we (probably) depart from the rest of nature? One obvious reason is that our language and our symbolically structured life world is itself a product of our biological nature and evolutionary history. Now and then philosophers say things like this; “Yes, we are biological beings, but we are also more than that”. When asked what more, the answer is likely to be that we also think, reason, pass moral judgements and so on. This answer of course implies that this is not part of our biological nature. Organisms do not pass judgements or enter into moral obligations! But is this really true? Don’t they? Humans are biological organisms, and they pass judgements and make moral commitments. Doesn’t it follow that there in fact are organisms that do these things? And is the fact that it is organisms that perform these activities irrelevant for the understanding of the activities themselves? It would seem natural to say that the burden of argument is on those who say that our intellectual capacities have a non-biological origin. And this burden will not be relieved by mere declarations that material organisms cannot possibly have such capacities. Neither will it be relieved by declarations that it would be a “category-mistake” to assume that they have such, since a very slight switch of perspective makes it seem obvious that some of them indeed do have it.

The part of our biological make-up that is most directly related to our mental life is the brain and its neural system. Any description of the brain’s topography, its neural pathways and circuits, the way information is coded in receptor cells and transmitted to higher levels of the neural system, any such description that could also be recognised as relevant to understanding the structure of mental phenomena, is therefore of special interest. This is especially true if such descriptions also provide some explanation of the system of

discriminations or distinctions characteristic of mental and semantic phenomena and of the way mental phenomena are related to our interaction with our physical and social environment.

One good reason to pay attention to the biological foundation of conscious phenomena is that this allows us to direct the question of relationship between rationality and emotions at the proper aspect or level of emotional life. It therefore allows us to give more specific answers to questions like: “Can we alter our emotions voluntarily?” “What sort of cognitive structures are contained within emotions?”

One fact about us is that we have certain cognitive capacities. We are able to use concepts to identify things as things of certain types belonging to certain classes of things. We are also able to describe these things propositionally. Another fact is that we live within a symbolically structured cultural environment in which things and events get a significance that is not readily identifiable from their physical properties. There are some facts about humans that are such that they make us conceive of something as something else, a sound as a promise or a verdict, a piece of paper as money and another human as a friend, an enemy, a rival, a partner and so on. We also form inter-human relations and institutions (stabilised and formalised patterns of interaction) that we endow with various types of values of various nuances. We form certain attachments and engagements to which we ascribe special significance. These things do not exist as such apart from the human perspective on them. But the fact that such things do not have an existence separate from the human perspective does not make them less important for the understanding of human emotions. Many human emotional varieties are in fact constituted and developed in relations to such objects. Such emotions should themselves be regarded as belonging to the

class of objects not existing separate from the human perspective. Understanding them would include understanding how this perspective is structured and how it is experienced by the beings that have it.

Part of this perspective is determined by our cognitive and rational capacities. That is why I believe we should consider human emotions as closely related to these capacities. One aim of this essay is to illuminate the relation between the human emotions and the human rationality and cognitive abilities.

I find it important to make one point unambiguously clear: I do not mean that emotions *as such* are rational in any reasonable meaning of the term “rational”. We have feelings and emotions before we have a language. Animals that never develop language have feelings and in many cases probably emotions. I even believe that it is possible to have emotions without knowing that one has them, at least in the sense of “knowing that”. Such “knowing” denotes a mental state connected with some propositional content. And mental states with propositional content are presumably only possible in language-users. It would indeed be odd to imagine a newborn child that, in addition to being hungry also is the state of knowing that it is hungry. What I am claiming is that humans, at a certain stage of their lives, i.e. when they have acquired the ability of using language, have emotions that are related to rationality in both senses mentioned above.

Before we proceed I will comment shortly on the seeming paradox in the claim that it is possible to have an emotion without knowing it. The paradox might disappear when we consider that consciousness is not an either-or entity, but a graded phenomenon. Consciousness (as we know it) makes a kind of hierarchy with conscious focused awareness of propositional knowledge at the top. Perhaps the lowest level (which is often

considered as a state of lack of consciousness) is our mental state during δ -sleep (the deep dreamless sleep). Consider the following: We have all met strangers that we are instantly attracted to and others in whose presence we immediately feel uncomfortable, “the chemistry doesn’t fit”. In cases of the last sort we might be in a state of suspicion without being clearly aware of it. Further reflection might clarify that this is the case. Such reflection comes temporarily after the initial response. I am later (Ch. 5) going to comment on Damasio’s model of consciousness. In this model there is a top-level consciousness (extended consciousness) based on a process wherein past responses are being reactivated through memory (long or short term), and are then being responded to like any other object.

How should we conceive of the relation between the propositional content and the affective mode in language-users? Does an emotion “strike” an adult in the way that some known propositional content is connected to, or “synthesised with” some affective quality? Maybe! We might know something and suddenly be struck with jealousy or envy because of some known fact that had formerly been indifferent to us. But this probably happens to people who already have considerable emotional experience. More originally, emotions are likely to “strike” us as undifferentiated wholes. The analytical work of distinguishing between propositional content and emotional mode or quale, and of identifying the different component of the content, requires some reflective distance.

There is a very obvious methodological problem connected with a study of emotions: Are emotions a natural kind? There are some evidences that there are states that we usually conceive of as emotions that make a natural kind. As mentioned above, the physiological profile of fear is similar in all mammals (Cf. LeDoux and Damasio). But that doesn’t give

us very good reasons to assume that emotions like, say, benevolence, indignation, insult and love are instances of one and the same natural kind and far less that everything that we give the name “emotion” makes one natural kind.

The prima facie features that we come across when we consider emotions are probably that they are responses to some event and that these responses are modes of being engaged in, attached to or repulsed from something, and that these modes are in some way pleasant or unpleasant. The absence of “indifference towards something” thus seems to be common to all states that we usually take to be emotions.

At least at a basic level the responses in question should be conceived as effects caused by how we are affected by the corresponding event. Fear is, for instance caused by some perceived threat to the organism in its environment. What do such responses do for the organism, in what way do they influence the organism’s interaction with its environment? I will claim that at least part of the answer is that the emotional responses are ways to detect, map and amplify situations that we find ourselves in. I will return to the concept of “mapping”. Presently it is enough to say that the various dimensions of the emotional maps can be made conscious and analysed separately. Making emotions accessible to rationality is just this analytical process.

What kind of situations do emotions detect and map? To answer this question I find it necessary to distinguish between emotions that are basic in the sense that they are genetically determined, and emotions that are determined by cultural codes. Fear, as response to dangerous situations is an obvious instance of the first. Fear is a kind of mapping of danger-spots in our environment, natural as well as social. Envy, jealousy and

pride instantiate the last. I take such emotions to be a kind of detecting and mapping of social situations. They map the significance of such situations in relation to the systems of social hierarchies, values and behaviour-codes. They are, in short, responses to certain features of social situations, responses that inform us where we are in the social landscape. This social landscape, which is made up of the dimensions of hierarchies, values and codes, is represented as an emotional “mindscape”. The activation, in form of consciously felt emotions, of certain points in this mindscape informs us that the social situation to which this activation is a response, is located in a certain place in the social landscape and that we ourselves hold a certain position within this situation. From this, the following can be inferred: If it is the case that what we call human self and person in some way is socially constituted and situated, then emotions play an essential role in such constitution.

The general conceptions of the human emotions that I want to explicate are the following:

1. Emotions are mapping responses to situations in which the organism, and later the person, finds itself.
2. Human emotions, as indeed all conscious human awareness, starts with diffuse feelings of something taking place inside or outside the body, something that is felt as influencing the organism. From the start such feelings appear as states of various degrees of comfort or discomfort.
3. As we develop, we learn to interpret, conceptualise, categorise and explain what is taking place. We also make distinctions along lines of different qualities and degrees of

comfort and discomfort. As this process involves development of categories and concepts, it is related to our linguistic development.

4. The way we learn to conceive of inter-personal relations are of special importance for human emotional development. Such conceptions are based, not only on concepts and categories, but also on mental images of situations and sequences of actions getting symbolic significance. The significance of such situations and sequences can only be properly understood in terms of culture.

5. This process also changes the way we actually feel, or better, broadens the spectre of nuances of our feelings.

6. This process should not be regarded as one where the importance of one faculty, the emotions, taken as a constant, is diminished while another that of reason, also taken as a constant, is enhanced. Emotions and reason are levels of the same basic phenomenon, situation-mapping and response-preparation, more generally, adaption.

7. The study of human emotions can be conducted as a study of the result of this process, regarded as a result of a process.

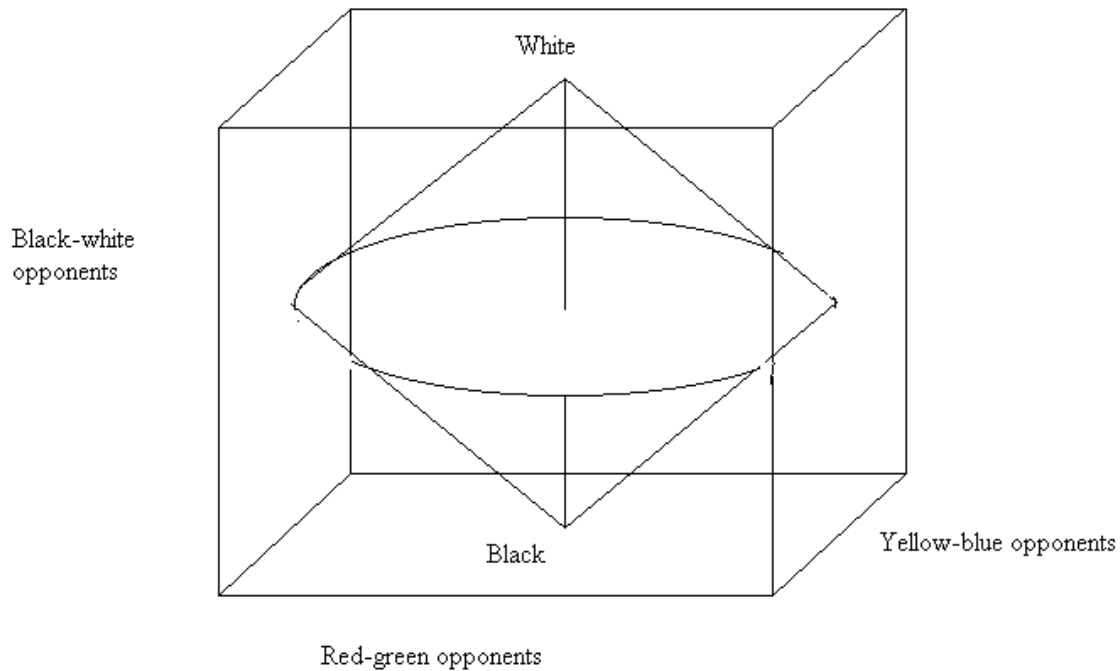
Chapter 2

Neural and Mental Mapping

I have occasionally been using the term “mapping”, mainly with reference to neural systems and conscious mental states. I am now going to discuss more explicitly the application of this concept in the context of mental phenomena in general and emotions more specifically. In a later chapter I will focus on emotions from the perspective of their formal, grammar-like, components. That is the perspective from which emotions are regarded as being capable of propositional articulation. The order of presentation is motivated by the theory that the formal component perspective may be an expression of a kind of multi-dimensionality that has deeper roots than preparing emotions for linguistic articulation. The formal-component perspective may be regarded as intermediate between a neural map perspective and a rational, linguistic articulation perspective.

The concept of neural maps has been elaborated in great detail by Paul Churchland (Churchland, 1995), so for a detailed analysis I refer the reader to his works. Churchland’s map-concept is elaborated from a theory of the working of our neuronal system and often demonstrated through models simulating such systems. An informed discussion of details about such systems is far beyond the scope of the present investigations. The basic principles on which the models are built are, however, quite simple, hence I will attempt to make a very brief outline of the underlying neuro-physiological and model-theoretical principles. An instructive (because of its simplicity) example of a neuro-map is the one of colours. Let’s consider the visual system as consisting of tree layers, a reception layer (in

the retina), a medium layer (in the lateral geniculate nucleus) and an out-put layer (in the visual cortex). Each layer consists of a population of neurons with a number of axons. Each axon can make synaptic connections with the dendrites of other neurons (in the next layer) whose axons in turn connect with the dendrites of still other neurons, so as to constitute a neural pathway. The input layer consists of three kinds of receptors, each being sensitive to electromagnetic waves within a certain, but for each kind different, range. These three range areas are the ones we conceive of as Red-Green opponents, Yellow-Blue opponents and Black-White opponents. The input layer sends impulses to various neurons in the mid-layer from where they are forwarded in different compositions to the output layer. Add now that the stimuli of the reception cells are vector-coded, that is they are stimulated along a non-discrete scale of intensity. The end result is that the input layer discriminates between a small number of different colour quality-types, while the output layer, that is the layer of conscious experience, discriminates between 10 000. Now, this output situation can be represented as a three-dimensional “space”. The dimensions will be the shades between yellow-blue, red-green and black-white. The space represents all colours that we can possibly conceive. Any specific colour can be represented as spots within the space.



Because this colour-space consists of only three dimensions and hence has a formal resemblance to space in the normal sense, it is easy to imagine. The next step will however be to extend the space-metaphor to cover all other kinds of mental phenomena and to imagine these spaces as consisting of far more than three dimensions. One of Churchland's favourite examples is face-recognition, probably because it can be simulated by computers and therefore demonstrated. Face-recognition operates along an unknown number (n) of dimensions. We can however imagine a n -dimensional space that is the space of possible recognisable faces. Each face would hold an exclusive place or point within that space.

Even if Churchland's map-concept is arrived at from a set of premises formulated as neuro-physiological descriptions, I believe that its, at least heuristic, value does not depend on the correctness of all these descriptions. In fact, I believe that we could arrive at such a

concept also from a top-down account, at least given the conception that mental states serve some orientation and navigation functions in human's interaction with their environment. This is not to suggest that the description in neuro-physiological terms is false. The point is that if identical conclusions can be reached from premises formulated in different kinds of terms, then, all the better. My claim here is that it involves no contradiction to conceive of a mental state as an intentional attitude, for instance involving desires and beliefs, and conceiving of it as performing some navigation-function. Both regard the states in question as something that can be analysed as complexes involving various dimensions. The intentionalistic approach points upward to a potential linguistic articulation while the mapping approach points downwards towards the some basic functions that must be performed by any organism in accordance with the sort of organism it is, the function of mapping the environment in which it finds itself situated. Whether it does so in virtue of the focusing of conscious awareness or at some other level isn't of decisive importance.

Except for a few general remarks, I have avoided the problem concerning the relationship between mental states and brain-states. I am not going to discuss this on the general level as a "mind-body problem". That approach tends to give us something that is either a problem that allows of no solution or a quasi-problem. I believe that Searle raises a timely question to which he also provides a correct answer: "Why do we still have in philosophy and psychology after all these centuries a 'mind-body problem' in a way that we do not have, say, a 'digestion-stomach problem'? Why does the mind seem more mysterious than other biological phenomena? I am convinced that part of the difficulty is that we persist in talking about a twentieth century problem in an outmoded seventeenth century vocabulary." (Searle, 1984. P. 14).

I believe that the concept of “mapping” provides a better starting-point than the general mind-body approach. The reasons are the following: First, it focuses on what the organism does rather than on its extension (in the Cartesian sense). Second, the concept of “mapping” can be applied to neural as well as conscious mental phenomena, independent of whether or not the first ones are conscious. Third, mapping taken as a conscious mental phenomenon is based on a system of discriminations and relations between the units discriminated. Such discriminations and relations are also established on a purely neural level. In fact it would be impossible on a conscious level if it were not sustained on a neural one. Fourth, instead of explaining our conscious mental discriminations as a function of neural discriminations, we could think of them as a function of linguistic ones. But since the number of discriminations tends to decrease along the line, neural-mental-linguistic, this seems to be a weak foundation for a general theory. I believe nevertheless that linguistic discriminations play a primary role in relation to a certain kind of mental phenomena, those related to symbolic intentional objects. This presupposes however that consciousness or mentality already exists and can therefore not explain its origin. Rather, consciousness should be regarded as a feature of the mapping systems of certain species, having a certain kind of mapping functions for these species.

The next step is of course to apply this mapping/space metaphor to emotions. The number of dimensions of which the space of possible emotions is constructed must of course be decided empirically, which is not going to be attempted here. My point here is that the formal components or categories (Cf. Ch. 4), if you prefer that term, can be conceived, not as open formal spots to be filled with empirical content, but as dimensions of an emotional space-map. This would not invalidate the formal investigation of emotions. It would rather

show that this approach applies to emotions of fairly rational or potentially rational language-users. It would also show that it applies to these because of the basic make-up of their neural system. It would in a way realise the Humeian naturalist program of showing the basis of our experienced mental states in the human nature. It would also be coherent with a certain notion developed by emotion-theorists like Solomon and deSousa, the notion of myths or paradigm scenarios. I will return to a more extensive treatment of this later. Here it is enough to point to the fact that neural networks have a recurrence-factor. That means that they are not in-put-feed-forward systems. In-put information is structured according to prototypes or paradigms that recur from the top-level back to the bottom-level. This is the neural network correspondent of the phenomenon that in hermeneutics is known as “Vorverstehen”.

One last remark: It seems rather obvious that consciousness has to do with our orientation and successful behaviour in our environment. I see that there is a wall in front of me, so I change direction. I feel the heat of something, so I withdraw my arm. I hear someone calling, so I direct my attention to where the sound appears to come from, and so on. These very trivial facts make it appropriate to think of vision, audition and other sensory modes as mapping-systems. Now, there are more phenomena than walls, calls and heat among which we need to navigate successfully. Among these are other people, their evaluations, social position, opinions, resources, etc. Were we not able to determine the various types of social situations, where they are located, what features they have, who are involved and in what way, we would be unable to act under the presumption that there is a social world at all. I can see no better candidate for the, indeed necessary, capacity that enables us to perform our social navigation than certain kinds of emotions. I shall however return to this below.

Inner States

Conscious states are usually thought of as “inner” to explain the kind of privileged access individuals seem to have to their own minds. And indeed, it seems obvious that each person has some kind of access that others do not have to her own inner states. This fact has been taken as evidence for different sorts of claims. Descartes, as we know, took it as evidence for the truth of certain kinds of propositional content of mental states. It has also been taken as evidence for the privileged kind of knowledge that each individual has of her mental state (Cf. Chalmers, 1996).

Whatever characteristics our special relation to our own mental states may otherwise have, I believe its “exclusiveness” to be of special significance. I call someone’s relation to her own mental states “exclusive” in a sense similar to the one we apply when we say that every material object occupies an exclusive place in space. The fact that some state is mine excludes that it is someone else’s. The term “exclusive” has certain advantages in this context. It can be used to remove certain apparent difficulties concerning what I can and cannot know about other people. I can know what kind of situation another human is in. I can also *imagine* what it would be like to be in that situation. I can therefore through analogy reasonably infer what kind of feelings she has; generally what condition she is in, what it is like to be her in a loose sense. But I am not that person! Now, does this narrow the range of my possible knowledge or imagination? Before we answer, we should consider that part of a person’s experience of being in a certain situation and responding emotionally to this situation, is the experience of being *an individual person* in that situation and mental state. Feelings are properties of individual organisms. The very logic

of feeling prohibits substitutive feeling. I take (social) emotions to be properties of individual persons. I will return to the concepts of “social emotions” and “person”. One of my claims will be that an essential part of what it is to be a person is to hold an exclusive place within a system of social situations. Presently I shall, however, restrict myself to a few comments on my use of the term “exclusive”.

The meaning in which I use the term “exclusive” has a certain reference to my conception of mental states as a kind of “mapping”. To put it very simple: If a mental state is conceived as a map of certain dimensions in the subject’s environment, the subjective mode of the mental content serves a locative function, the “you-are-here-spot” function. Let me give a very simple and trivial example with reference to visual perception. Visual perception obviously informs me about certain kinds of features of physical objects in my surroundings. Equally obvious: It also informs me of my location relative to those objects. Visual perception, like any other sensual perception comes in a double modality, objective and subjective or inner. In virtue of the first, it conveys information about the objects. In virtue of the second, it conveys, among other things, information about my position relative to those objects. This position is recognised as exclusive, as “my” in the way that it excludes that it could be any other person’s or any other position. It is the “where I am”.

