REPLY TO SCHROEDER, CLARKE, SEARLE, AND QUANTE

Jürgen Habermas
Translated by Joel Anderson

I would like to thank my colleagues, who have been familiar with the debate on free will for longer than I, for their instructive criticisms. What got me working on this topic was the public attention that neuroscientists had been receiving on these topics. My aim was to bring together arguments against the tendency to jump to philosophical conclusions from a successful and undisputed scientific enterprise. Hard naturalism is not science, but philosophy—and not a particularly good philosophy. A research strategy that takes its lead from physicalism allows, at best, a wager on the correctness of the deterministic thesis. It is in this connection that I also criticize the program of criminal law reform, which is supposed to lead to the elimination of the concept of guilt.¹ This was the occasion for my philosophical confrontation with approaches that aim either to downplay the problem of determinism, along the usual compatibilist lines, or to eliminate the problem, along the lines of ‘hard’ or scientistic naturalism. What I see as mistaken about the ‘scientistic’ approach is its objectivist attitude, which ignores the epistemic relevance of the participant’s perspective. This is why I suggest an epistemological turn in the free will debate, which must not get stuck, however, in an unsatisfactory dualism. I indicate, at the end of the essay, a possible alternative to the prevailing physicalist image of the world. But I leave no doubts as to the tentative character of these latter considerations.

This train of thought suggests the sequence in which I take up, first, Timothy Schroeder’s critique of my phenomenology of free will (Section 1), then address Randolph Clarke’s defense of compatibilism (Section 2), go on to engage critically with John Searle’s concise criticisms (Section 3), and finally debate with Michael Quante regarding the meaning of ‘determinism’ (Section 4).

1. Reply to Timothy Schroeder

Timothy Schroeder misunderstands my phenomenology of free will as a ‘theory.’ My view is that philosophical work starts with the task of reconstructing the intuitive knowledge that actors and speakers performatively employ. For that, we have to adopt the perspective of the participants who take part in the relevant practices, instead of doing as if we were scientists—amateur psychologists, say—whose business is to develop research hypotheses. The three ‘principles’ that Schroeder critically examines are what results from an analysis of the semantic content of the presupposition of free will that actors reciprocally attribute to one another when asking for an account of their actions.
The philosopher who describes communicative actions operates at an analytic level different from that of the psychologist who explains mental states. The philosopher takes into consideration the further point that those reasons, which often acquire action-motivating force for an individual actor, are not just ‘something in the head’ but are the same reasons that are embodied in cultural traditions, anchored in institutions, and processed in communication. With ‘culture’ and ‘society,’ dimensions of an ‘objective’ mind are considered—‘objective’ in the sense that here the mind is embodied in symbols. This kind of a symbolically materialized mind eludes a mentalistic view. I conceive of persons as beings who exist ‘in the space of reasons’ (Sellars) and are receptive to reasons. Having made these introductory remarks, I turn briefly to Schroeder’s objections to the three supposed ‘principles.’

(a) For the attribution of a free will that is presupposed in the language game of responsible agency, it doesn’t really matter to what extent the actual preparation of any action depends on explicit reflection. Prima facie, all motives for actions count as attributable reasons for acting, whether we are speaking of habits and dispositions, spontaneous moods and vacillating preferences, or the conscious results of deliberation. The tacit assumption is that the reasons for which a person performs an action are always his or hers, whether they actually emerge from reflection or not. Of course one can’t choose one’s own personal character traits; they are the product of nature and nurture, in the sense of life-history and socialization. But when adults ask to be excused for an offence by appealing to such personality traits and thereby seek to distance themselves from causally relevant features of the character that they have in fact acquired, they can expect to meet with a sympathetic response only in exceptional cases of, say, pathogenetic personal or social circumstances. Normally, we identify persons with their character and hold them responsible for it, on the (often counterfactual) assumption that they could have made up their mind to revise this or that problematic attitude. This is an observation of what people mean when they engage in a specific language game, not a psychological description of what simultaneously happens ‘in the mind’ of those people.