It is probably quite easy to recognise visual perception as a kind of mapping involved in our navigation as we move around in our physical environment. Now, my assumption is that this is analogous to the way other modes of mental content, among them emotions, work. Fear, as an intentional state, informs us of danger-points in our action-space. The feeling of fear, the quale with which it is typically experienced, informs us of the relevance

that these points have for our security and well being. This relevance is, in turn, a kind of information about our location and position relative to those points.

I am later going to elaborate the distinction between basic emotions and social emotions. My hypothesis is that this distinction corresponds to a distinction between two aspects of our environment, the natural and the social. To anticipate a little: Social emotions map our social environment in the way that they are responses to social situations. They inform us about the various kinds of social situations we are in. Their characteristic qualitative features, the feelings of being envious, jealous, proud, embarrassed etc inform us about our own exclusive position within the corresponding situations.

I have no ambition to present these remarks as an exhaustive theory about the subjectivity or internal character of mental states. I am merely pointing to a feature that I find coherent with the conception of mental states as mappings. It seems quite obvious that a mapping system is useless, at least for navigation purposes, without some kind of “you are here” marker. And I can see no better candidate for this function than the exclusiveness signified by the subjectivity of experience.

What I have said here might be taken as a functionalistic account of our mental states, and hence subject to the same counter-arguments as, at least some versions of those theories. Such theories are typically based on descriptions of mental states in terms of their causal profile, that is antecedent conditions, and output results. Two interrelated objections can be made against this approach. It disregards the qualia accompanying mental states, and it is indifferent to the special characteristics of the physical features in which it is realised. If some device made of beer-cans was made to respond to objects that typically cause fear,

and if the output of this response-process were one that is typical of fear, then this device would have instantiated a state of fear.

I am however saying that the qualitative feelings are important, not just to avoid the objections mentioned, but for the operation, and generally for the significance of the mapping system. First we have the obvious need for the “position on the map” function that is maintained by the quality and intensity of the mental state. Second, the fact that some objects are suited as objects of, say fear, joy or sadness, depends as much on characteristics of the human body, that is the human biological system, as on features of those objects. Rattlesnakes are dangerous for humans, but not for beer-cans. Even if the beer-can device were made to avoid rattlesnakes this would not be because of some sense of imminent threat for the simple reason that it’s not threatened. An opposite assumption seems to confuse the description and the fact described.

There is however a certain problem as to how mental maps relate to the mapped landscapes. In what sense, if any, do the maps represent the relevant features of the landscape? I believe that the right way to approach this question requires a (pragmatic) rephrasing of it. I suggest the following: What would be the criteria of successful application of mental maps? This reformulation shifts the emphasis from the question of the correct representation to the successful application. The idea is again very simple: A road map is a good map if the correct use of it gets you where you want to go. This implies that the quality of the map should be measured, not only by its representative accuracy, but also by the success of the application of its implicit action-directives like: “If you want to go to Paris, take the second exit from where you are now and then turn left!” There is however a crucial difference between roadmaps and mental maps. Mental maps are

integrated parts of the map-users. A certain spot, or a configuration of spots on one or on a set of mental maps also signifies the mental state of the user. There is no homunculus outside that state, looking into it. Before we proceed from this assumption, we shouldn't however entirely rule out every possible implication of the homunculus-theory. Mental states are composite and parts of such compositions in fact monitor and oversee other parts. I take this last assertion to be no more problematic than one saying that we are in fact able to reflect on certain parts of our own mental states, i.e. make them into objects of self-awareness and self-investigation. But even so, the desires and wishes that the map might help us to fulfil, must, in some way, be internal to the mapping system itself, or else there would be no way to explain why certain maps are, in certain situations, picked out as focal points of attention. The mental state must therefore represent, not only some existing features of the surroundings, but also some desired or anticipated state resulting from the application of directives also represented in this state. For this reason I believe that the mapping-perspective on mental states must be supplemented with an intentional-content perspective. A certain emotional state is characterised by more than the features that distinguish it from other possible emotional states. It is also related to some substantial content and some substantial aim. I therefore believe that the term "emotional content" must be taken to refer to something at the conscious, not merely neural level, of the emotion. I take it to refer to the meaning of a potential description given by an subject of her emotion. Such a description would, among other things, identify the emotion-object. Now, how do such objects exist? I believe that they in a way represent something. But that does not necessarily mean that they should be regarded as mere pictures of things, events and conditions. In a way they also represent the subject's projects and anticipated conditions. John Searle (Searle, 1983) uses the term "intentionality" to signify "the conditions of satisfaction" of a mental state. That means that the intentional content of a

mental state represents the conditions that should be satisfied for this state to be, among other things, correct or adequate. Searle specifies this with reference to perceptions. If I see a certain object, say a house, the conditions of satisfaction for this perception are, 1. That there is in fact a house there, and 2. That 1. is the cause of my perception of a house. A propositional description of the intentional content of a perception does not only relate to something as existing independent of the perception, it specifies the condition for this perception to be successful. The term “conditions of satisfaction” can also be related to intentional actions, understood as actions intending to achieve or cause something in the world. Part of the intentional content preceding such an action is that the intended changes actually take place. Another part is that it takes place as an effect caused by this action. In this way we can say that an intentional state specifies conditions of its own successful application on the external world. In the concept of intentionality we therefore have a tool enabling us to see the connection between representation and action-directives.

In my opinion, Searle’s terminology is highly useful for the study of emotions, because it makes us focus on the inter-actional character of emotions (and other mental phenomena) and also because it makes it clear that such states have their internal standards of the success or failure of interaction. They are conceived as applied to, or applicable to, situations and they contain the criteria for the successful application to those. An emotion contains some conditions according to which it is an adequate or correct response to some situation. Let’s say I fear the monster under my bed. This fear has a content that may be articulated propositionally. This articulation will specify certain conditions for the fear to be a successful mental state. One such condition is of course that there is in fact a monster under the bed. As fear is related to danger, another condition is that this is dangerous. Of course this matter is more complex. Fear of something that is in fact harmless may be

functional and some times even rational. After all, we always live under conditions that are to some degree uncertain. But broadly speaking, in the long run we are better off if we correct our fear of monsters under beds.

The concept of “condition of satisfaction” may also be used to illuminate the relationship between emotions and motives. Motives are the action-directives implicit in the mental maps in the sense that they specify the correct way to apply the map to the landscape. To be sure: Emotions and motives come as part of larger mental webs. Envy, for instance, contains a motive to obtain something that is in the possession of someone else. The correct application of this emotion would therefore be some measure that led to this goal. Envy is however part of a mind consisting of other kinds of emotions, opinions, allegiances and so on. The correct application of the envy-map to the environment is therefore not necessarily also the correct application of the mental web taken as a whole.

I should add that I understand the concept of intentional content, not as something grasped through introspection, but as something that is clarified by a description of some (desired) external state or situation. This also applies to the subject whose intentional content it is. The way to clarify one’s own intentional state is, if my interpretation is correct, not to take a closer look into the soul, but to specify some external state of affairs.

In chapter 4 I will claim that human emotions can be analysed as having a sort of underlying grammar-like structure. If this is correct, one might ask whether or not they are also governed by some sort of rules analogous to the rules that we often take to govern our language. To be more precise: We usually assume that meaningful sentences are formed according to certain syntactic rules. Is it also the case that significant emotions, that is emotions that we recognise to be of certain kinds, are also formed according to certain

rules that are in some sense constitutive for their significance? My answer will be that in a certain sense they are, but also that this applies to emotions at a far more developed stage than we usually think.

The analytical approach that I will follow in chapter 4 justifies, in a certain respect, a conception of emotions as rule-conform responses. The reason is that such an approach tacitly includes the observable behaviour, not only the neural and mental events in the concept of “emotional response”. To be more precise, it views the neural and mental event under the perspective of its potential linguistic articulation and accompanying bodily expression. It therefore conceives of the emotist as participant in inter-personal schemes of co-operation and inter-action. As such she is subject to rules regulating intelligible speech and rational, or in other ways acceptable behaviour. But however justified this approach might be, its results do not entail that such rules are also constitutive for emotions as they first appear as situation-responses.

Chapter 3

Interlude: The Concept of Rationality as applied to the Emotions

In the philosophical tradition the question of the relationship between emotions and rationality is usually approached as a question about the relationship between two mental capacities or two sorts of mental entities each of which can be given an independent separate description. It is true that the question concerns mental entities, but that does not mean that this fact should also be the unquestioned starting-point for an investigation of the relationship between them. At a more basic level, all mental capacities can be regarded as ways that the human organism functions. The differentiation of functions is something that has an evolutionary history. Concept-formation, inference-conducting, reflecting on alternative ways of acting, culture-building as well as emotion-responding are all aspects of the human organism's way of managing the situations in which it lives its life. Some of these functions are evolutionary older than others. But that does not mean that these older functions remain unchanged through the evolution of the younger. And it certainly does not mean that the later developed functions have an existence independent of its evolutionary ancestors. To be sure; I do not believe that the evolution of reason changes all the older functions of the human being. But some of them are changed to various degrees. And some of our emotions are among these. Our attitude towards them is changed to a certain degree and so is their complexity and spectre of variety. Emotions are also changed by being objects of cognition, social norms and by their role in sustaining social norms and

even cognition (Cf. Elster, 1999, Ch. 4). In short, emotions change when they become part of a web of other functions.

What I will have to say about this does not necessarily apply to everything that can be termed “an emotion”. The aim is more modest, I want to characterise some distinctive aspects of some of the emotions of beings that also have the abilities that we think of as rationality.

“Rationality” however interpreted, involves some kind of abstraction, disregarding something as unessential, looking for some common or general patterns and formal similarities. I have said above that our emotional evolution from early childhood is a development from a diffuse state towards a mental life consisting of a variety of emotional states that can be identified as different from each other. The introduction of rationality implies a development that is, in a sense contrary to this. For an emotional state to be considered rational, it must be possible to measure or evaluate this state by standards that are to some degree inter-subjective and that can be given some kind of linguistic articulation. Now, our emotional language is much cruder than our actual ability to feel emotional changes and nuances, as is our ability to make linguistic distinctions in general much poorer than our ability to make sensory discriminations. Linguistic communication, therefore, encourages a kind of standardisation of expression, emotional as well as sensory.

Humans certainly have mental capacities that other minds do not have. But what seem to be distinctively human is not only this, but also the ability to co-ordinate a vast complexity of functions in a relatively coherent way over time so as to construe a life-story. For our emotional life this means that this functions as a part of a system that also includes our

cognitive and rational capacities, not, at least not only, as a faculty detached from the other mental faculties. It is therefore worthwhile to consider how our cognitive and rational functions are integrated in the functioning of the human emotions.

Human emotions have content, that is, they are about something and they are directed towards something. The same is probably the case also with non-human emotions. Antonio Damasio traces the origin of all consciousness back to the basic feeling that there is “something to be known” (Damasio, 1999). There is something going on outside or inside the organism that is in some way important to the organism. But the content of human emotions can also, at least partly, be articulated by the beings that have them, as statements about objects having certain properties. Because such statements are truth-related and often value-related, human emotions are also phenomena capable of being cognitively evaluated. More precisely, emotions may be, and in fact very often are, rationally assessed, they are criticised by others, and they are, at least occasionally, corrected by the subjects that have them. Such criticism and correction may follow different standards of evaluation. One of these concerns truth-value, we may get things wrong. Another concern values and norms, our emotional reactions may be inappropriate in some way or other.

Not all writers on the human emotions agree that emotions per se have content. There has been a discussion among psychologists over this topic. Some say that emotions essentially are affections; they are the affective mode that is characteristic of each of them. P. N. Johnson-Laird & Keith Oatley (Johnson-Laird & Oatley, 1989) claim that certain basic emotions (happiness, sadness, anger, fear and disgust) are these affects. We may describe situations that we conceive as proto-typical antecedents and consequences of these emotions. Such descriptions are however, according to these authors, not descriptions of

the emotions in question. “In short, the members of a culture have a prototype for the sorts of events that cause an emotion such as sadness, and for the sorts of events that ensue; but they do not have a prototype for the subjective feeling itself.” (Ibid.) If this were correct, the terms “affect” and “emotion” would be synonymous when used to designate basic emotions. (It should be added that, according to J-L & O, the case is different with complex emotions.) The arguments with which J-L & O support their hypothesis are basically semantic. Statements like, “I feel sad, but I don’t know why” or “I feel sad even though I don’t show it in any way” are meaningful. Therefore they refer to something that may be distinguished from causes, objects and behaviour.

Against this, several things could be said. First, of course, that it is not the case that meaningful statements necessarily are also true statements. Second: What could “a prototype for the subjective feeling itself” possibly be? And how do J-L&O know that “members of a culture” don’t have such? Is it because no such prototype is reported? Should it be reported it would of course have to be linguistically. But no linguistic description can report “the feeling itself” or its prototype. But that aside; it could also be the case that the terms we use to describe our own condition confuse phenomena that it would be rewarding to keep apart. Criticising J-L & O, Ortony and Clore claim that “... the linguistic test of the awareness criterion may perhaps depend upon judges tacitly giving the candidate term a mood reading. “ (Ortony & Clore, 1989). They argue that it is important to distinguish between mental conditions that should be termed “moods” and those that should be termed “emotions”. In their interpretation, moods are part of emotions, but emotions also include more: “... (I)n the normal course of events, the appraisal of the antecedent condition is causally responsible for the psychological consequences, which together lead to an inclination to act (or not to act) in a certain way. Although, together

these components cause the experienced *feeling* of an emotion, the *feeling* of an emotion can be talked about independently of the appraisal of the antecedent conditions and the dispositional consequences. In other words the experienced feeling of an emotion is *part* of the emotion, but not the emotion itself. The emotion is the whole package, of which the feeling is a necessary, but not sufficient component.” (Ibid.)

The important question here, as I see it, is not whether or not there are pure affective states, but whether or not there are more developed mental phenomena that essentially consists of cognitive as well as affective components, and that should be regarded as sequences rather than instant states. I believe that we should think of the distinctive human emotions as instances of the second sort, and I think it convenient to reserve a special term, “emotions”, to designate them. Emotions are, according to this opinion, not separate affects that are connected to certain contents. Affect and content are aspects of the same phenomenon.

There are however other arguments that indicate that emotions are states that are independent of cognitive identification and classification of objects. Zajonc (1980) has argued that the distinction between affects and cognitions are based on different perception-modes and different ways of processing information. He claims that affective responses work relatively independent of cognitive identifications. The following (presumed empirical facts) represents for him arguments against the assumption that emotions contain an element of object-identification:

1. Affective reactions are inescapable.
2. Affective judgements tend to be irrevocable.
3. Affective reactions are difficult to verbalize.
4. Affective reactions need not depend on cognition.

5. Affective reactions may become separate from content.

Zajonc assumes that there are two separate systems processing affective and cognitive relevant information. The first system operates quicker and is ontogenetic and phylogenetic more basic or primary. Zajonc suggests that the ability to respond emotionally to stimuli goes deeper than the ability to discriminate cognitively between different properties of objects. He therefore introduces a distinction between *diskriminanda* and *preferenda* and ascribes emotional responses to our perception of entities of the last group. Of this he says: "I cannot be very specific about *preferenda*. If they exist they must be constituted of interactions between some gross object features and internal states of the individual-states that can be altered while the object remains unchanged, as for example, when liking for a stimulus increases with its repeated experience." (Ibid.) Such *preferenda* are, according to the theory, conceived temporarily earlier than *diskriminanda*, and they are relatively resistant to cognitive or rational correction.

There might be some sound intuition at the bottom of Zajonc's conception. It is however blurred partly by a vague terminology and partly by his tendency to regard perception of discriminate properties of objects as a sort of value-free registration of objective properties. He also tends to regard an emotion as a momentary state rather than a sequence of events. I believe that we will be better equipped to grasp the characteristics of human emotions if we think of them as typical series of events including typical antecedents, immediate responses, the possible correction of these responses and the typical dispositions for action. I do not deny that there are important elements in our emotional responses that we share with other mammals. Zajonc's theory might work with

regard to these elements. But if our concern is to understand the *human* emotions, into which these elements are integrated, I think it leaves out important aspects. As I have said above, I believe that human emotions have cognitive as well as affective aspects. Zajonc's theory grasps, at best, the last aspect.

There is also in my opinion a very serious weakness shared by the approaches of J.L. & O. and Zajonc. They both seem to take our ability to make linguistic discriminations as an indicator of our ability to make discriminations in general. The fact is, however, as I have already remarked a couple of times, that our ability to discriminate between sensory nuances is far superior to our ability to express such nuances linguistically. The seeming separation of judgement and content and the seeming resistance to correction of some judgements may therefore be due to the inadequacy of language to express differences that nevertheless are consciously felt.

In addition: Some of the theses listed above are in fact hard to understand. For instance, what is meant by "judgement" in 2? If it means what it usually does, an articulated ascription of a property to an object, it is simply untrue. We have all changed our minds with regard to evaluation of people and objects. How does 2 go together with 3? 3 seems to imply that it is difficult to pass affective judgements at all.

To put it very simple: Zajonc assumes that we have two separate systems. There is the intellectual or rational system that, among other things, deals with identification of properties, which, he believes, necessarily takes place in a propositional that is language-like fashion. There is, in addition to this, an affective system that determines our preferences and guides our choices. He attacks a model that pictures the process of

emotional reaction as a sequence starting with a cognitive identification of things to which we later respond with some sort of affect. Zajonc can be, and has in fact been, criticised for having a too intellectualist conception of the mental activity that identifies properties. In a reply to Zajonc, Richard Lazarus (1982) claims that Zajonc is founding his model on an inadequate conception of elementary cognitive processes. According to Lazarus, Zajonc equates cognitive processes and rationality. Lazarus' own theory is that cognitions of a certain type, the one's he calls *appraisals*, are basic in any response to events in our surroundings. With "appraisals" is meant immediate evaluative perceptions. "Immediate", because they do not follow as a result of intellectual reflection or elaboration of perceptions given independent of such reflection. The standard of evaluation is the well being of the organism. Put differently, it is simply wrong that cognitions make their first entrance as value-free recordings. In Lazarus' theory; "(E)valuations or cognitive appraisal also begin at the start ... (E)motion or feeling is never totally independent of cognition, even when the emotional response is instantaneous or nonreflective, as emphasized in Arnold's (1960) use of the term *appraisal*. This is the real import of the expression "hot cognition". The thought and feeling are simultaneous." (Lazarus, Ibid). Lazarus' position is that emotions have cognitive and affective components, but not in the way that these exist as separate systems that are later integrated. According to him, our cognitive discrimination of qualities and properties expresses evaluations from the very start. These aspects may later be dissociated. But such dissociations are results of special circumstances and special processes, and they should be explained as such. It is the dissociation, not the association, that need special explanation: "(I)solation and intellectualization (or detachment) which are aimed at regulating feelings, can create a dissociation between thought and feeling. Moreover, attack occur without anger, and avoidance without fear ... (S)uch separations are less the rule and more often a product of coping under special

circumstances.” (Lazarus, 1982). This approach conceives of the dissociation as the problem that needs explanation. Special cultural, institutional and other inter-human situations would be circumstances that could explain the sort of dissociation in question.

I will draw two general conclusions from the foregoing discussion. The first concerns the, already mentioned, use of semantic entities like concepts and judgements as indicators of emotional, and more general mental states. If we consider the fact that our emotionally relevant neural networks are capable of far more discriminations than our emotion-vocabulary, it would seem rather futile to use the last to determine the first. This argument also has some consequences for the way we should conceive of the relationship between cognitive discriminations and factual statements. I can discriminate between the various shades of red on the carpet in the room I am now sitting in, but I am unable to describe them to you in a way that I myself, and far less any other, would be able to recognise as an accurate description.

The second is simply that I will take Lazarus’ as a support for my own assumptions that evaluative responses always are about something and that the evaluative feeling and the discrimination of this “something” makes up one and the same mental state.

Chapter 4

The formal Structure of Human Emotions as they relate to Rationality

I am now going to look at emotions and their composite nature as they present themselves at a higher level of conscious awareness than we meet when we approach them from the mapping-perspective. The formal structure of emotions is discovered through a kind of phenomenological awareness of our own mental states. It is the (top-down) description given by rational beings of their own emotional states.

The analysis of the formal structure of human emotions has been a subject of philosophical investigation at least since Aristotle. Indeed, Aristotle may be regarded as the founder of the analytical approach in the study of human emotions. By “an analytical approach” I mean an analysis of the formal structure that makes emotions into meaningful entities, of their basic grammar, so to say. This implies that an analytical investigation, in this sense, primarily applies to emotions of organisms that also interpret the meaning or significance of events.

It is foremost in the *Rhetorics* that Aristotle undertakes this kind of approach to emotions. Such is his definition of anger: “Let us then define anger as a longing, accompanied by pain, for a real or apparent revenge for a real or apparent slight affecting a man himself or one of his friends, when such a slight is undeserved” (*Rhet.* II. I. 2)

Definiens lists a series of properties that must be present for something to be counted as anger. The emotion must motivate a certain kind of action, a certain quality of affection must attend the emotion and the object of the emotion must have certain characteristics. First, it must be human; an emotion directed towards bad weather or other natural phenomena does not count as anger. The human object must be a particular. An emotion towards a collective is not anger. The emotion must be directed towards some person because of something this person has said or done. This “something” must be conceived as an expression of slight, and the slight must be considered illegitimate. Anger can thus be seen as expressing a set of opinions and judgements, and these may be more or less correct. You may be wrong about what people have done, about who has in fact performed a certain act. Something may also be construed as slightful that was not intended that way. The emotion in question may thus be considered more or less rational according to the correctness of the opinions that are underlying it.

Aristotle’s definition is part of the *Rhetoric*. The aim of Rhetorics is decision, and as emotions “are all those affections which cause men to change their opinion in regard to their judgements ...” (*Rhet.* II i), it becomes important to understand what it is that determine emotions and emotional variations. This context is worth noticing for two reasons. First, because it could imply that there is a connection between the rationality of emotions and the fact that emotions can be manipulated. If emotions can be manipulated, they can be changed, and if they can be changed, there is at least a possibility that they can be changed on the initiative of the person who has the emotion. If the person who has the emotions, can change them, there is a chance that this change could be the result of consideration according to some rationality standard, e.g. an adjustment to external facts or

an alternative interpretation of such. Second, Aristotle's approach to the emotions in the *Rhetoric* instantiates what might be called "practical", it investigates emotions as they relate to human action.

Aristotle's approach seems to imply that emotions are structured according to some categorical formal patterns. And in fact he says something that may be interpreted that way: "And each of them (the emotions) must be divided under three heads; for instance, in regard to anger, the disposition of mind which makes men angry, the persons with whom they are usually angry, and the occasions which gives rise to anger." (Ibid. II.B.9) In this brief statement we can reconstruct three categorical forms; mental antecedent, object and situation, which indicates that emotions may have some general formal grammar or structural components.

If emotions are organised according to formal categories it might be possible to reconstruct or explicate them, or parts of them, as propositional judgements. Robert Solomon is a contemporary writer who has developed this idea. He regards emotions as "constitutive judgements", judgements that constitute a "surreality in terms of values and self-conceptions" (Solomon, 1993 p.194). An implication of this approach could be that an investigation of the rationality of emotions should not be limited to investigating whether or not such judgements are true and adequate. It will also concern the conditions of the possibility for such surreality. Inspired by Kant, Solomon calls the object of his investigation *logic of emotions*; "I am employing the term "logic" as Kant used the term, to signify the employment of categories and concepts." (Ibid p. 194). Solomon lists thirteen such categorial emotion-concepts: Direction, Reciprocity, Power, Evaluations, Scope and Focus, Intersubjectivity, Strategies, Criteria, The nature of the object,

Distance, Desires/Intentions and Personal status. The last category is called “Mythologies: The synthesis of our Emotional Judgements”. This has a special status to which I will return. For our present purpose it is not necessary to comment separately on the different categories. Neither do I want to take a stand on the question whether or not emotions are underlying categories in such strict a sense as is often associated with the Kantian ones. In fact I believe that Solomon ties his analytical tools too close up to a linguistic conception of emotions. I have few problems with that as long as we speak of certain highly developed social emotions, experienced by sophisticated language-users. The main point of interest for me at this point is however that Solomon’s scheme tells us that there are certain things we should look for when we want to understand emotions, and he gives a suggestion about what those things should be. He tells us that we will probably understand something significant about an emotion if we consider various aspects of its object-directedness, or if we ask what sort of relation between the self and others that is implied in the emotion. Love implies another sort of reciprocity than does contempt. It also implies another conception of the distance between oneself and the other. Other emotions correspondingly have certain strategies of action that are adequate expressions of the emotion. Certain strategies are compatible with hatred, but not with friendship. Every emotion also expresses certain evaluations and is at odds with others. Such evaluations imply certain forms of arrangement of oneself and the others within a value-hierarchy.

There is no need to enter further into each category. I will merely point out three main consequences of Solomon’s scheme. One concerns the question about what sorts of entities one should look for in an analytical investigation of emotions. Emotional changes and differences will, according to the scheme, be variations within the basic concepts regarded

as variables. The second, which is implied in ascribing the basic concepts constitutive status, is that entities like “self”, “objects that the self is engaged in or attached to”, “action” and “value”, are entities that are elements in emotional complexes. Given that this is true, one might ask if this also means that such entities have no existence separate from their function in the formation of emotions. Third, the categories could be thought of as a set of mapping-systems, co-ordinates, regulating the interaction between individuals and their environment. Such co-ordinates are vital in this interaction and the emotions are ways we feel or sense them.

As I have already mentioned, I believe that Solomon is wrong in thinking that emotions in general have a linguistic origin, that they *are* judgements. I do not deny that there are emotions that could be analysed in linguistic terms, but we are then talking about relatively developed and sophisticated emotions of human beings who are already competent members of a language-using society. We are talking, not of the origin of emotions, but of emotions reconstructed and measured against other aspects of human capacities and social scenes on which they are displayed. On the other hand, if we take a different perspective, we could think of the entities that Solomon conceives as categories as a kind of vectors, and emotional states as results of what Paul Churchland has termed “vector coding”. This would allow us to see a similarity between emotions and other kinds of mental phenomena like perception of colours, taste, identification of faces etc (Cf. Churchland, 1996). It would, in short, give us a conceptual tool that could help us to understand how we discriminate between different emotional states at a pre-linguistic level. When we find ourselves in an emotionally relevant situation, there must be some features of the situation that determines it as emotionally relevant. Put in another way, we must process the situation along certain lines that construe it as emotionally relevant. I find it very plausible

that this is in fact part of what is going on when we respond emotionally to certain situations. Conceived this way emotions emerge as a kind of multi-dimensional mapping system, with as many dimensions as there are “categories”. Together the dimensions could be represented as constructing a space wherein any possible emotional state has its place. The “space” would be a multi-dimensional “emotion-space” analogous to the colour-spave described in chapter 2. What is being mapped are social situations (and for some basic feelings like fear, other kinds of situations) and their importance to the emotist. To allow for enthusiasm for a moment, I actually think it is a wonderful idea to create a map of possible social situations, a map built on the vectors that determine our responses to such situations. It would indeed provide an important tool for social-psychological analysis.

One component that must be identifiable for an analytical approach is the *object of the emotion*. If we are to assess the rationality of an emotion, it would seem necessary, in one way or other, to consider its object. But the concept of the object of emotion is not univocal. It could refer to “the content of the emotion”, “the cause of the emotion”, “that towards which the emotion is directed” and so on. Ronald deSousa (deSousa, 1987. ch.5) has suggested as many as seven different meanings, all being relevant for an assessment of the rationality of emotions. These are: 1. Target 2. Focus, focal property 3. Cause 4. Motivational aspect 5. Aim 6. Proposition 7. Formal object.

This enables us to ask seven different questions (which presumes seven answers with different levels of specificity) in a situation where an emotion is to be assessed. If someone were angry such questions would be:

1. With whom is he angry?
2. What is it about this person that makes him angry?

3. What incident is it that enangers him?
4. What is it about this incident that motivates him?
5. To what sort of action does it motivate him?
 6. What true proposition describes the situation that enangers him?
 7. What is the criterion that this emotion is successful?

The concept of formal object and the corresponding question need some explanation. The formal object of envy would be “something desirable but not possessed”. According to Aristotle’s definition of anger, “undeserved slight” would be the formal object of this emotion. In fact there are as many formal objects as there are types of emotions. To use the words of deSousa: “An emotion means a formal object, that is, a property characteristic of a paradigm scenario and ascribes it to an object” (Ibid. p. 242).

The concept of formal object illustrates a connection between emotions and speech acts. Searle remarks that there is a resemblance between speech acts and intentional states consisting in that they both represent the conditions of their success (or conditions of satisfaction). Searle remarks that it is “... crucially important to see that for each speech act that has a direction or fit the speech act will be satisfied if and only if the expressed psychological state is satisfied, and the conditions of satisfaction of speech act and expressed psychological state is identical (Searle, 1983 p. 10-11).

It should be added that an emotion might be rational even if its conditions of satisfaction are not satisfied. But it cannot be rational if it doesn’t have any such conditions at all. And it is often not rational if it remains uncorrected when the person who has the emotion realises that the conditions are not satisfied. Now, one might ask if such emotions without

possible conditions of satisfaction are themselves possible, given that emotions are psychological states and that such states are characterised by their intentionality. I think they are if we understand “conditions of satisfaction” as “possible conditions of satisfaction”. In fact I believe that much of the pain that emotions can inflict upon us is due to some intrinsic contradiction in their conditions of satisfaction. Twenty years ago, in a Norwegian parent-run kindergarten, I observed a decoration on a wall. Under a painting it was written: “Don’t do what your mother tells you to do!” Let’s assume this to be written by the mothers. What should the kids do to act according to this message? Let us then assume that this message was transformed to be a part of the children’s emotional attitudes towards their mothers. What would be the conditions of satisfaction for such an emotion?

Formal object could also be called second-order objects. This term is meant to signify that they are the properties of first-order objects that determine the type of emotion that has this object. If I am afraid when someone threatens me with a gun, the threatening person with his gun is a first-order object, while danger is the second-order object.

Our ability to distinguish between first and second-order objects is obviously a function of our generalising capacity and, as such, probably distinctively human. Without this there are reasons to believe that we would have a much simpler way of responding to certain objects. To some objects we still have this direct, not generalised, response-system. LeDoux shows that we respond with fear-reactions towards some objects much faster than we are able to identify them cognitively. Nevertheless, we do in fact possess a generalising ability giving us more possible histories for a single emotion-reaction.

The distinction between first and second-order objects also gives us a way to understand why the same object can cause different types of emotions in different people and also different emotions in the same person in different situations and against different backgrounds. And because a single object can carry different second-order objects for one and the same person, it can contribute to our understanding of emotional ambivalence.

The concept (if not the term) of second-order objects is not a modern invention. In the philosophy of Plato it plays a leading part. A passage from the *Symposium* deals with the significance of the distinction between first and second-order objects for our emotional development. Socrates recapitulates a conversation he has had with Diotima from Mantinea. Diotima describes the education of our ability to love as a process in which this emotion changes direction. It is awakened as directed towards one single object. Steadily it grows more independent of the particularity of the object, to end up being directed towards “the formal object of love” which, according to Plato, is “the good” (Sym. 205 D). Plato always regards the first-order object of emotions (as well as cognitions) as copies of the things towards which the soul is really directed, i.e. the ideas.

The distinction between first and second-order objects is an important one. The line between them should however not be drawn so sharp as to blur differences between certain emotional nuances. One might well ask if love for one’s children is the same emotion as love for a spouse. Do these emotions have the same formal object, hence being the same type of emotion? Is pride of a successful bank robbery the same emotion as pride of a good university degree? I think not! Perhaps we should say that it is two different species of pride. Another problematic example: In chapter 1 I said that fear instantiates the mapping of certain features (danger-points) in our natural surroundings. But, as we know, fear can

also be triggered by certain features in our social surroundings, like fear of making a fool of yourself, fear of the condescending looks of other people etc. I do however find it reasonable to consider such instances of fear as a kind of second-order emotions that is related to another (first-order) emotion, in this case shame. The formal object of this kind of fear would then be shame, the first-order emotion would be the second-order object of a second-order emotion. The general lesson we might learn from this is not that we should abandon the distinction between first and second-order motives. It is rather that the amount of formal emotional objects, and therefor the emotional repertoire within a given culture, varies, among other things, with the amount of different first-order object-types in the same culture. To put it plainly, the greater differentiation within a culture's amount of possible first-order emotion-objects, the more fine-graded nuances of emotional experience within the same culture. This will have some consequences for what I will have to say later on.

At least in one interpretation there is a connection between something being rational and the possibility to ask questions about this something. What is rational is open to questioning as a test to its rationality. Asked in connection with emotions, questions may reveal that an emotion is founded in some sort of mistake or failure, something has "gone wrong" in some way or other. I believe that we should distinguish between two different kinds of possible wrong-goings. The first is based on getting the facts wrong. The person with whom I am angry has not done the things that make me angry. Perhaps these things have not been done at all. I may misconstrue someone's actions as say, insulting. That type of wrong going has to do with the truth-value of the propositional content expressed in the emotion.

The second type has to do with the reasonableness of the emotional reaction and its expression. Let's say I get things right. One may still ask whether or not these things give me a good reason to respond emotionally as I do. There are nuances in this problem-field. One may focus on a totally unimportant aspect of someone's behaviour. One may ask if a certain type of event is a good reason for a certain type of emotion (or emotional expression). Or one can be more specific and ask if it gives *me* a good reason for such an emotion. Is poverty a good reason for shame? Would it be a good reason for me to be shameful? Is beating Brazil in soccer a good reason to be proud? Does the fact that Norway has beaten Brazil twice give *me* a reason to be proud? After all I am not part of the team. To distinguish these kinds of questions from the first (that concerns truth-value) I will say that these are about the *relevance* of the emotion. Obviously the relevance-question (in contrast to the question of truth-value strictly speaking) can only be raised within a framework of established and stabilised social values and codes of emotional exhibition or exposure.

Most people probably agree that we may go emotionally wrong in (at least) these two ways – truth-value and relevance. And most people would agree that we are best off avoiding such failures. From these assumed facts, I believe that we can learn something about some properties of emotions that are specific to language-users, that is humans.

LeDoux has an intriguing example of a seemingly wrong reaction that is in fact the “right” one. You walk down a forest-path and suddenly you see something crooked on the ground before you. Spontaneously your brain initiates a fear-reaction that prepares your organism for fight or flight. A moment later you discover that it is not a snake, but a crooked branch. Let's say you make this failure nine times out of ten. Is it not irrational to waste energy in

this way? It is neither rational nor irrational, but it surely is a good way to survive. LeDoux takes this to be an evolutionary old heritage that we share with a lot of other species. This sort of reaction is useful, today probably more so in connection with car driving than with snake-detecting. As a little aside, this shows us something about the human ability to generalise reaction-patterns. Humans are presumably not genetically coded to detect potential dangers before they identify them, as they move along a highway at a speed of 150 km/h. Nevertheless, experienced car-drivers do just this.

But automated stimulus-response patterns like this do not exhaust our emotional repertoire. When it comes to emotions carried by symbolic structures, the situation is different. Such emotions are part of a symbolically structured human interaction-system, and they are based on the fact that such systems work, not only spontaneously, but over time. They tie together various parts of interaction sequences and they are based upon the relative stability of values and interpretation-systems. In fact, many of them are based upon the fact that certain interaction-systems are institutionalised. And the value of institutions, as far as they have any, is always based on their relative stability, their ability to uphold a formal framework over a period of time. Spontaneous, as opposed to reflective emotional responses are therefore often dysfunctional to such systems. I believe that it is vitally important not to think of this as something foreign to human emotions. It is rather part of what is specific to these.

Aristotle is perhaps the first to see relevance as a parameter in the investigation of emotions. For him relevance has a moral significance. In the *Nichomacian Ethic* he tells us that: “It is possible to feel fear, confidence, desire, wrath and pity, generally to feel pleasure and pain too much or too little, which is, in each case, not good. But to feel these

things when one ought too, towards the right people, from the right reasons and in the way one ought to, that is the middle and the best and this is exactly what is virtue.”

According to Aristotle, emotions are tied up with a set of descriptive and normative considerations. If these are wrong in some way or other, the emotion is inappropriate, and expresses some weakness of character. Such failures may among other things consist in ascribing importance to something insignificant or to expose too strong or too weak emotions in a given situation.

In an attempt to elaborate the concept of relevance applied to emotions we may utilise David Hume’s distinction in the *Treatise* between direct and indirect passions. Passions as sorrow, fear, joy, hope, attraction and aversion are direct. The prototypes of indirect passions are pride, humility, love and hatred. Such passions may also be named “social passions”, because they occur within a relatively stable system of values, interaction-systems and ascription of relationships. If it were at all meaningful to ask about relevance for the direct passions it would be about the probability of consequences of the cause of the passion. Is the object causing fear really dangerous? Is the object causing hope really likely to lead to the desired consequence?

As for the indirect passions the question of relevance can be raised in another way. To show this, it is necessary first to sketch the meaning of “indirect” as applied to these passions. These passions are not caused by the presence of an object or a property alone. In addition, the object must stand in a certain form of relation to a person, a self, mine, or someone else’s. These passions are cognitive and performative attitudes towards people and they are caused by something people have or do. The question of relevance can be

raised in two ways. First it can concern whether or not the property or quality really is a good reason for the emotion. Is the quality/property an adequate formal object for the emotion in question? It can also be posed as a question whether or not the relation between the object and the person is such as to rationalise the emotion. If someone is proud because his child have passed an exam with excellent marks, it would of course be possible to ask if he has a good reason to be proud. Why should he be proud over someone else's achievement? Maybe he would answer that he is proud to *have* a child who has done this. Be that as it may – the fact that it is meaningful to ask questions like this, and the fact that the answers give a basis for intellectual reflection, shows the rationality-relatedness of these emotions. Emotional reactions and expressions may be tested and assessed against standards of truth-value and relevance, cognitive correctness and norms regulating personal expressions.

Given that emotions combine different elements or variables, one might ask how, or according to what principles, if any, these elements are combined. One possible answer would be that they are combined according to certain rules comparable to the grammatical rules that are often taken to operate when we form intelligible sentences. I believe that an answer based on such an assumption would be wrong at least if it is intended as a theory on how emotions generally come to be felt as unified mental phenomena. It may be right if applied to some highly sophisticated and developed emotional nuances experienced by people who are already capable of analysing emotions and manipulating their different components, like language-users who can use the formal grammar of their language to create fine-graded distinctions of meaning. Such distinction-making on the basis of formal analysis certainly has an important application in the field of justice which requires a formal competence similar to that of grammarians. But at a more basic level I believe that

emotionally relevant elements are rather combined in imagined proto-typical situations or paradigm scenarios. I will however return to this later.

The basic point that I want to bring out of this analytical approach to the emotions is that the picture of emotions as being accessible to analysis into different components, yields a picture of emotions as a sort of situation-images or, better, prototype pictures of situation-types. These situation-images are experienced in certain qualitative modes, namely those pleasant and uncomfortable modes characteristic of the experience of different emotions. I will later approach the concept of “situations” more directly. For the present it is enough to bear in mind that we have shown that emotions contain elements that can also be thought of as building-bricks or constitutive elements of complex situations

Emotions and Motives

What is the relationship between emotions and the complex will/motives/actions? My preliminary answer will be that one aspect of the development of human emotions is that the originally diffuse affections that represent the origins of emotions are consciously sensed as directing attention to something in a certain performative way. One important addition seems necessary: This does neither mean that an emotion determines the empirical performance of an action, nor the strategies employed or the way the action is carried out. It means, at most, that it determines the *kind* of action. In cases of developed emotional lives, I take this to mean that it determines what I will call the formal motive under whose perspective the action can be properly described and understood.

The analytical conceptual tools we have dealt with so far seem to imply that there is some kind of logical relationship in the sense that some sort of disposition to act and some sort

of aim-direction is part of what it is to have an emotion. This may be correct as far as the already developed human emotions are concerned and as far the question of logical relationship is an interesting question here at all. If it is, it must be because we have already established an emotional web, or emotional webs in certain ways and have already started to conceive of these webs as conventionally established patterns of action. “Logical” would then mean that we conceive of the meaning of an emotion in light of its place in such a web. Actually I believe that it is a bit beyond the point, at least if it means that the questions about the relationship between emotions and motives can be answered a priori, that is independent of empirical investigations. But before we proceed from the assumption that emotions and motives are part of some kind of webs, it is worth mentioning an alternative opinion, i.e. of David Hume. According to Hume motives are contingently connected with emotions. In his discussion of, on the one hand, love and hatred and on the other, desire and aversion, he claims that love does not necessarily imply a desire for the well being of the loved one, even if it normally produces such a desire. The reason he gives is that “... these desires arise only upon the idea of the happiness or misery of our friend or enemy being presented by the imagination.” (Hume, 1978, p. 367) Passions are, according to Hume, impressions, but the desired state can only be represented as an idea. The relation between them can therefore only be one of association, not an essential one. The bottom line of these considerations is that emotions *cause* motives and actions and cannot therefore entail them logically. It would probably sound rather odd to say that the cause of John being unmarried is that he is a bachelor. Anyhow, Hume’s discussion is within the framework of asking if a quality is implied in, or caused by, something else, which means that it tends to be a discussion of emotions as static units (particular impressions) rather than about dynamic processes. The sound intuition at the bottom of Hume’s theory is that no single fact as such logically implies the existence of some other

positive fact. Descriptions may entail that other descriptions are true (or untrue). Facts don't have that quality. The impressions which are emotions are, in their origin, taken as pre-linguistic facts.

This conceptual framework has been criticised by the American social-psychologist and philosopher George Herbert Mead (Mead: 1934). According to Mead, Hume and his followers hold a conception of mental phenomena as something substantial and static. It starts from a given experience of pain or pleasure that is associated with certain things, events or the like. In contrast to this, Mead develops a theory of consciousness and mental phenomena as something dynamic. According to this theory a mental phenomenon is part of a complex social act, in the sense that it is made and shaped within the context of this act. The mental phenomenon is an anticipation, or an opening of, a social act. This act should be understood within a sequence of acts in the way that it aims at calling forth a response in an addressee to be adjusted as a result of such a response. Social actions go through a process of mutual adjustment and adaption and should be understood as parts of such a process. The consequence of this theory of mental processes for a theory of emotions will be that emotions are mental states being given certain expressions that should be conceived as introductions to social acts. This implies that emotions are shaped within frameworks of sequences represented in the imagination. They are parts of mentally represented "stories".

As indicated above, one way to proceed in the understanding of the relationship between emotions and motives might be through a distinction between first and second-order motives (formal motives). If we take this distinction as given, we obviously start with emotists with a high degree of generalising capacity. Given that someone is angry with

someone, and desires to avenge himself. Revenge may be achieved in different ways. He might inflict some physical harm on the other, he might embarrass him in public, he might try to make his affairs difficult, and so on. "Revenge" would be the second-order motive for all alternatives, while the different options would be different first-order motives. Given the emotion in question, there might be different options for first-order motives, but only one possible second-order motive. My suggestion is that the second-order (or formal) motive is part of the emotion's internal structure part of the complex that is an emotional state in a developed human being. That would mean that the emotion could not be given an adequate description without some description of the formal motive. The first-order motive should, on the other hand, be considered as an effect caused by the emotion. Considering the formal motive as part of the emotion also allows us to describe the emotion as a sort of performative attitude, a disposition to perform certain types of acts. That will in turn enable us to utilise the conceptual tools of language-pragmatism. Emotions could be considered under the aspect that John Austin calls "illocutionary" (Austin: 1962), and this could be used partly as one (out of several) differentiativational criteria and partly as a connective between the emotion and the action.

The analogy between an emotion and a speech act becomes especially clear if we consider the illocutionary aspect of an articulated emotion. The articulation of real anger or hatred to the person towards whom these emotions are directed will, with all reasonableness, be taken as a real or potential threat. Taken as something else would probably mean that the articulation was interpreted as another kind of emotion, for instance fear.

The distinction between first and second order motives also relates to the classic distinction between understanding and explaining human action. To explain something we have to

understand what this something is. To explain the action of some human agent, we have to understand what *kind* of action she is performing. I might see that someone is cutting wood, but that doesn't necessarily imply that I understand that she is really helping her grandmother, performing an act of felt obligation or gratitude. I might hear that someone is misinforming someone else, but that doesn't imply that I also understand that she is performing an act of, say jealousy. So, it seems that we would need some notion of a formal motive to understand what kind of action that is performed. But understanding this in light of such a notion is also to understand the emotion motivating the agent. Once this is understood, we are in the position to assume some causal chain between the formal motive and the first order motive as it is carried out in action under the specific circumstances at hand. Such an assumption would then be an attempted explanation of the action.

As to the connective function, the clue has already been given. The formal motive is an essential part not only of the description of the emotion, but also of the action. A certain action may be described as an act of revenge that partly makes it intelligible and partly implies that it is motivated by, say hatred. An action may be taken as hostile or friendly, which means that it is construed in the light of different formal motives and correspondingly as expression of different sorts of emotions. If someone's action toward someone else were taken to be an act of revenge and it later turns out that the actor wasn't at all angry, then we would have reasons to suspect that the action was given a wrong description. Admittedly, one might perform an act of revenge without being in a state of anger in the same way that one might continue to avoid dangers after one have stopped fearing them. But even if the *state* of anger is absent, one must still have the *disposition* to perform acts originally connected with anger. And without some reference even to the

relevant emotional disposition it would be hard to maintain a conception of a certain action as, say, one of revenge. To summarise: The hypothesis of an internal relation between the emotion and the formal motive is based on the assumption that an emotion is, among other things, a certain performative attitude, and that the action adequately expressing the emotion is intelligible as a realisation of, and caused by, the formal motive that corresponds to that emotion.

I have said that the formal motive is part of the internal structure of the emotion. To be more precise; it is part of the emotion's intentional structure. Achieving a goal or state of a certain type is part of what makes up the condition of satisfaction of the emotion. That would sometimes, but not always, mean that part of the intentional structure of an emotion is the state that would bring the emotion to an end. Shame would disappear when honour is re-established, fear when danger passes. On the other hand, the intentional structure of symbolically based emotions often does not contain specific ways in which this should be achieved (at least in non-ritualistic cases). That is why first-order motives usually are not parts, but effects of emotions.

The distinction between first- and second-order motives also makes us see that a consciousness with second-order motives as part of its intentional structure is a consciousness that has alternative options for action, a consciousness that is not restricted to respond along rigid stereotyped patterns.

I shall now try to argue that the first-order motive of an action should in fact be regarded as an effect caused by the emotion. A description of an action in light of a first-order motive would be a description at a lower level of generality. Descriptions like, "he shot him", "he

ruined his business” or “he gave him false information”, are to be considered less generalised than “he avenged himself”.

First-order motives are caused by second order motives applied under the conditions of certain beliefs that make the causal conditions under which the second order motive causes the action in question.

But how can a non-physical entity like a motive cause anything at all? First of all, if we consider that what we are talking about is a (human) organism in a certain state of performative attitude, it is not all that non-physical. A formal motive thought in abstraction from this would of course not be capable of causing anything. Thus comprehended it would however be hard, not only to understand how it could cause anything, but also how it could exist at all.

What is vitally important here is to understand that the logic of emotions is such as to make emotions properties of something, that this something is an organism and, in the case of symbolically founded emotions it is a human organism. Such organisms do in fact feel and reason and this fact is part of the explanation why they behave as they do.

The desire to harm someone as revenge is an intentional desire to harm this person. This intention implies a desire to be the one who causes the harm, and probably often also that the offender knows who caused the harm. So, to start with, causation is part of the intention. One reason why we are able to form such intentions is that we have experienced that we are able to cause something by acting in certain ways. In this case the intention is in fact an intention to cause something.

I assume that there is an important (not only semantic) difference between the following propositions:

1. He hit because he wanted revenge.
2. He wanted revenge because he was angry.

1 represents a causal relation. 2 does not. 1 represents a temporal sequence. 2 does not. 2 describes a pattern of mental states where one part is defined in terms of the other, anger in terms of revenge and vice versa.

Finally: I take these considerations on first and second order motives to bear some impact on a classic problem in anthropology; how can we understand other cultures? Maybe the problem is to understand what kinds of second order motives it is that causes the first order motives that we can see followed in action. If so, that wouldn't imply that the second order motives, once understood would be unrecognisable. It would simply mean that we know too little about the circumstances (cultural and/or natural) to understand the causal chains between the second and the first order motives.

There is of course also another alternative source of the problem. The interpersonal relations in another culture might be so different from our own, and for that reason give rise to emotions (and hence formal motives) unrecognisable for us. If what I have said above is correct, that would imply that there would be bad prospects for our understanding of action in such a culture.

The Shaping of Social Emotions

Most writers on emotions make a distinction between basic or in some sense fundamental emotions and other emotions. Physiological evidence supports such a distinction (Cf. Damasio and LeDoux). I see no reason not to accept that there are some emotions, emotional reactions and ways of feeling affected that are more basic than other. We don't necessarily have to draw a sharp line between basic and non-basic emotions. We could rather imagine a scale where we, on the one side have emotional reactions that occur independent of reflective cognitive and evaluative judgements based on symbolic code-systems and, on the other, emotional reactions where such judgements play an essential role. The formation of basic emotions may perhaps be regarded as an instance of the formation of basic-level concepts in general (Cf. Lakoff & Johnson, 1999. p. 29). Basic-level concepts are concept on the medium level in a three-level categorial hierarchy. "Chair" would be an instance of such a concept, placed between "furniture" and "rocking-chair". Correspondingly "anger" could be said to hold a level between "indignation" and "hostility". We hit some kind of median before we specify or generalise. Because fine-graded emotional distinctions often have references to social rather than natural properties of things and events, it seems reasonable to assume that the higher degree of emotional nuances, the higher the dependence on symbolic code-systems. It might be tempting to talk of such nuances as an emotional functional differentiation on the intra-personal level. The reason is that emotional nuances of the sort in question are highly specialised to fit into special sorts of social situations. Consider for instance embarrassment, shame, indignation, insult and so on. These terms could be predicated, not only of mental states, but also of certain features of social situations based on complex symbolic code-systems. Described in another way, they can be said to represent means for the individual's adaption to the *social* world. I will call such emotions "social" because they typically are caused by situations

whose elements are signifiers of some social value or code. These emotions typically occur within contexts of “social realities” in the sense used by Searle. A social reality, in this sense, is something that is there because people believe and accept certain things about them. Such belief and acceptance include three elements, assigning of a function, that the object is part of some co-operative system (collective intentionality) and status function, something performing a function in virtue of being ascribed a certain status.

There might, however, be a problem here. What I have said about the possible functional role of social emotions seems to imply that the individual identifies these emotions in terms that could also be applied to social situations. But if that is so, it might be hard to understand the origin of the mental quale that seems to distinguish one emotional nuance from another. The feeling of being embarrassed can certainly not be predicated of social situations. A possible approach to this problem might go like this: Emotional qualia might be regarded as ways in which the special significance of social situations or events, and the subjects position relative to other elements in these situations/events, are detected and sensed by the individual. That would be analogous to a claim that fear is the way danger is felt or sensed. But even if this would work for fear/danger, it is not obvious that it would do the same for, say embarrassment/a situation where one misunderstands some social code. The reason is that the first would have a more straightforward relation to our evolutionary history. It seems that we would have to say that, not only are social codes and culturally determined ways to describe social situations learned, the mental qualia that characterises the adequate response to the various situations are also acquired in the same learning-process. Is that plausible? I believe it is! After all there are many sensory nuances that are acquired through some kind of “cultivation”. Consider the ability to distinguish the taste of red wine made from Cabernet, Merlot or Pinot Noir, or even the ability to taste, which are

best on different occasions. The same phenomenon could also be instantiated with reference to food and special occasions. It wouldn't be all that exceptional if it should turn out that the spectre of our feelings undergoes a similar process and that what we learn to conceive of as certain emotional qualia really are subjective experiences of significant features of social situations and our position within such. Maybe Jean Jacques Rousseau had this in mind when he said that culture teaches us to mistake the cultural qualities of things and needs for natural ones. Anyhow, the quale that typically accompanies an emotional nuance would serve the locative function (the you-are-here function) with which I have dealt earlier.

Some emotional sensations, physiological reactions and expressions are given with our biological nature. There are numerous investigations, beginning with Darwin, of facial expressions in certain animals and human infants which clearly indicate that facial expressions of fear, anger and joy are genetically determined and even very similar in humans and some other animals. The visceral physiological reactions that are connected with these emotions are also common to humans and other mammals (LeDoux, Damasio). But it seems unlikely that *all* human emotional experiences and expressions should be genetically directly determined. Where that to be the case, it would also have to be true, not only that we are social animals, but also that all the nuances of a complex society originate in our genetic heritage. Fear is obviously connected with dangerous situations. If we had not been born with an ability to react adequately on such situations, we would not have been born at all. Natural selection would have eliminated our ancestors long ago. But there are, as mentioned above, other more specialised emotions that cannot be so directly related to our evolutionary history, emotions that could only operate within highly symbolically structured human interaction.

There are two things I want to emphasise in this connection: 1. There are some emotions that can only be identified and described in terms that also apply to social situations or relations. 2. I have said that there is a difference between having an emotion and knowing that one has that emotion. The sort of emotions mentioned in 1. are more closely connected with “knowing” than is the case with more basic emotions. They are, so to say, more intellectual.

As already mentioned, John Searle tells us that the intentionality of any mental state, emotions as well as other perceptions, consists in their having “conditions of satisfaction which are determined by the content of the state.” (Searle, 1983. p. 409.) If I see something as a bear climbing up a tree, this cognition contains the condition of satisfaction that there is in fact a bear in the perceived place, that it is in fact climbing up a tree, and that this state of affairs is the cause of my perception of a bear climbing up a tree. If I feel fear anticipating that it might climb down, this emotion contains the additional conditions of satisfaction that it might in fact climb down and that I would in fact be in danger if it were to do so. This way of thinking would suggest that understanding the development of emotional nuances would require understanding of the development of nuances of intentional states. This would entail learning nuances of “conditions of satisfaction”. The conditions of satisfaction of an emotion are, among other things, that the emotion is an adequate response to a situation correctly conceived. The adequacy of an emotional response often cannot be understood apart from the cultural codes of the society in which it is expressed. One will for instance have to master the unwritten rules of construing social situations and interactions as demeaning, debasing, honourable, virtuous, vicious, elevated, respectful and the guidelines for how to respond to such situations.

Now this would perhaps not imply that such construals represent something real. I might construe something as dangerous that is in fact completely harmless. But there is a difference here, viz.: Something might be dangerous or not, quite independent of my construal, whereas there are other formal objects that are constituted by being construed as respectful, honourable etc. (on the precondition that such construals express some collectively shared value-system or symbolic code, not only individual idiosyncrasies).

Forty years ago Norwood Russel Hanson (Russel Hanson: 1961) formulated the, still quite useful distinctions between three perceptual levels, *seeing*, *seeing as* and *seeing that*. This could be thought of as corresponding to an increasing degree of linguistic structuring of the perception-process. *Seeing as* may be regarded as a perception of objects that implies subsumation under concepts, while *seeing that* may be represented as an introduction to a propositional statement.⁴

Corresponding distinctions may also be utilised in connection with emotion-objects as well as with emotional content. As for the first, the various modes would be similar to the modes of other perceptions with the difference that what one sees *as* or *that* would be value-relevant properties. As for the second, we could distinguish between pure sensations of pleasure or pain, the conception of these sensations *as* pride, contempt, envy etc. and the experience of these states as situations of certain types where certain types of actions are likely to occur, actions that are relevant to the formation of certain motives. The mental activity of conceiving something *as*, could be termed “construing”. If someone responds with respect at the sight of someone performing a self-sacrificing act, he construes this object as, say, virtuous. “Construing” or “seeing as” is the mental activity of determining

the formal object as being of a certain kind. It may well be asked how we learn to do this and how we learn to predict probable actions and form relevant motives in the given situation.

We have all experienced mental states in which we are not clearly aware of, or even confused about the conditions of satisfaction of those states. Such states could be called “diffuse mental states”. We may feel unpleasant without being able to locate the source of this feeling or to form an opinion about what we ought to, or would like to do. In absence of a better option we may symbolise the conditions of satisfaction in a chocolate bar, a beer or something of that sort. Such conditions are, as all parents know, often observable in children. When we grow up we develop a capacity for emotional nuances. How do we do that? How are diffuse feelings of comfort or discomfort systematised into a repertoire of distinctly different emotions with highly specialised conditions of satisfaction?

Modern culture offers a large variety of terms designed to interpret desires and needs, to keep some types of desires apart from other and to verbalise the corresponding feelings of frustration or satisfaction. This state of affairs applies to desires that are experienced as physical as well as those to experienced as emotional. The condition of satisfaction for thirst is something to drink, but the desire to drink wine is not satisfied by beer, neither is the desire to drink red wine satisfied by white wine and so on. This phenomenon can be regarded as an experience of the differentiation or specialisation that is part of modern society and culture. This culture, as is suggested by Bourdieu, actually encourages the development of the ability to make ever more sophisticated distinctions. In fact, the very ability to make fine-graded distinctions is often taken to signify social competence. Of course it is not all that obvious that the ability to make distinctions is actually accompanied

⁴ The fact that Russel Hanson wants to show that there is no pure seeing, is of no importance here.

by a corresponding distinction of mental qualia. But when it is, it would seem more reasonable to regard such felt qualitative distinctions as a result of the first mentioned, rather than the opposite. What I suggest here, is that there are certain cultural codes and symbol-systems to signify hierarchies and labour-division systems that are internalised as felt emotional nuances and that these nuances are quite fine-graded in modern societies. The relation between such emotions and their formal objects is less straightforward than is the case with basic emotions like fear, whose relation to danger is as direct as it can possibly be. Sophisticated social emotions tend to disguise their formal objects and their conditions of satisfaction. They may therefore be regarded as resulting from a sort of double coding, viz. the system of coding given in our neural system and the one that is based on cultural value-hierarchies.

This, or related problems, have been dealt with by classical sociologists and social philosophers, most noteworthy Rousseau and Durkheim. These writers have obviously made great contributions to our understanding of some of the mechanisms of this process. Their contributions mainly concern the modernising process on the level of functional differentiation of labour and, in Rousseau's case, the growth of inter-human dependence. It seems to me that Rousseau's basic intuition here is that the progress of civilisation is accompanied by an increasing symbolisation of inter-human dependence, recognition and rejection and that these changes also have impact on the subjectively felt emotional life of the individuals. I will however leave these contributions aside and comment on another aspect of the human emotions, an aspect that relates them to significant social situations.

Imagining a Drama

I have said that our emotional development partly is determined by our learning the significance of different inter-personal situations. I have also said that our knowledge of the significance of these is acquired, not only as concepts naming them, but primarily as proto-typical mental images of situations and sequences of actions. If this is correct the significance of the general emotion-terms are learned through experience of some particulars that come to be accepted as representative. To state my point in the words of David Hume: “ ... some ideas are particular in their nature, but general in their representation.” (Hume, 1978. p. 22)!

It is possible that some mental states like fear and pain (maybe also joy and sadness) are learned by a sort of inner ostention, that is, learned by recognition of the feel of them in instances that are taken to be typical. But it seems implausible that all emotions are learned that way. What about feelings of insult, betrayal, debasement and annoyance? Such states seem to be learned through a more complicated process than a mere directing of awareness towards a definite modification of the mental. But maybe some sort of ostention still plays a role in learning such emotional states. Let's say that someone tells us that the feeling of insult is what you feel when someone treats you with disrespect or disregard in a way that you have not deserved. Or, the feeling of betrayal is the emotional state you are in when someone has gained your confidence and later uses this against you. If you recognise this state, it would of course not be through a pure inner ostention, you would have to understand the significance of the words “disrespect”, “disregard”, “deserved” and “confidence”, which could hardly be done through inner ostention. But let's assume that he says: Imagine yourself in a situation were you have told someone some about your intimate secrets. The next day this someone finds out that he will reap some advantage by revealing

the secret to a third person. This situation would instantiate “betrayal”. I might answer that I have experienced such a situation. I might say that; sure, I had some feelings, but I didn’t recognise them as feelings of betrayal, I just felt a bit confused and disappointed. Well, he says, you are still too naive and inexperienced to understand what is really going on. You ought to have felt betrayed!

This story could be interpreted as one where someone presents a paradigm of betrayal. It has three crucial elements:

1. It gives a paradigmatic description of a social situation of a certain sort.
2. It gives a normative evaluation of such situations.
3. It appeals to imagination as a mean to learn and accept a certain propositional attitude (that of betrayal) toward some event.

Now, this is not an instance of inner ostention in the proper sense. There is an appeal to a social code and a cultural hierarchy of values. Even the sort of imagination that is appealed to is acquired. So is the authority that is ascribed to this imaginative capacity and to the example that is (presumably) accepted as paradigmatic. What I am suggesting is that such learning is a stage in a process of growing into a society and culture, a stage that presumes that we are already fairly well acquainted with, and emotionally integrated in, the same society and culture. Once this is done, and the possibility of appealing to imaginary situations conceived in a certain way, is established, it seems to be possible to establish a finer spectre of emotional nuances by some sort of imaginary ostention. This is presumable what is done in exemplary stories that children are told, in movies, literature etc. What is done is something similar to what Descartes does when he establishes a criterion of truth. “Establish the background of the process of methodical doubt. Then imagine that, as you

do this, you suddenly realise that, in this doubt, you can be certain of your own existence. There you have the paradigm situation of being in a state of certainty!” To make my point even more recognisable I will mention that I regard the classical theories of social contracts as instances of the same way of delimiting a phenomenon, in that case duty or political allegiance. “Imagine a situation where people have no duties towards one another, but where each individual have a natural right to defend its own life. Imagine then that they, for some reason, find this situation to be contrary to their real interests. As a civilised being you surely know what a contract is and what is implied when you enter into a contractual relationship. Imagine therefore that these people enter into such a relationship with each other and with a political institution that they establish through this contract. There you have a paradigm situation of duty and political allegiance!” Note that this way of representing duty/political allegiance does not explain how these phenomena come to exist. It presupposes that those to whom it is directed are already acquainted with cultural codes that make the described situation intelligible as one of establishing a certain type of relation. What it does, is to establish a scene with a high degree of dramatic, and hence emotional appeal, a scene that can be used to identify and classify real situations as being of a certain type and exclude other ones from the same type. As an aside it could be mentioned that there are slight differences within western culture as to what situations should properly be understood as instances of this mythical scene. American children would for instance probably fairly early recognise their relationship to their school as being of such a kind, while Scandinavian children would hardly be able to see any such connection at all.

There is a neat combination of descriptive, normative and emotional elements working together in this sort of imaginative activity. Descriptive of course, as a situation is

described. Normative, for two reasons: First, the feeling of being betrayed cannot be detected as a pure fact, it has to be conceived as resulting from some wrongdoing. Of course, the sequence of events could be given another description, but that would not have been a description of a betrayal. Second, the story claims to have some superior authority as to what is to count as betrayal. Emotional, because this situation provides a code for the individual to interpret his affective state as being of a certain kind, that of feeling betrayed. Paradigm examples of emotional significant situations often have a dramatic or scenic structure. This probably accounts for their emotional appeal.

Characterising emotions through descriptions of certain situations and sequences of events is not a modern invention. In fact it can be traced back to Aristotle. In his book on *Poetics* we read that a tragedy aims at evoking pity and fear. Aristotle tells us that the plot as well as the characters must have a number of determinate qualities for this emotional effect to be obtained. We may reconstruct Aristotle so as to describe, not only the tragedy as a cause of the emotional effects, but also of the intentional structure of these emotions themselves. About the hero of the tragedy he says that he must be; “a man who is not eminently good and just, yet whose misfortune is brought about not by vice or depravity, but by some error or frailty. He must be one who is highly renowned and prosperous ... (*Poet.* 1453 a). The reason is that “pity is aroused by unmerited misfortune, fear by the misfortune of a man like ourselves” (*Ibid.*). He also directs certain demands to the structure of the plot. There is no need to enter into details here. The main point is that he describes a relationship between certain emotions and a certain narrative structure, a structure that needs certain character-types to fit in. Of course he deals with a certain elevated type of pity and pain, one that purifies the soul. The point is, however, that he describes these as emotions, which

have intentional objects that are peculiar to them, and that these objects (or characters) should be of a kind that makes them fit into certain dramatic sequences of events.

A similar point is made in the *Rhetoric*. An emotion-type is characterised by being connected with certain scenes, populated by certain personalities that take part in a certain type of chain of events. The definition of benevolence, for instance, goes like this: “Let it then be taken to be the feeling in accordance with which one who has it is said to render a service to one who needs it, not in return for something, nor in the interest of him who renders it, but in that of the recipient.” (*Rhet.* II. Vi.) The emotion is defined through a description of a certain social scene where certain types of actions are performed out of certain types of motives, and where certain types of intended outcome is excluded.

We would however not have needed Aristotle to recognise the idea that different emotions are related to scenes that are typical to them, and which are in fact appropriate to characterise them. The war and the courtroom are scenes for a number of emotions. Places where you can be seen by everyone, or hidden from everyone, are also such scenes. The same may be the case with situations where you help someone or are yourself helped. Who has not staged imaginary scenes where one accuses people with whom one is angry or jealous? The accusations are targeted against the villain, but aimed at the ears of the judges. The defendant is ascribed stupid or naïve attempts to excuse himself, only to be annihilated by the impartial judges.

Ronald deSousa has introduced the concept of “paradigm scenarios” underlying any specific emotion-type. The concept of paradigm scenarios is connected with the biography of the individual and with single or particular emotionally significant experiences as part of

this biography. The starting-point is not a complex, but the emotionally relatively undifferentiated state of the infant, a situation largely characterised by genetically determined responses being elaborated along the way through interaction with the responses from the infant's social environment. Such is his hypothesis concerning the function of paradigm scenarios: "My hypothesis is this: We are made familiar with the vocabulary of emotions by association with *paradigm scenarios*. These are drawn first from our daily life as small children and later reinforced by the stories, art and culture to which we are exposed. Later still, in literate cultures, they are supplemented and refined by literature. Paradigm scenarios involve two aspects: first, a situation type providing the characteristic *object* of the specific emotion-type ... and second, a set of characteristic or "normal" responses to the situation, where normality is first a biological matter and then very quickly becomes a cultural one." (de Sousa, 1987. P. 182).

I find it basically correct to assume that emotional differentiation above the genetically determined level reflects cultural phenomena as well as individually acquired ways of integrating the emotional significance of the various situation-types.

The image of emotions as something being maintained by an inner scene can illuminate the structural components of the emotions. In his book *The Theory of Moral Sentiments*, the 18th century Scottish philosopher Adam Smith develops this idea. In Smith's conception this inner scene also has an imagined audience. Smith calls this the "impartial spectator". The concept of the impartial spectator plays a double role. One function has to do with the conception of (social) emotions depending on the existence of certain sorts of social situations and phenomena for their formation. Such emotions also need to be exhibited in such situations. The impartial spectator is the informed and competent audience. The other

function is to provide a standard for the correctness and reasonableness (in Smith's terms "propriety") of the emotions. The impartial spectator's sympathy with the subject who has the emotion provides a confirmation and an acceptance of the propriety of his expression. So, the imaginary scene with its audience is a precondition for the subject's evaluative attitude towards his own emotions.

Smith's conception of emotions and emotional expression as something being sustained and regulated by imagined social scenes has an important implication concerning the multi-perspectivity of emotions. In my discussion of emotions as mapping-system I said that the subjectivity or subjective "feel" of the emotions informs us of our position relative to the other elements in the mapped situations. Smith's theory, which is about emotions of the kind I call "social", tells us that such emotions presupposes the existence of other perspectives than my own on the same situation. It also tells us that the "feel" that signifies my position reflects the "feel" that signifies a certain other position within that situation. Expressed in terms other than Smith's this means that the intentionality of such emotions is more complex than is the case with more basic emotions. The condition of satisfaction for fear is often quite simple. The conditions of satisfaction for, say benevolence, insult or admiration include that the adequate expression of these emotions would be conceived in a certain way from a perspective other than one's own. Isn't it partly this state of affairs that makes social emotions "social"?

I shall now proceed to comment on the significance of the concept "mythologies" for the understanding of how we establish and maintain emotional differences. The concept of mythologies as relevant in this context is taken from Solomon. He claims that certain myth-like situations and figures are connected with certain emotions in such a way that an

emotion is characterised by the mythology that exemplifies it and makes it vivid in the imagination. For instance, the mythology that instantiates *anger* is normally a “(c)ourtroom or Olympian mythology; oneself as legislator and judge; the other as defendant. Oneself as the defender of values, the other as offender.” (Solomon, 1993. p. 229). *Innocence* can carry the picture of “(l)amb among wolves, the beautiful soul. (Ibid. P. 273). Indifference is entrenched through “(t)he pariah mythology, the man whose name is stricken from all records, who is allowed no quarter, no conversation, not even punishment ... (Ibid. p. 270).

Similar applies to other emotions. One of the consequences of Solomon’s approach is that it allows a sort of structuralistic and even dramaturgic investigation of the emotions. Such an investigation would illuminate not only the structure of the myth, but also of the intentionality of the corresponding emotion. One condition of satisfaction for anger would, for instance, according to this be that the other really is an offender of values and oneself really is a defender of the same values. Another consequence is that learning of emotions and emotional nuances, as far social emotions are concerned, is interwoven with the learning of the symbolic significance of mythical situations. We learn to imagine ourselves into myths, as partakers in significant dramas.

According to Solomon, mythologies function as “the synthesis of our emotional judgements”. That gives this concept a special place among the emotional categories. Let us make a thought-experiment where we consider each category as a variable that can have different values. If we give each of the twelve categories/variables two possible values, we will get 4096 possible combinations. If we give them three, we get 1594323 combinations that are potential emotional nuances. Thinking the categories as possible combinations will however give us a highly inadequate picture of human emotions and the way they are

established. An emotion is not an aggregate of different components, but a unified response to a situation, wherein the situation itself is conceived as unified from the perspective of the subject. Solomon's theory seems to be that a theoretically possible combination is a possible emotion if and only if we are able to form a mental image that combines the categories in a, to us, significant situation or scene, a scene to which the emotion applies as a description. The special status of the concept of mythologies compared to the other emotional categories is that it functions more like a picture or rather a plot, than a concept. Its unifying function rests on the ability to form emotionally significant pictures of situations or scenes as part of plots. This would imply that our emotional repertoire is connected with our ability to form such scenes or situations in the imagination.

When I discussed the categories relevant to emotions, I gave each category a discrete value. We saw that a very limited number of values gave a large number of possible combinations. If we think of the categories as vectors with non-discrete values, the number of possible combinations is virtually infinite.

Paradigm scenarios and mythologies could be thought of as signifiers and codifiers of emotions. Signifiers, because they come to stand for the corresponding emotion and because stories and scenes provide very strong identifications-objects, codifiers, because they in a way express the objective existence of situations as, say, unjust, honourable, pitiful, degrading etc, and so rationalising the corresponding emotions as proper responses.

There is probably some limit to the number of emotional nuances that the social interaction-system of a given culture is able to handle, in the sense that the agents within

the system would be able to recognise them. In a given culture there will therefore be a limited number of mental state-types that could be counted as, or recognised as, emotions. If this assumption is correct, the amounts of myths that are considered significant within a given culture would tell us something about the emotional code underlying the interaction in the same culture. If this is true, we should however use the concept “myth” in a very broad sense, meaning every situation-type or type of action-sequence to which we respond in some emotional way. The main point here is that there is no emotional state that could not be referred to some situation-type or action-sequence type. Emotional characteristics of a given culture can therefore be taken as an indicator of corresponding social characteristics. Elster (Elster, 1999 b) points out that Aristotle’s theory of emotions also serves as a source to our understanding of peculiar features of the social organisation of the Greek city-state.

So far I have said that the concepts of “paradigm scenarios” and “mythologies” play an important role in two respects: 1. They contribute to the conceptualisation, and hence the identification of certain emotional states. 2. They model the emotional states so as to fit into a culture’s code for social situations and interaction. Now, there is also a third function that can be attributed to these concepts, a function related to the maintenance of the individuals’ autobiographical selves. There is a need, not only to identify, but also to re-identify emotional states. The emotional life of an individual displays a certain degree of continuity and coherence at, at least, two levels: First, each individual has, or acquires, some character-traits or dispositions that shape patterns of response and action which come to be typical for this individual. Second, an individual with some degree of emotional experience is able to recognise his emotional states as states of a kind that he has had before. This rather obvious fact raises two questions whose answers are not all that

obvious. The first is this: How is it that we identify an emotional state as one we have formerly experienced? The second is this: How can a particular experienced emotional state have transfer-value to other emotional particulars? My answer to both these questions will involve the concept of “paradigm scenarios”. An emotional state comes to be classified as being of a certain kind because it is interpreted as an instance of a paradigm. The paradigm delivers a sort of tentative structure in light of which the state is interpreted. The paradigm also provides a pre-evaluation of states of this kind and also a prediction of what kind of actions, outcomes and so on which are to be expected in these types of scenarios.

To conclude: Paradigms or prototypes have references in two directions. On the one hand, they picture features of social situations such as to give them a special kind of significance. On the other hand they are used to organise emotionally relevant information into recognisable emotions.

Chapter 5

Self

Dealing with the subject of human emotions we will have to address the difficult problem of the self. From Plato to Descartes the philosophical tradition treated this problem as concerning a substance capable of independent existence, i.e. existence independent of the body in which it was temporarily imprisoned, but not subject to the same changes as was the flesh. In this conceptual edifice the emotions were able to distract the self, but not really to affect its essential properties. This line of thought was broken in the 18th century, most noteworthy by David Hume. There are different and conflicting interpretations of Hume's writings on the self. Some say that he altogether denies the existence of the self. Others say that he denies the existence of a substantial self. Others again say that his writings about personal identity in Book I of the *Treatise* only prepare the way for treating the self as a product of the development of our emotional life. Either way, what is fairly beyond doubt is that Hume himself found his own treatment of personal identity unsatisfactory or at least incomplete. His comments on his own shortcomings are, in my opinion, so instructive that I think they are worth a closer consideration.

It is in the *Appendix* to the *Treatise* that Hume revisits the problem of personal identity with which he had dealt in Book I. He declares that he has been entangled in a sort of contradiction or "in a labyrinth" from which he can't find his way out. The problem as he states it, is this: "... all our distinct perceptions are distinct existences ... the mind never perceives any real connexion among distinct existences." (*Treatise* p. 636). So how can it be that these perceptions are in some way unified in a mind or self?

In my opinion Hume's problems have the following sources:

1. The concept of mind as "a something" that is facing the task of putting different, mutually independent elements together into a unified whole.
2. A static, as opposed to a dynamic, conception of self.

Hume invites us to conduct a thought-experiment. Let us consider a mind that is at the level of an oyster. This mind, as Hume asks us to imagine it, has only one perception, e.g. that of thirst. Here you can find no self, only thirst. Add one more perception, e.g. that of cold. Still there is no perception of self. Add as many particular perceptions you like, the introduction of the unifying self remains a mystery.

First of all; I believe that Hume would be better off had he asked us to conceive of an *organism* instead of a *mind* with one, two ... perceptions. What should we say if he had put the question that way? Let us assume that this organism has a perception of thirst. Let us then assume that it also has a perception of water. It would in fact not be an open question whether or not the organism is able to establish some sort of relation between these two perceptions. If it were not, it would not be an organism at all. The point here is simply that organisms come down to us from evolution with the capacity to unify perceptions like these, or else they do not come down to us at all. If we focus on the system of synaptic connection and neural circuits instead of mental units, there wouldn't be a problem of unification of "thirst" and "water" at all. The point is that if there is a problem like this, it is solved at a deeper level than the conscious one.

But maybe it might still be a unifying problem. People can wake up after a “night out on the town” and ask themselves “Am I the person who did so and so?” We also have the pathological cases of split brains and dissociated personality syndrome. But these situations present different kinds of problems than the one Hume asks us to consider. His problem is not how unity can fall apart, but how it is established.

If we accept that the problem is not how a mind may have a self, but how an organism may have, or in fact has a self, we no longer have to accept the terminological and methodological platform upon which Hume erects the problem. It might or it might not be correct that it is impossible to discover a “real relation” between two “perceptions of the mind” if you look at one of them at a time. It might or might not be correct that one already existing perception cannot produce any real new perception. And it surely is correct that the self cannot exist as a separate perception prior to or independent of other perceptions. But it is even more certain that organisms are able to co-ordinate themselves on the basis of more than one perception at the time. It is simply a brute fact that organisms, human and others, present relatively coherent responses to a huge amount of information, and that they do this all the time.

Antonio Damasio (Damasio, 1999) has suggested a model depicting different levels of self-hood and consciousness. In fact it is intended not only as a model, but also as a description of distinct neural systems. According to Damasio, understanding the evolution of the self is a question of understanding the complexity of our neural system. Every organism can be described as having a *proto-self*. This self is concerned with upholding the boundaries between the organism and the environment in which it is situated and with mapping events relevant to the organism’s survival and also to respond to such events. This mapping depends on a neural system processing relevant information for the

organism. The proto-self does not require consciousness. There is also a *core self* based on three neural systems, one mapping events affecting the organism, one mapping the organism's visceral responses to such events and one supervising and mapping the relation between the two others. Damasio calls this third system a second-order system. The core-self is the conscious awareness of what is going on from moment to moment. The awareness of the relation between events in the environment and the organism's responses to these events, has the form of feeling or is presented in an affective mode. If Damasio's description is correct, we have the tools we need to solve Hume's problem. It is simply not true that there is no "real connection" between distinct perceptions. The core-self and the conscious awareness that it produces, *is* this real connection. If we were to look for the origin of basic cognitive categories like causality, distance, resemblance and other categories relevant for interaction between the organism and the environment, the core-self would be a good place to start looking.

It may be added that core consciousness, according to Damasio, also results in an enhancement of the image of, and sharpening of the focus on the object affecting the organism. It also strengthens the awareness of the feeling that the object causes in the organism, even to a level of a *feeling of knowing*. This seems to have two important consequences. First; it strengthens the awareness of the division between the objective and the subjective. Second; it seems to imply that conscious states as propositional attitudes have a deep biological basis in the origin of consciousness.

In humans there is also awareness, or rather a conception, of a self that can properly be termed the *autobiographical self*. Now, given that there are emotions that are distinctively human, what is the relationship between these and the autobiographical self? Addressing

this question, it is important not to forget that the human autobiographical self is built upon the foundation of the proto-self and the core self. That implies that it is not built out of nothing, and that the terms we should use to grasp the distinctive features of this type of self-conception, do not signify anything that could alone explain the construction of this self. What we are looking for is a necessary, not a sufficient condition for the autobiographical self.

Like all levels of self-hood, the autobiographical self is concerned with internal stability and borders. Now, what tools do we have that enable us to fulfil such a task? I believe that part of the answer regarding stability is that we have a grammar or formal structure (being a linguistic elaboration of some basic dimensions of our perception-systems and response-systems) that enables us to conceptualise a vast variety of situations, narratives about ourselves which are relatively coherent, and the institutionalising of our inter-personal interactions. The tools we have for border-drawing are, in addition to these, certain means of defining or conceiving of ourselves as unique individuals or persons. To be sure, our stabilising and individuating endeavours engage more than our emotional repertoire. But this fact represents no reason why we should not ask what part is played by the emotions in the creation of the autobiographical self.

It is quite possible, and even probable, that our biological make-up can account for the functions and the formal structure of the tools mentioned. The reason is that the functions fulfilled are part of our organic survival. Nevertheless there is also a dimension of self-construction involved in the making of persons. And in the case of making an autobiographical self, our biological make-up is such that it also requires the element of conscious self-construction to fulfil its task. This construction has the form of

interpretation and explanation, interpretation of the significance of the events taking place in the person's life and causal explanation of the relationship between the various events. I am here suggesting that the making and maintaining of an autobiographical self is also a question of self-interpretation and self-construction. The intriguing thing here is that this is obviously a biological process, but that it works in cultural terms.

The important point is that the culturally and conventionally determined semantic meaning of the content of the autobiographical self cannot be reduced to their biological underpinnings. Stated in other terms: the formal structure determined by our neural system cannot generate the semantic meaning of the terms necessary to maintain an autobiographical self through a lifetime. John Searle (Searle, 1984) has claimed that "syntax is not sufficient for semantics". That might or might not be true. But even if it should be untrue taken as a universal claim, there is one reason to believe that the kind of semantics that could be generated by our formal neural system would be insufficient to uphold autobiographical selves. The reason is this: When we think of, and reconstruct, our lives as relatively continuous wholes, we use words like "will", "decision", "responsibility", "freedom", "reasons", "belief", to mention some central terms. We also use terms like "justice", "shame", "guilt", "deserved reward", "pride", to mention some slightly less central. Two things are important here: 1. Such terms derive their meaning, not only from the fact that they serve to individualise the persons of which they are predicated, that is, define the borders between this person and other persons. They also get meaning from their function as co-ordinators of action in the way that they define the social and cultural space within which action takes place. 2. People have not always ascribed the same significance to these terms (or rather the concepts that they express) as regards the interpretation of the continuity of their lives. The ancient Greeks did not

distinguish between reasons and causes the way we do and they would not have recognised the significance that we often ascribe to this distinction. Their concepts of “will”, “freedom” and “justice” were different from ours. “Shame” played a different role in the lives of men. “Guilt” had a different significance during the European Middle Ages from what it has today. The semantic meanings of the terms which would be central in a man’s or woman’s life-story were different. But the ancient Greeks and the medieval Europeans had the same neural make-up as we have. If our neural system could explain the semantics of the terms able to hold together a life-story, it would therefore also have to explain why a system of meaning can function as the spinal cord of a life-story at one time, but not at another. This is not to say that our neural system is insignificant in explaining such changes. If it didn’t have the plasticity and the adaptive capacity that it actually has, we would obviously be much more vulnerable to changes in our social as well as natural environment. But nevertheless, the reasons for the changes of the semantics of individual life-story building should be sought in such changes, not only in the capacity to adapt to them. As a result of such changes, the efforts needed to construct and maintain an individual life-story have slid from adaption to rigid social institutions and stereotyped ritual patterns of interaction to the individual’s own imagination.

One addition: One might ask whether or not words like “will”, “decision”, “belief”, “guilt” and so on really refer to anything. However interesting this question might be, in the present context it is irrelevant. The present point is just that some concepts work, that is, they function as central concepts in the construction of autobiographies.

As opposed to our core consciousness, the maintenance of this self essentially makes use of culturally determined semantic meaning and conventionally established conceptions of

values. Our neural make-up is probably the key to the grammar we use to construct interpretations of events and the continuity of our personal lives. That means that the grammatical and categorial basis upon which we build semantic meaning goes deeper than the conscious construction of such meaning. Auto-biographical selves depend on the ability to use the formal system of consciousness to construct semantic meaning, in short transform the formal system of awareness into grammatical rules in the linguistic sense of this term. An important implication of this assumed fact is that our language has pre-linguistic preconditions and that these are operative in our daily use of language. If we really accept that we are biological beings with an evolutionary history, this is hardly surprising. Language is a transformation of processes that already exist, and this transformation is made possible by the symbolic representation of these processes. The alternative would be to say that language is constructed according to rules and systems that in their origin are linguistic, that is, the evolution of language is possible on the presumption that language already exists. To be sure, this does not mean that language introduces nothing but a different way of representing ongoing processes. Language makes possible the construction of a new level of self-hood, the autobiographical self. Language, however, does much the same as what is intentionally contained in any experience, ascribing something to something. In addition it builds up the formal system and rules of such ascriptions. Such rules are used through a lifetime to ascribe experiences to one and the same self and to construe the connection between these experiences as a relatively coherent narrative. I tend to regard the self as a concept with certain conditions of satisfaction, one of which is that the experiences I ascribe to myself are in fact mine. That would mean that it is one of the conditions of satisfaction of human intentionality that there is in fact a self. Another condition is that the connection I conceive to be between these experienced events are in fact the actual connection between them, at least to some degree.

All this seem to imply that the possessive pronouns “mine” and “my” in a way goes deeper than the personal pronoun “I”. Without “my experiences” there can be no “I”, while the opposite is not necessarily the case. If it is not my arms and legs that I see and feel below my head, then it isn’t I who is sitting here. But I don’t believe I need to have the concept of my autobiographical self in order to conceive the pain in my right leg as my pain. Consider a situation that suddenly makes you very happy. First there is the feeling of happiness. There is an additional reflection needed for the conscious awareness that “this is happening to me”, or even more, “I am the one this is happening to”. The conditions of satisfaction for this last mental state are indeed more complex than for the one that identifies a pain in the leg as *my* pain. One reason for this, is that the “I am the one”-state also includes that this pain could be ascribed to the same self that went to a movie yesterday, who recognised the story of the movie as similar to one he/she had read five years ago, and so on.

It is obvious that one significant feature of the consciousness that goes along with the autobiographical self is memory. Memory may perhaps be thought of as a sort of coded storing in our neural and motor system of certain events. But it can be no doubt that this is greatly supported by language. A single word spoken may in fact cause the release of neurotransmitters leading to the activation of motor- as well as perceptual memory that in many instances may present itself as a sort of emotional memory. Language also enables us to construct extensive narratives where the narrative as a whole makes the parts potentially present. Just consider the way you remember songs or poems. You don’t have to feel that you remember every single line in a song the moment you start to sing. One line seems to introduce the next. Similarly, a good way to try to remember something is to recall the context in which you first experienced it. It is interesting to note that Damasio uses the term “word-less narrative” to designate the most elementary awareness of

something going on affecting the organism (Cf. Damasio 1999). What he seems to assume, is that narratives, or rather narrative structures, are more basic than language. I find this term and the theory at the basis of it, extremely interesting because it implies that elementary consciousness starts with construction of connections between events/states. But nevertheless, it is obvious that the verbalisation of narratives allows awareness of incomparably more complex narrative structures, even so complex as to integrate the events of a lengthy human life.

I will have more to say about the connection between narratives and human emotions later. At present it is enough to say that human selves in their normal condition are selves of language-users.

To conclude, human self and the sort of consciousness that is essential to it, are based upon a biological system that can be described as a formal system. This system can be, and is, transformed into linguistic rules. The human emotions depend upon this sort of self. It therefore seems reasonable to assume that at least some human emotions can be described as having a formal structure that is similar to the structure of these rules. The foregoing chapter was an attempt to show that human emotions can indeed be described as constructed according to such a system, and that this description is recognisable as a description of the phenomena we experience as human emotions. By this I mean that it conforms to our reflected awareness of our own emotions.

Social Emotions and Autobiographical Selves

In chapter 3 I tried to establish a connection between emotions and situations, hence also between social emotions and social situations. The essential features of this connection, as I described it, are: 1. Social emotions are responses to social situations in the way that they map them and inform the subject of her location and position on that map. They also inform the subject of the significance and importance of the situation and of her position in it. Generally, it tells her what is at stake and how. 2. Social emotions construe situations as being of certain kinds. 3. This is done according to paradigm examples of such situations. 4. Such paradigms (or prototypes) are activated in what could properly be termed “the imagination”. 5. Paradigm situations can be represented as temporal sequences of events and actions, as stories or dramas. That adds a temporal dimension to the map.

In chapter 4 I sketched the concept of mapping. I will now try to knit the topics from chapters 3 and 4 together, claiming that certain phenomena, and therefore the mapping of such phenomena, are of special significance in the construction and maintenance of autobiographical selves.

This knitting together will include some clarification of the concepts “person”, “institution” and “narrative”. My underlying assumption is that the meaning of these concepts presupposes that they be constructed within a conceptual scheme which is also a “social-situation mapping system”. I will therefore first have to elaborate the concept of such a system.

Emotions and Social Mapping

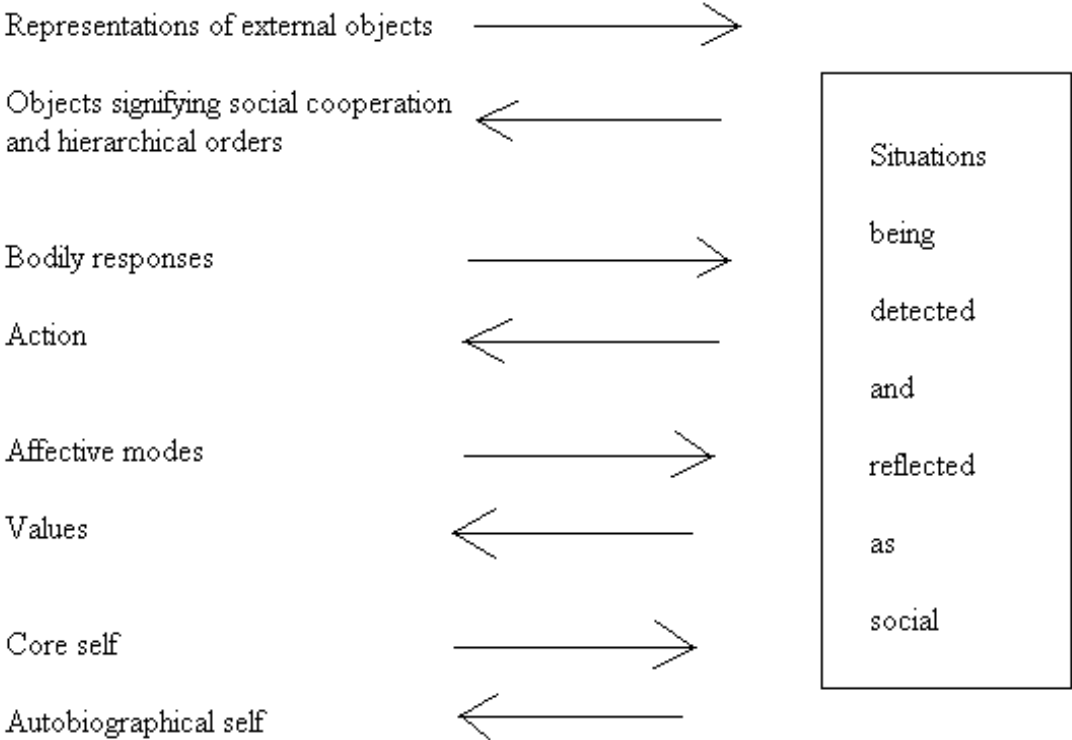
The approaches of deSousa and Solomon give us an account of the structural elements of emotions as a potential for their propositional articulation. For Solomon this even seems to be more than a potential. He is in fact very explicit about his opinion of emotions as “judgements”. I believe that this account, however valuable it is for the understanding of our ability to relate rationally to our own and other’s emotions, disregards another important aspect. What I have in mind is the following: The theory holding that emotions are organised according to categories as well as paradigm scenarios and mythologies can be elaborated in terms that relate emotions not so much to their potentially rational articulation as to their characteristic structure as mental phenomena. The categories and other concepts relating to the structural components of emotions can be regarded as dimensions of “maps” used in our orientation in our natural and social environments. A paradigm scenario may be regarded as a sort of emotional recurrence pattern or recurrence model.

I have already sketched the concept of “mapping”. I will now suggest that certain emotions, those that I have called “social”, in fact function as part of a system mapping a special kind of phenomena, namely social situations. If that is right, we should expect that it would be possible to reduce emotional differences and nuances to different variations along some socially relevant basic dimensions. What would these dimensions be? As an approach toward this question, I will suggest the following four dimensions as a tentative answer: Object, Self, Action and Value. This answer is however problematic in certain respects. First of all, it is given from a phenomenological perspective. It is therefore a kind of top-down approach, initially devoid of empirical support as to its neural basis. It will therefore rely on the possibility of finding something on a lower level to sustain it. Its

rationale is simply that it seems to cover basic dimensions of our conscious emotional experiences. Second, taken as mapping dimensions they differ from dimensions of, say, colour maps or taste maps in that they are more complex or composite. The social features of a situation are not directly accessible to any sensory modality. The social character of phenomena related to one dimension depends on its relations also to the other dimensions. An object cannot, for instance, be seen as “social” isolated from its relation to the self-, value- and action determinants, and correspondingly with phenomena related to the other dimensions. The dimensions, therefore, cannot be based on sensation alone, but must relate to something else. If Damasio is right in his assumption that there is a kind of second-order neural system as described, it is tempting to think of the dimensions of our social map as in some way related to this system. This system, as we recall, produces representations of the relationship between the external environment and bodily events and, according to Damasio, it is the system in which the affective modalities originate. Such representations occur as elementary narratives or categories. What I suggest, is that the system sustaining social mapping may be regarded as a third-order system, relating to the second-order system. Its input units are the packages “external objects-bodily responses-affective modes-self”. These, I suggest, are re-evaluated in light of their relevance to the interpersonal (social) sphere in such a way that they are re-interpreted as to signify social situations and systems. This re-evaluation or re-interpretation transforms the response-factor into an action-dimension, the affect-mode into a value-dimension, the object-factor into a symbol related to inter-personal inter-action or hierarchy-systems and the core self into a kind of self-interpreting and self-constructing unit.

The radar metaphor that could be taken to be a connotation of the term “navigation” as used in this book indicates that navigational aspect of social emotions work through some

kind of projection-reflection system. Let's try to elaborate this and see if it might hit some essential features of the workings of social emotions. We have, as mentioned, the four constituent elements of the core-self, representations of external objects, bodily responses, affective modes and self. Imagine now that we project each element unto our social environment in such a way that each projection hits certain kinds of features in this environment. Imagine then that these projections are reflected back to us. How are they then received, having thus been reflected through this social mirror? Now, I do realise that I can't give any demonstrative arguments on this topic. All I can do is to make a suggestion that is meant to appeal to the phenomenology of social experience and ask the reader to consider if this model might illuminate some of the structures behind our experience of certain situations as social. The suggested model for a social navigation system is this (where the right-directed arrows represent the projection-element, while the left-directed arrows represent the reflection-elements):



An essential feature of a possible system of social mapping would be that it relates to phenomena as relatively durable and stable. If it is correct, as Damasio assumes, that the core-self has a kind of momentary existence, monitoring what goes on from moment to moment, it would therefore be insufficient. I will later claim that such relative permanence or durability is, in part upheld by institutions and narratives.

To summarise: I have suggested a possible mapping system representing certain dimensions to which we are emotionally sensitive, dimensions which at the same time represent features which are relevant for conceiving a situation as “social”. I am here suggesting a possible overlapping between an emotional and a social space-map. This theory of “double mapping” is obviously founded on the idea that the identification of situations as social ones is connected with certain kinds of emotional responses to such situations.

Now to the paradigm/recurrence-factor. “Recurrence” shall in this context refer to the sort of feedback mechanism which is known to be operative in all perceptual systems. The essence of this component in such systems is that the incoming information at an early stage (the bottom-level) is selected and shaped according to patterns and structures which have already passed through that stage, and are fed back. What we see at one moment is for instance selected and shaped by what we have formerly seen. Initially this process is performed by our neural network and therefore does not depend on conscious interpretation. It may be thought of as an elementary form of memory shaping new perceptions on a pre-conscious level. What I am getting at here is that paradigm scenarios

may be carriers of emotional memory, structuring emotionally relevant information according to certain stereotypical situations, and giving emotional memory transfer-value. If this is so, paradigms pick out something as, say envy, admiration or hatred in a chaos of emotionally relevant information. Let's consider a situation where we experience a set of emotionally relevant information with a profile we have never before experienced. After all, if the vectored category-model is correct (or, to be more accurate, have representational value), this would probably happen all the time. The prototype pattern recurrence model allows us to understand how this does not result in emotional chaos. For a set of emotionally relevant information to be experienced as an emotion of a certain kind, it must be organised according to a prototype pattern, a paradigm. Such patterns fill out what is missing and adjust atypical features of a situation. Prototype patterns work in virtue of two features. First, and fairly obvious, they establish distinctions between situation-types. In addition I believe that they are based on a semantic system. The first feature provides something that is essential to any semantic system, a system of meaning-discrimination. In addition to that I believe that for a pattern to become established as a prototype emotion-pattern, the meaning discriminated must also have some sort of significance, it must exemplify a situation-type that for some reason or other is taken to be important. They must, in short be instances of experiences of some importance in the lives of those who have them, small or large dramas where something is at stake.

To summarise: I suggest that the relationship between emotions and prototype patterns (paradigms) works through a combination of two levels. One concerns categorisation and discrimination, the other concerns the significance of the units thus categorised/discriminated. The first gives the logic or grammar of emotional experience, the second transforms the units into a sort of signals for appropriate responses, value-

reactions, action-strategies, adequacy in relation to antecedent events and so on. This is done by placing the discriminated emotions in a context of typical (and proper/improper) action-systems, social institutions and cultural codes. Such contextualisations transform the discriminate entities into semantic-like entities and are therefore probably the contribution to the evolutionary history of emotions which is distinctively human (given that humans are the only species with a semantically based communication-system).

Persons, Institutions and Narratives

I have suggested an overlapping between the mapping of social emotions and social situations. What kinds of connections are there between these two mapped “landscapes”, the “mindscape” and the “socioscape”? This question can be raised on the phenomenological as well as on the neural level. An attempt to give an answer on the phenomenological level would have to look for experiences, which in an essential way involve both phenomena, experience of subjectivity as well as of sociability. I believe that the concept of a “person” signifies such an experience-type. A person is a self that is defined or experienced in relation to its social context, which in turn is determined by its inter-*personal* relevance. I also assume that an autobiographical self is built around the concept of “a person”, and that person-hood is developed on the basis of the type of configuration mentioned. Person-hood is a phenomenon that relies strongly on the symbolic power of certain emotion-objects, that is, the ability of such objects to signify aspects of the identity of the person whose emotion-object it is. Here are my guideline assumptions concerning that symbolism: 1. The relevant emotion-objects symbolise the person in a possessive mode. I will explicate this below. 2. The stories supporting such emotional configuration are conceived as more or less extensive parts of the biography of

the person in question, in the sense that she comes to identify herself with them, conceiving of them as parts of her life-story.

Thinking of oneself as a person, not only an organism, involves conceiving oneself as a possessor of certain things. Another way to put this is to say that it involves the ability to develop the sort of consciousness that is expressed in terms connected with the possessive pronoun. Among those things that in modern western cultures are considered most inseparable from the person are rights, experiences, the body and certain events like birth and death, all connected with the person through a very existential meaning of the possessive pronoun. “My birth”, “my death” and other important or essential events to which that pronoun can be connected, signifies “me” in such a way that without it, there is no “me”. The “my/mine” keeps up the borders between me and them. The things, which are “mine”, are parts of my person in a certain way, as far as violations of these things, to a greater or lesser extent, are violations of my person. Many of the connections between a “my” and some term should be interpreted as moral or legal claims, others as claims of accountability (my actions, my promise, my opinion). The important things to emphasise are that the concept of “person” is a way to maintain borders, and that such border-keeping is, to a large extent, articulated in moral and legal terms. It is also articulated in epistemic and existentialist terms, “my knowledge”, “my opinion”, “my experience”, “my life”. These terms, as person-delimiting terms, serve an appropriating function.

If we as individuals are to think of ourselves as one rather than two, three, fifty or hundred persons (each with its own borders), there must be some sort of connection between the “I” who claims ownership of this house, the “I” who admits responsibility for this action, the “I” who has experienced the death of her parents etc. There is probably a natural

connection between the subjects of different experiences as long as we talk about direct bodily encounters with the environment. The I on whom this wound was inflicted is the same as the one whose wound is now healing. The I who experienced a burning of his arm, is the same I that now spontaneously withdraws from a hot object. But when it comes to experiences mediated by symbolic code-systems like events embodying social norms or institutions, the case is not so straightforward. The continuity of the I must in such cases be maintained by a stable pattern of symbol-decoding. Neither is there much, when it comes to more than objects that bear direct impact on the bodily survival, in the individual's organism that let us determine what should be counted as essential to the person and what should be regarded as more accidental. There are even no natural phenomena to discover, which would tell us how close to the person various things should be regarded to be, to be regarded as relevant to her person-hood. During the European Renaissance the individual's capability to distinguish between true and false statements came to be regarded as part of his or her person-hood. There had been a time when the Roman Catholic Church had been entrusted that task. Similar things could be said with regard to private property, moral decisions, personal opinions and so on. In the eighteenth century, Jean Jacques Rousseau describes a culture that is felt as a threat to the individual's possibility to experience its life as its own. A century later, Søren Kierkegaard and John Stuart Mill describe the same. There arises a feeling that there is a dimension of authenticity to individual life and that this requires a unified person that is in possession of all important aspects of her own life. As a brief aside, I will mention that I consider the authenticity-ideal (Cf. Taylor, 1991) to be distinguished from other ideals of person-hood basically by the kinds of connections that it takes to be essential for the person's identity. Connections to things that involuntarily can be lost are insignificant. The authenticity-ideal can therefore be regarded as a sort of neo-stoicism. There is however no need here to go further into this. The main

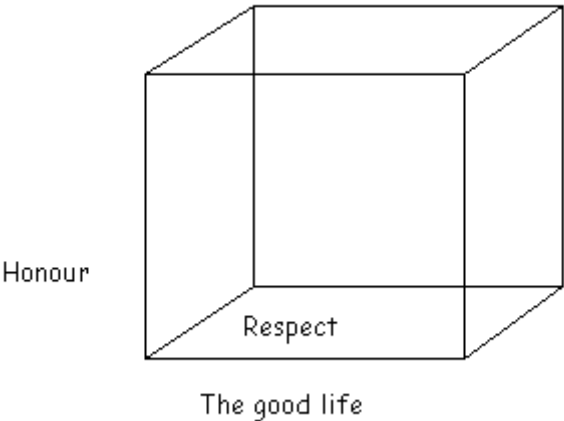
point is that the requirements for person-hood are historical and cultural variable and so is the kind of complexity that must be united into one single biography.

A few words about the concept of “identity” as applied to “persons”. It is not like we should first make up our mind about the meaning of “identity” and then apply this meaning in an adjective way to the meaning of “person”. The meaning of “identity” in this context is rather specific to the meaning of “person”. Applied to the organism it would mean something like keeping up the borders between the organism and its external environment, the maintenance of the internal homeostasis and the establishing and strengthening of neural patterns facilitating individual adaption (memory). Identity of the organism in the sense mentioned is certainly prerequisite to the identity of the person, but only as a *sine qua non*. In the identity of the person something is added. It must be sustained by conceptual edifices that establish relations between elements that would otherwise be unrelated. Such relations depend on the semantics of these edifices. It must also be sustained by certain formalised patterns of human action and interaction. I will call such patterns “institutions”. And finally, these factors must be supported by various kinds of emotional prototyped configurations.

The concept of “the person”, as I have treated it, comes very close to the notion of “the moral self” as this has emerged through a philosophical tradition counting, among others, Locke, Rousseau and Kant. The person should therefore, in some way or other, be conceptualised in moral terms. Such conceptualisation should also include some conception of how these terms relate to each other in the determination of person-hood and the individualisation of single persons. This notion, or rather *our* notion of our moral selves, has been extensively investigated and analysed by Charles Taylor (Taylor, 1989). I

mention this because I believe that in his exposure we can find, maybe unexpected, support for the space-map concept and the concept of navigation (or orientation) employed earlier.

Taylor assumes that our moral identity is established in relation to “strong evaluations”. Such evaluations relate to three different dimensions that together make the framework of our moral identity. He calls this our “moral space” and baptises its dimensions; “the good life”, “honour” and “respect”. Taylor thinks of these as dimensions of a map used partly for knowledge of the “moral” landscape and partly to identify one’s own location. Taking his words literally, it doesn’t seem improper to represent the moral space as three-dimensional spaces are often represented:



To be sure, the map-space-orientation metaphors are basic in Taylor’s theory, and he makes coherent use of them. He says that there is an “essential link between identity and a kind of orientation” and “(t)o know who I am is a species of knowing where I stand”. (Taylor, 1989. p 26-7). Taylor’s view seems to be that any possible “self” could be represented as a location within such a three-dimensional “moral space”.

Taylor may be right or wrong in his identification of the dimensions of moral space. His account may also be incomplete. However that may be, I find the space-map approach itself valuable. First it maps a kind of moral landscape in which we orient ourselves. Second, it provides us with a conceptual tool to understand a certain aspect of identity, the “who am I” aspect. Translated into the map/space-metaphor it would be “where am I”. If what I have formerly said is correct, then the key to the understanding of the connection between “who” and “where” should be found in the subjective experience of being on a certain “location”. An identity-problem should, according to this methodological approach, be regarded as a kind of dislocation. Maybe it is also possible to read another identity-aspect out of Taylor’s approach, the one concerning identity as sameness over time. If so, it would be because the three dimensions could be taken as a framework, not only for moral orientation, but also for construction of life-stories, the basic determinants of the topics of such stories. Since a life-story shouldn’t be imagined as a stand-still on a certain point, he stories would have to make the patterns of movement between different points intelligible.

Taylor’s approach has of course nothing to do with neural networks. It is cast strictly within a phenomenological mould. It would seem far-fetched to think of the three dimensions of the moral space as three separate neural pathways. It has no reference to the neural system whatsoever. Nevertheless it is formulated within the logical structure of an organism mapping its environment as part of its orientation and navigation in this environment. It is essentially the logic of organisms being located in a space relative to various types of phenomena and objects among which it has to navigate. The three dimensions make this orientation a kind of triangulation. Knowing your position relative to three co-ordinates will of course give you the exact, absolute as well as relative position of yourself and other objects in a three dimensional space. And this is in fact the one and only

point of the formal construction of a space through co-ordinates. So, even if it has no reference to neural systems, it is hard to see how it could describe the situations of beings other than organisms with such a system.

My notion of the identity of the person, as I have treated it, is a notion of something that is not given in the organism. It must therefore, in some way, be created. I believe that there are two phenomena whose functions deserve special attention here, institutions and narratives. Both can be regarded as kinds of prototypes. Institutions obviously serve to stabilise the conditions of action, and to some extent they do this in their capacity of functioning as formal prototypes of certain interaction-forms. Narratives serve to create some meaningful connection between the different events that makes up the experiences of an individual human life.

I will not try to give any comprehensive account of institutions. I will only focus on certain features that are relevant for the inter-dependency between social emotions and inter-personal situations. Institutions may be regarded as formalised inter-personal situations. Institutions like schools, families, hospitals, governmental organisations contractual relations etc are all structures that to a certain degree define, for the agents, what kind of situation they are in. They are a kind of formalised and standardised situations. They set prototype forms to be filled. The prototype forms can be specified as the rules that should govern interpersonal interaction within the various institutions. My assumptions are, first, that we, as we become participants in institutions, also develop corresponding emotions which are also, in some sense, formalised, second, we participate in institutions as “persons”. I believe that David Hume makes a correct observation saying that we often confuse such emotions, which he calls “calm passions”, with reason, which he takes to be a

kind of value-free theoretical insight. I am however uncertain as to the question whether or not we have a capacity for such a kind of value-free pure theoretical insight. Actually I suspect that “reason” is in fact a kind of “calm passion”. The reason for this suspicion is that “rationality”, like emotions, can be regarded as a response-modal to situations in which we are in some way engaged. This similarity between rationality and emotions can be regarded as more basic than the differences in response-style. Regarding the conception of rationality as a response-modal to situations, I will make the following assumptions:

1. Some kinds of responses to social situations can be called rational.

2. These responses relate to, and map, other situation-features than do responses we conceive of as emotionally relevant. Responses of that kind are responses to certain general (and maybe formal) features of situations.

3. There is a link between institutions, or the institutional aspect of situations, and the kind of responses that could be conceived as rational.

Our ability to construe or describe situations is, according to the theory I have defended, an important determinant of our emotional life. The development of that ability is therefore an important aspect of our emotional development. Now, rationality can also be regarded as a type of situation-construal and -response. Taken in itself, these claims seem to be rather obvious and trivial. I believe, however, that they have at least one possible implication not quite so obvious and trivial: It seems to follow that the difference between emotions and rationality, as related to the topic at hand, is a difference of types of situation-features to which we respond. There is, in short, a different story behind a rational response than behind a “hotter” emotional one. But that shouldn’t blur the basic similarity and

evolutionary continuity between them, both being kinds of situation-responses. The life-forms which enable us to respond rationally to situations are those structured according to formalised types of social encounters and inter-action.

For institutions to function as stabilisers in a human individual's life, this human must in some way be attached to, and engaged in, these institutions, that is, she must be able to respond to institutionalised situations in light of some relatively highly formalised prototype paradigm. Any felt forms of engagements and attachments should be considered an emotional response to the situation or object towards which the engagement or attachment is felt. The bottom-line of these considerations is that rational responses to social situations in important respects are not different in kind from those conceived of as emotional.

Now to narratives: By "narrative" I here mean a chaining together of events in such a way as to make the sequence intelligible, recognisable, significant or meaningful in some way and also make the parts significant in light of the sequence in which they partake. My remarks on narratives will be even more narrow-focused than those on institutions. Their purpose will merely be to sketch the idea that there is a concept of narratives that relates to autobiographical selves, and that such narratives construe these selves as *persons*. My remarks on narratives are therefore relevant only as related to the autobiography of persons.

The narrative that is my autobiography shall make it intelligible how the links in a chain of events are "mine", not other people's. That means that they must be conceived in some

possessive modality that excludes other possessors than me. One way to do this is to give some non-replaceable emotion-objects a privileged status in my life-story. I will return to this in the next chapter. When we “interpret” ourselves, be it as we act in single episodes or our lives as a whole, we create narratives. According to Damasio even the simplest core consciousness shows an elementary narrative structure. Core consciousness is constantly engaged in a sort of story-construction, whose elements are significant events in the environment, such events in the organism itself and categories that connect these two chains. The autobiographical consciousness depends on the ability to construct a far more complex story, one that can bring together the external and internal chains of events of a lifetime, chains that at each moment of life also includes projections of the future. Typically, this is done, not only in terms of causes and effects, but also in terms of culture and morals. Culture and moral, because the my/mine connections very often cannot be conceived as purely natural. Now, such terms might be inadequate in some way, they might simply get things wrong, they may have no other reference than they construe themselves, they may, as Marx says, in fact be made to obscure some facts. They may also, as Max Weber suggests, be made to create some comforting meaning in an objectively meaningless world. Nevertheless, the fact that we construct narratives remains, and the narratives form some of the structure of the understanding that we, rightly or wrongly, have of ourselves. If there would be no autobiographical self without such story-construction, then the study of narratives will tell us something important, or even essential about human self-hood, at least as it is under social and cultural conditions similar to our own.

So far I have said the following about the more highly developed human emotions: They can be articulated in accordance with a certain category-scheme (or basic grammar). They

are connected to certain paradigmatic scenes or myths. They map social situations, are conceived as parts of social situations, responses to social situations, as, at least potentially, exposed in such situations and adjusted in accordance with the conception of such situations. This implies that, as conceptualised, they are already structured so as to fit into certain types of social situations. And they very often do that in one out of two ways: They either express the distinction between me and what is mine and, on the other hand, others, or they counteract destructive consequences of this distinction, such as loneliness, helplessness and isolation. This does not imply that nature has been so kind as to give us a standardised emotional repertoire to meet all kinds of social situations. More likely it means that our emotional development takes place in interaction with our social and cultural experience.

There is a striking similarity between life-story construction and social emotions, so striking that it is tempting to regard life-story construction as a kind of emotionally relevant prototype situation-application on a larger scale. Both are based on underlying paradigmatic dramatic stories. And both take these stories to have some important significance, they exemplify something of more than ordinary importance.

These features of our emotions should lead us to conceive of them, at least the highly developed symbol-based ones, as sequential rather than static or substantial phenomena. By this I mean that emotions should be viewed as forms of our awareness of ourselves as partakers in various forms of interaction-sequences and lasting interaction-schemes. These human emotions are therefore not only to be taken as instantaneous affections or feelings, but rather as phenomena that in a way represent the structural continuity between the experienced past, the present and the anticipated future. I am suggesting that the ability to

conceive social situations in light of prototypes, and the ability to construct life-stories, are stages along a common line of development. Our ability to view ourselves from the perspective of a life-story is, according to this, developed from our ability to identify social situations. In a way similar to the one in which we apply prototypes to situations, we also learn to apply prototype stories of what an individual relatively coherent and meaningful individual life is like.

What I am getting at here, is that there is a certain emotional repertoire that enables us to maintain our person-hood through a variety of social encounters, and that the construction of this repertoire takes place in cultural and moral terms, because these are the sort of terms that define such encounters and situations and define the personal roles for the agents. These encounters are, in turn, related to each other as incidents in a more or less coherent life-story.

I believe it is important to stress the “more or less” aspect here. Persons are complex beings and are normally not capable of explaining some logical connection between all their actions and experiences. Owen Flanagan (Flanagan: 1991) points to the discrepancy between, on the one hand, the philosophical conception of a moral person as a being whose actions in all spheres of life are guided and determined by the same basic principles, and, on the other hand, our common-life based conception that allows of considerable inconsistency and multitude of action-forms. The first kind of concept comes close to a concept of “a saint” rather than of a “person”.

What we call memory obviously plays a central role in maintaining autobiographical selves. Whatever memory may be, it implies that something past influences the present.

Memories are not only past experiences as passively present. They have active impact on the present and the way we project and prepare for future actions and events. This is done in various ways. One of these concerns the emotional continuity of our selves. Here we can see the significance of our emotional paradigms. Human emotions must have some sort of transfer value, or else they could not contribute to the maintenance of our autobiographical coherence.

The transfer value of past experience involves a bi-lateral relation between the past and the present. Not only is the present interpreted in light of past experiences, but also the past is continuously reinterpreted and re-evaluated.

I believe that the concept of “recurrence” may be utilised also to understand the function of narratives in constructing and maintaining a life-story. I have spoken of paradigms as recurrent forms. Now, narratives may also be regarded as recurrent patterns, organised in certain sequential ways, in a person’s life. Certainly there is no contradiction between recurrence and variety as long as the variety can be interpreted within a recurrent scheme. What I am saying is that narratives function as such recurrent enlarged schemes. These schemes are obviously dynamic and sequential, rather than static. Such sequential forms may follow different types of logic in different persons. They may for instance be homeostatic or morphogenetic. In the first case they will tend to confirm and stabilise the harmony of different significant elements in a person’s life. In the second, they will contribute to the interpretation of the dynamics in a life-story as a continuous renewal.

What is the relationship between recurrent emotional prototypes and life-story narratives? After all, recurrence itself is no more than mere repetition, which seems insufficient to

sustain a narrative. A possible answer might be this: Recurrent emotional prototypes can develop into parts of a life-story because such prototypes are changed during the recurrence-process. The prototypes themselves therefore come to get a history, the history of their application under different circumstances and continuous adjustment according to these. This answer would be consistent with a very common experience. Most of us have probably experienced that the way we are affected by emotions like envy, jealousy, humility, ambition, etc. changes over time, even if we can recognise such emotions in the changed feel of them. Thus, what I am saying is that these emotions themselves have a history which comes to be part of the life-stories of the persons who have them. I take it that the history, or rather histories, of an individual's emotional prototypes is what creates her personality. A personality cannot be defined in terms of emotional states or particular emotional experiences. On the other hand, it seems very reasonable to define it (at least partly) in terms of emotional *dispositions*. We talk about angry, happy and jealous people. Now, angry people are not angry all the time. Happy people are not happy all the time, nor are jealous people jealous all the time. When we still call people angry, happy or jealous, we are talking about their tendencies to activate certain kinds of emotional prototypes and to suppress other. The history of an angry man's prototype for anger is different than a happy man's history for the same prototype. It has, among other things, a broader application. This might possibly account for the differences in *personality* between the two.

I have assumed that there is a connection between certain emotions and certain situations. I have also assumed that some situations obtain paradigmatic status and serve as prototypical forms signifying the corresponding emotions. I am now saying that such situations can be enlarged so as to signify a life-story. Life-story situations are more open than

single-emotion situations, they have a larger sequential potential. Now, here comes a hypothesis: To these enlarged situations, the life-story narratives, correspond certain emotions or emotional structures that could similarly be regarded as enlarged. I take these “enlarged emotions” to be the feelings of the experience of having applied the emotional prototypes over time in a variety of situations. The emotional experiences themselves become objects of what might be called “higher order emotions”. Among those feelings are the ones we often term “moral”. By “moral feelings” I mean the feelings that are associated with the character-traits that are usually considered morally good, like kindness, honesty, generosity, courage, mercy and so on. The reason why I call them “feelings” is that they are also learned through paradigm scenarios which come to be accepted as models of prototypes which should be activated as responses to as many situations as possible, prototypes that are to suppress alternative, not so desirable, response-types. Hence, they instantiate topics around which a life-story should be constructed. The paradigm scenarios in question are stories rather than isolated scenes. The story of “The Good Samaritan” might serve as an example. In a prototype way the story instantiates, not only a certain kind of action and feeling, but also a certain kind of person. Most of us know other stories that give prototypes of the courageous person, the generous person, the person who never gives up and so on.

Now, no one, or at any rate, very few of us, is such a person through and through. Hence, if this is what it is to be a person, it seems to follow that there are in fact very few of them. To phrase this as a slightly weaker claim: Each of us would not be *one*, but many persons! In fact, I believe that this is true in a certain sense. Our character-traits are to some extent situation-relative. Certainly there are truth-conditions according to which the sentence “I am not the same person as I was 20 years ago”, is true. On the other hand, there are also

truth-conditions according to which it is false. In forensic contexts it would, in most cases, be false. And even if I consider my present personal identity, the “who I am now”, it would not be unconditionally true. Were it to be true, this would obviously contradict the individualising function that I have ascribed to the concept of person-hood. We shall therefore have to look for something that serves to individualise persons in their own conception. In the next chapter I shall discuss one possible candidate for this function.

Chapter 6

Replaceable and non-replaceable Emotion-Objects

They say that everything can be replaced

Bob Dylan

I have tried to establish a connection between autobiographical selves mapping their social surroundings through some kind of emotional system that can be made accessible to rational reflective awareness by the emotists themselves. Does that mean that all aspects of human social emotions are rational in the sense which links “rationality” to “universality”? The occasion to raise this question is given by my assumption that the concept of “a person” serves an individualising and border-drawing function and that our personhood depends on a certain emotional development. It might seem that this is at odds with a conception of emotional development as an increasing rational reflection on our emotional states, at least if rationality is equated with universality. I am now going to discuss this by approaching the following question: Are some human emotions uniquely tied up to certain particulars in such a way that they cannot be transferred to other objects? If so, would this mean that such emotions were irrational? If we were to answer yes to the last question, the reason might be that we usually (and correctly) associate rationality with some form of universality. It is obvious that universality is also often relevant in the assessment of the rationality of emotions. If I am proud to be rich, I might rationalise the pride by saying that everyone who is as rich as I am (everything else equal) would have the same reason to be

proud. In fact I might feel insulted if they were not, taking this to be an insinuation of my pride be irrational.

Before proceeding on the question of the replaceability of emotion-objects, we should distinguish between posing the question in the context of first-person experience and third-person description. In a third-person context any emotion-object is in a sense replaceable. If not, it would be impossible to say anything general at all about such emotions. The problem we are dealing with here is whether or not emotions that in first-persons are experienced as having irreplaceable objects are always at odds with rationality-criteria.

The universality that is often ascribed to rationality can be expressed through a formalisation-procedure. We may e.g. formalise a proposition expressing/describing an emotion and say that such an articulate expression will be a good reason for the emotion if and only if it will still be the same good reason when we, certain conditions given, substitute certain words within the form with other words. If there are good reasons for A to be proud to be the owner of x because x has the property y, then the same ought to be the case with B if he were to be substituted with A. With certain reservations, A would have the same reasons to be proud if he was the owner of z, if z also had the property y or a property with a value equal to the value of y. Similar things could be said if C was shameful because of r, and so on. If fear is an adequate reaction when I believe that a lion is waiting around the next corner, it would still be adequate if it is really not a lion, but a crocodile. The reason is, of course, that they are equally dangerous. If it turned out to be a mouse, hanging on to the fear would be irrational.

But does this hold for emotions in general? Would any emotion that cannot stand such a substitution-test be irrational? What about erotic love? If I love *her*, should I mean that every other man ought to do the same? And should I love every other woman who has the same properties just as much? It seems that Plato would answer yes to all these questions. As I have mentioned above this could be a conclusion of the argument in the *Symposium*. In fact Plato goes even further and tells us that love ought to change direction from the first-order object towards the formal object (the object's *eidos*), which, according to him, has a separate existence independent of the first-order object. It is reasonable to take Plato to conceive of this as a development towards greater rationality.

Two things seem here to be in conflict. First, any emotion, no matter how unique and irreplaceable it is to the subject who has it, has a formal structure which it shares with other emotions and other people's emotions. Secondly, individuals often experience objects as if they express the uniqueness of the emotion-object. One dimension in this conflict is, of course, grounded in the difference between the first and the third-person perspective.

Now, this situation characterises not only the perception of emotions. I may say that the anatomy of my body is the same as the anatomy of other people's bodies, and it is still not irrational to mean that my body has a unique and non-replaceable place in my life. Our physical existence as well as our conception of self is essentially based on the fact that each of us is a unique particular. But maybe the analogy isn't as good as it might seem. We may experience emotions of attachment to things exterior to ourselves as irreplaceable, in spite of the fact that they may go away while we stay. It might be our children or parents. The body can obviously not disappear leaving us behind. Nevertheless I believe one aspect

of the analogy to be illuminating. This has to do with what it is that creates our sense of individuality. I will return to this below.

DeSousa has a lengthy discussion of a Greek myth, illustrating the problem at hand. The myth goes like this: Alkmene loves her husband Amfitryon. The problem is that Zeus falls in love with Alkmene, and one night, while Amfitryon is away, he disguises himself as her husband and goes to bed with her. Later Alkmene finds out, and feels angry and insulted. Has she any good reasons to feel this way? After all there was no conceivable difference between the man she loved and the man she had slept with. He had the same properties and the same ways of responding to her. If she loved his qualities or properties, she had got exactly these. Nevertheless, some of us probably feel that she has good reason for her hurt feelings. After all she had been cheated. It wasn't the man that she thought it was, but another who had the same properties. But is Amfitryon more than the sum of his properties and his ways of responding to things? Even so, Alkmene has been fooled, and one of the reasons is that her emotion, in this case, had an irreplaceable object.

What sort of problem is raised by this case? The question is not whether or not Alkmene ought to feel insulted, but whether or not she has reasons to feel this way. It isn't hard to imagine that the episode might be a rather good experience. After all you don't go to bed with gods (or goddesses) every day. Let's equate the following two assertions;

1. She has a reason to feel insulted.
2. Anyone else would have a reason to feel insulted in an equivalent situation.

In that case I believe that we are making the mistake of confusing particulars with universals. The starting-point of our problem was that Alkmene in fact felt insulted, not whether or not any rational person in such a situation should feel insulted. The rationality

in her emotions cannot, or should not, be put to a universalising-test. Rather it should be assessed on the basis of the intentional content in her way of acting.

One more myth, now from the Old Testament, and, seemingly, with a different morale. The Lord wants to put Job to a test, and as a part of the test he sends a desert-storm that kills Job's seven sons. After operation desert-storm has been terminated and Job has passed the tests, the Lord makes things up to Job. He gives Job seven new sons (and even three daughters – probably interest). The end of the story is that Job dies 140 years old. A good life has come to an end. The Lord has put things even; Job has no reasons to complain. Is this a story of paternal love with replaceable objects? In that case one might well ask if it is emotionally credible. What about the dead seven? Maybe the story isn't about this at all. Maybe it tells us that there is only one irreplaceable object of love, God.

In that case the morale would find support in the writings of St. Augustine. He can be read so as to mean that mundane earthly love has replaceable objects. There is a demand that "thou shalt love thy neighbour". Are we to abide, we would of course need to know who that is. Augustine's answer is that it is everyone, and that this implies that we should love everyone as much as anyone else (*De Doctrina Christiana* Ch. I). That does not mean that we shall do the same for everyone, or help everyone, but the only reason for this is that this is impossible. His basic point seems to be that love is real only when its particular object is in a way indifferent, which means that it is replaceable. Every object of real love is in reality a stand-in for God. This might perhaps seem as a possible description of justice, but to Augustine it is, strangely enough, a description of love. I do not mention this only to exhibit a philosophical relict. In fact I believe that the Augustinian conception has had deep impact on western thought-models. In my opinion Augustine confuses the

significance of closeness and distance. The significance of closeness is underrated in areas where closeness is important and it is overrated in areas where it is improper. If this is correct, moral philosophers stressing the importance of closeness in all human relations as well as value-bureaucrats, are both carriers of the Augustinian legacy. I believe that the core of the Augustinian strategy of generalising emotion-objects is to create a public sphere where emotional conflicts are neutralised or rather delegitimised, but where authority still is founded on the basis of emotional support. Augustine is therefore an important source in the construction of a public sphere uniting neutrality, rationality and emotional cooling.

Let's put historical differences aside and look at Alkmene and Job as our contemporaries, modern people with modern emotions. Is Job's love more rational than Alkmene's? Maybe, but if so, I suspect that we should assent to all the following three sentences:

1. We ought to be rational.
2. There may be a conflict between emotion e and the criteria for rationality.
3. Emotion e ought not to be adjusted according to these criteria.

There are two interesting questions involved here. The first is: What is the function of emotions having non-replaceable objects? The second is: What would be the consequences of confusing replaceable with non-replaceable emotion-objects?

As to the first: I believe that emotions with non-replaceable objects have to do with certain preconditions for human rationality. To be more precise, they are important in the formation of our sense of individuality or self. They are in other words parts of the process in which we create an individual personal identity, i.e. they serve to draw borders between

my self and other selves. Generally I find it useful to relate the word “identity” as used in the term “personal identity” to “identification” and “identifying”. What we call “personal identity” is formed in a process where we identify ourselves with different things. Some of these identification-objects are conceived as unique to us, as non-replaceable. If that were not the case, the alternative would be to identify oneself with the formal structure of the emotional relation between subjects and emotion-objects, not with the substantial characteristics of the object. This alternative would imply that developing personal identity would presuppose that we were already rational, rather than that the development of rationality presupposes subjects with personal identity. What I am saying is simply that part of the way in which we are individualised is by identifying ourselves emotionally with things that we conceive as unique to ourselves. How else could we hold our own biography apart from that of other people? Now, many things are unique to every single one of us. I am the only person in the history of the world who occupied seat 29 A on the plane that left Tromsø September 8 1998 on flight SK 376. I believe that very few people would consider this unique position to be a basis for what we conceive of as personal identity and far less rational individuality. On the other hand, in any culture, at least one slightly similar to our own, there are certain other criteria of individuality or individualisation. It might be that someone is the oldest son of certain parents, that he or she has written a certain book or scored the decisive goal in the WC. It might well be the case that there are no universal criteria across time and culture. But in any culture consisting of individuals who conceive of themselves as human individuals, there must be some criteria. In a paper, written as a comment on Hume’s *Treatise*, Donald Ainslie calls such criteria “existential connections”. I believe that it is important to realise that the relevant differences between existential and non-existential connections are identified as emotional differences. And emotional involvement in existential connections is experienced as relations to non-replaceable

objects. I might add two things: First, I take this to support what I have formerly said about the semantic that is necessary to support the maintenance of autobiographical selves. Second, if we correctly describe some human emotions as attachments to non-replaceable objects, we are probably not describing the human mental system as it must necessarily be. At most we are giving a phenomenological description of the way the sense of person-hood is developed in cultures similar to our own and hence, how autobiographical awareness is established in such cultures.

If we are specific enough when we classify emotions, there are in fact certain emotional attachments that have non-replaceable objects (as there are attachments that only seem to have such objects). In the *Antigone*, Sophocles describes a situation of this type. In spite of the King's prohibition, Antigone buries her dead brother under the prospect of being put to death herself. For her the decision to do this is an existential decision in the sense that it cannot be avoided without jeopardising her identity as a particular individual or, more accurately, as the particular individual she wants to be. "Had I lost my child, and were my husband dead I might have a child with another husband. But as mother and father are hidden in the bosom of the earth, a brother will never be raised to me." The tragedy tells us, among other things, of the close connection (indeed an existential connection) between being a particular human individual and being emotionally attached to a non-replaceable object. Antigone's emotions towards her dead brother have an intentional structure that is such as allowing no replacement. The fact that this situation has been created by circumstances independent of her intentionality or choice is of no relevance to her. The basic thing is that this is in fact the situation she finds herself in. It has correctly been pointed out to me (by prof. Karl Halvor Teigen) that Antigone's remark shows that her brother is *in principle* replaceable. Even so, if we add that we live our lives in relation to

situations and events that are in principle non-reversible and non-repeatable, the *in principle* replacability functions as a reminder of the existential divide between the *in principle* and the *in fact*.

Back to Alkmene: The answer to the question whether or not she has good reasons to feel hurt or deceived is not to be found in an investigation in the properties of her lover, but in the intentional content in her emotions and actions towards him. The intentional content in these is not “making love with a man who has the properties x, y and z”, but “making love with a particular man, Amfitryon.” She was misled to believe that the conditions of satisfaction for her emotions were present. Should it be irrational to feel deceived in this situation, the reason must be that it is irrational to have emotions directed towards a particular that cannot be replaced by an identical particular in such a way that this would change the relation between intention and action. And if what I have said about existential connections is correct it would also be irrational for her to conceive of herself as a certain particular, namely Alkmene. The same would of course be irrational for the rest of us. It would then be irrational to form personal identity in the way that we actually form personal identity, living through a biography in which relations to certain particulars have an essential significance.

One addition: Even if a certain emotion-object has an irreplaceable position in my life, it doesn't follow that it does not have the same position in the life of another individual as well. However, in many such cases our feelings towards such objects are based on the, maybe false, belief that we ourselves hold a unique position in their emotional life. This would instantiate what Hegel described as an essential part of the development of self-consciousness, “the desire to be desired” (Hegel: 1970).

I have claimed that there are certain emotional operations that have an individualising function and that these work as emotional attachment to non-replaceable objects, or to be less absolute, to objects that are hard to replace. To be sure, these are not the only operations that function as “individualisers”. As we all know we also act in contexts where our individual particularity is not the focus-point, contexts that nevertheless function as mediators of our individuality. We act as citizens, as professionals and so on. As participants in political discussions or as teachers we are not (or at least we ought not to be) identified as the eldest son of NN or with other unique properties. In these circumstances we should be identified through competencies and ascribability-criteria that are general, not unique. Our performances in such arenas are neither emotion-neutral nor irrelevant to our personal identity, i.e. our conception of who we are. What I am getting at is simply that the emotion-objects in these spheres are characterised by their formal features, which implies the relative unimportance of their particular or individual traits. Their importance as emotion-objects is essentially connected with their status as representatives of universals. They are therefore to a large extent replaceable. It would presumably be relatively unproblematic to accept that we act and perform in many contexts where our individual biography, personality and emotional attachments are, as such, relatively unimportant or irrelevant. On the other hand, I want to point out some consequences of this, consequences that would be more controversial. First, it is inaccurate to represent the difference between primary and secondary relations as a difference between interaction being guided by emotions and interaction being guided by reason. Human interaction is, normally guided by both levels of our mental life. All human interactions have elements of emotional attachments and engagements, but there are differences between emotion-objects along the lines I have drawn. Second, it would help

us understand how our attachments to general emotion-objects contribute to the formation of our identity or, to be more precise, how the formation of our identity presumes an interplay between different sorts of objects. Third, it will help us formulate a critique of different forms of confusion between attachments to particular and more general emotion-objects.

What I have in mind when mentioning such form of confusion is something along the line of Max Weber's diagnosis of modern culture and politics. Two key-concepts in the formulation of this diagnosis are "disenchantment" and "charisma". A few words about the relevance of Weber: According to Weber the prehistory of modernity is coextensive with the history of rationality. One of the main characteristics of this historic process is the tendency towards secularisation and "Sachlichkeit", the bureaucratic version of objectivity. Individual phenomena come to be regarded and treated as "cases" not as unique particulars. This implies that they are processed administratively as well as cognitively within a formal framework that is considered right under the precondition that each case is replaceable by other cases that are similar in aspects conceived as relevant. Weber regards the modern bureaucracy as the most important institutional expression of this way of considering and dealing with things. The problem with rational bureaucracy is not that it is illegitimate or inadequate in the field of public administration and decision-making, but that its way of thinking also expands to other spheres of life, to the "management" of private and personal life. Parts of this life are organised through emotional attachments and engagements that have less replaceable objects than the cases of bureaucratic treatment. To be sure, many relations that more often than not have replaceable objects are built around an irreplaceability-ideology. One such relation is romantic love, and often marriage. But the fact that they are construed within this ideological framework shows that there are real

emotional needs here. Such needs are not *per se* irrational, but they may be transformed into irrational expressions. Weber's most famous case is the modern version of the charismatic leader. The irrationality of this phenomenon consists in his authority being founded on his presumed irreplaceability while his legitimacy is founded on the opposite.

Summary

I have tried to render a version of human emotions which takes into account the fact that they perform functions which must be performed for any organism, the function of border-drawing and navigation/orientation as well as the fact that they perform these functions for beings that are also rational and consciously self-reflective, beings living their lives in societies functioning through more or less shared cultural codes. The first, the general functions, can be described at a level where terms denoting phenomena that are essentially conscious may have a contingent status. As to the second, the phenomenological aspects of the emotions seem to be essential.

The conscious aspect of the emotions not only includes awareness of certain instantaneous emotional states. In addition it comprises consciousness of intentional content as well as a set of sequences of experiences. It therefore involves what we may call a “phenomenology of the emotions”. Concepts denoting what we usually conceive of as “persons” and “rationality” are part of such a phenomenology. I do, however, believe that any conscious emotional state also contains essential elements which are themselves not conscious. I take this thesis to be supported by a phenomenological as well by a non-phenomenological perspective. By the first one because the alternative would lead to the implausible assumption that a phenomenological investigation of oneself might come to an end because there would be nothing more to discover or be aware of. By the second one because this would lead to the implausible assumption that conscious phenomena are sufficient to sustain themselves.

I have also tried to extend the elementary mapping function that I take any emotion (conscious or not) to perform, to apply to human emotions as they work in social and cultural contexts. My thesis has been that they perform a double function. They function as a system mapping social situations and informing us of our exclusive position within these, and they develop into the spinal cord of individual life stories and hence autobiographical selves.

Bibliography

Ainslie, Donald C. *Scepticism about Persons in Book II of Hume's Treatise*. Forthcoming, *Journal of the History of Philosophy*.

Anscombe, G.E.M. *Intention*. Second ed. Oxford: 1963.

Armon-Jones, Claire. *Varieties of Affect*. Toronto: 1991.

Aristoteles. *Poetics*. Loeb ed.

Aristoteles. *The Nichomachean Ethics*. Oxford University Press 1992.

Aristoteles. *Rhetorics*. Loeb ed.

Augustin. *De Doctrina Christiana*.

Austin, J. L. *How to do Things with Words*. Oxford Uni. Press: 1962.

Chalmers, David J. *The Conscious Mind*. Oxford Uni. Press: 1996

Churchland, Paul M. *Matter and Consciousness*. MIT-Press 1988.

Churchland, Paul M. *The Engine of Reason, the Seat of the Soul*.

MIT-Press, 1995.

Churchland, Paul M. Conceptual Similarity across the sensory and the neural Diversity. *The Journal of Philosophy*. XCV, NO 1, Jan 1998.
pp 5-32.

Churchland, Paul M. Neurosemantics (unpublished).

Cooper, John M. *Reason and Emotion*. Princeton University Press: 1999.

Crumley, Jack S. *Problems in Mind*. Mayfield Publishing Company, 2000.

- Damasio, Antonio R. *Descartes' Error. Emotion, Reason and the Human Brain*. New York: 1994.
- Damasio, Antonio R. *The Feeling of What Happens*. New York, San Diego, London 1999
- Davidson, Donald. *Essays on Actions and Events*. Oxford: 1980.
- Davidson, Donald. Epistemology Externalized. *Dialectica* 45 (1991) pp. 191-202.
- Davidson, Donald. Subjective, Intersubjective, Objective. In: D. Hutto & P. Coates (ed.) *Current Issues in Idealism*. Bristol: Thoemmes Press, 1996.
- Davidson, Donald. Seeing through Language. *The journal of the Royal Institute of Philosophy*. 1997 pp. 15-27.
- Descartes, René. Passions de l'Ame. *Oevres de Descartes XI*. Paris, 1986.
- de Sousa, Ronald. *The Rationality of Emotion*. MIT Press: 1987.
- Elster, Jon. *Strong Feelings*. MIT Press, 1999 a.
- Elster, Jon. *Alchemies of the Mind*. Cambridge Uni. Press, 1999 b.
- Flanagan, Owen. *Varieties of Moral Personality*. Harvard Uni. Press: 1991.
- Greenspan, Patricia. *Emotions and Reasons*. New York & London: 1988.
- Habermas, Jürgen. *Theorie des kommunikativen Handelns*. Frankfurt.a.M. 1981.
- Hegel, G. W. F. *Phänomenologie des Geistes*. Theorie Werkausgabe. Frankfurt a. M.: 1970.
- Heller, Agnes. *A Theory of Feelings*. Assen 1979

- Hume, David. *A Treatise of Human Nature*. Selby-Bigge utg. Oxford: 1978.
- Izard, Carroll. *Human Emotions*. New York & London: 1977.
- James, Susan. *Passion and Action*. Oxford: 1997.
- James, William. What is an Emotion. *Mind*, 9 (1884) 188 – 205.
- Johnson-Laird, P. N. and Oathley, Keith. The Language of Emotions: An Analysis of a Semantic Field. *Cognition and Emotion*, 1989, 3 (2), 81-123.
- Kenny Antony. *Action, Emotion and Will*. London: 1963.
- Lazarus, Richard. A Cognitivist's Reply to Zajonc on Emotion and Cognition. *American Psychologist*, Febr. 1981. Vol 35, 222-3.
- Lazarus, Richard. Thoughts on the Relation between Emotion and Cognition. *American Psychologist*. Sept. 1982. Vol 37. 1109-24.
- Lazarus R. & Lazarus B. *Passion and Reason*. Oxford Uni. Press 1994.
- LeDoux, Joseph E. Cognitive-Emotional Interactions in the Brain. *Cognition and Emotion*, 1989, 3(4), 267-289.
- LeDoux, Joseph E. *The Emotional Brain*. New York, 1996.
- Llinás, R & Ribary U. Coherent 40-Hz oscillation characterizes dream states in humans. *Proceedings of the National Academy of Sciences*, vol 90 (1993).
- Locke, John. *Two Treatises of Government*. Cambridge Uni. Press: 1960.
- Lyons, William. *Emotion*. Cambridge Uni. Press: 1980.
- Nafstad, Petter. *Pasjoner og personer - En analyse av David Humes lære om de menneskelige følelsene*. Oslo: 1999.

Nagel, Thomas. What Is It Like to Be a Bat. *Philosophical Review* 83 (1974) pp.435-450.

Nussbaum, Martha C. *The Therapy of Desire*. Princeton Uni. Press: 1994.

Mead, G. H. *Mind, Self and Society*. Chicago: 1934.

Ortony, Andrew and Clore, Gerald L. Emotions, Moods and Conscious Awareness. *Cognition and Emotion*, 1989, 3 (2), 125-137.

Plato. *Symposium*. Loeb ed.

Plutchik, R & Kellerman H. (ed.) *Emotion. Theory, Research and Experience*. Academic Press: 1980.

Quine, W.v. O. Two Dogmas of Empiricism. *The Philosophical Review* 60 (1951) pp 20-43.

Rorty, Amélie Oksenberg (ed.). *Explaining Emotions*. University of California Press: 1980.

Rousseau, J. J. *A Discourse on Inequality*. Penguin Classics: 1984.

Russel Hanson, Norwood. *Patterns of Discovery*. Cambridge: 1961.

Ryle, Gilbert. *The Concept of Mind*. London, 1949.

Searle, John. *Intentionality*. Cambridge Uni. Press, 1983.

Searle, John. *Minds, Brains and Science*. Harvard Uni. Press, 1984.

Searle, John. *The Construction of Social Reality*. New York: 1995.

Searl, John. *Mind, Language and Society*. New York: 1998.

Shaver, Schwartz, Kirson & O,Connor. Emotional Knowledge: Further Explanation of a Prototype Approach. *Journal of Personality and Social Psychology*. 1987, Vol 52, No 6. 1061-1086.

Sherman, Nancy. *The Fabric of Character. Aristotle's Theory of Virtue*. Oxford Uni. Press: 1989.

- Smith, Adam. *The Theory of Moral Sentiments*. Oxford University Press: 1976.
- Sofokles. *Antigone*.
- Solomon, Robert C. *The Passions*. Indianapolis: 1993.
- Taylor, Charles. *Human Agency and Language*. Philosophical Papers I. Cambridge: 1985.
- Taylor, Charles. *Sources of the Self*. Cambridge Uni. Press: 1989.
- Taylor, Charles. *The Ethics of Authenticity*. Harvard University Press. 1991.
- Weber, Max. *Wirtschaft und Gesellschaft*, Studienausgabe, Tübingen: 1980.
- Young, Iris Marion. *Intersecting Voices*. Princeton University Press: 1997.
- Zajonc, R. B. Feeling and Thinking. *American Psychologist* 1980, Vol 35, No. 2. pp 151-175.