If we are not to talk past each other in the present discussion, we need to pay attention to what the standpoint is from which we are making our arguments. One can attribute revisionary power to the weighing of reasons only if one conceptualizes ‘reflection’ from the participant perspective as the performance of a person and not merely as an observable event. For Schroeder, processes of reflection come up for consideration only as events: ‘Every act of reflection starts somewhere, and the first thought that begins reflection is not chosen on the basis of reflection’ (Schroeder, p. 82). But wouldn’t we call that begging the question? Let’s modify his example: Dorothy notices that the person she told the joke was hurt by it, and she then regrets having spontaneously told it. That means that, upon reflection, she wishes she could undo what happened and resolves to act differently in similar situations in the future. By the way, such reasoning would, because of its behavioral impact, also raise the probability that her behavior will change in the future.

(b) The second criticism has to do with the presupposition that one could have done otherwise, which is attributed to an agent as a background assumption and which the agent takes for granted performatively, that is, in the course of deliberating and acting. The premise underlying one’s deliberations is not that one is not prevented from acting otherwise, but that one is able to do so. Without this premise, it would be meaningless, on conceptual grounds, to weigh reasons pro and con. What Martin Luther expressed at the Diet of Worms was an existential belief. His ‘I-can-do-no-other’ was the result of a
conscientious process of doubting. And, again, Luther was engaged in that process of deliberation only on the presupposition that the outcome of his wrestling was open. It is as the product of a reflective process of formation [Bildungsprozess] that the resulting decision to break with ‘Rome’ demonstrated personal freedom. For ‘freedom’ means that we can subject our will to reasoned opinions. In the case of ethical convictions, it sometimes even seems to us that we commit ourselves irrevocably.

(c) The third criticism is based on an example that is incompletely described. Granted, Sam has to get his work done, but why does he isolate himself from his friends? In order to resist his social proclivities, which keep distracting him from his work? Then Sam shouldn’t be surprised when his good intentions run the risk of failing again. Or is it in order to escape bothersome social obligations that keep him from getting to his work? In that case, Sam learns something about himself: he discovers that what keeps drawing him away from his work is not just these obligations but rather an underlying need for social contact. In the one case, Sam is aware of his conflicting motives, but he ‘lacks the strength’ to follow his long-term interests at the expense of short-term gratification. In the other case, he becomes aware of an unconscious motive that (we assume) can then be ‘worked on’ and influenced in reflection. Unconscious motives are sometimes the stronger motives, but often not the more rational ones. The suppressed intuitions can, however, sometimes turn out to be the more reasonable considerations. Then one benefits from the search for unconscious motives that explain [erklären] why these intuitions got no uptake [sich nicht zur Geltung bringen], for example, an obsessive work ethic that threatens to ruin Sam’s personal relationships.

2. Reply to Randolph Clarke

Randolph Clarke presents three arguments: he defends classical compatibilism, disputes the notion of freedom as absolute, and calls for a naturalism that is supposed to dissolve the puzzling normativity of the mind in an empiricist manner (apparently along the teleosemantic lines).

(a) Clarke begins by emphasizing that the presupposition of free will is relevant only for the formation and execution of action intentions. It has nothing to do with the spontaneous emergence of desires and emotions (which can, of course, be revised in the course of practical deliberations). The same goes for thoughts that merely occur to us. These observations don’t contradict the tight connection that, I claim, holds between freedom of the will and responsiveness to reasons. The fact that persons move within the space of reasons is a necessary condition for their subjecting their will to practical insights and, in this sense, acting ‘freely.’ Clearly, every case of forming an explicit judgment involves the kind of ‘rational motivation’ that is associated with taking a ‘Yes-’ or ‘No-’position regarding fallible validity claims. But the judging subject is well aware of the difference between heeding the ‘non-coercive force’ of the better argument and being subjected to the causal efficacy of a chain of events in which the body happens to be involved.

Clarke then takes the compatibilist position that we can engage in practical deliberations and act responsibly without having to presuppose the possibility of being able to act otherwise. In order to counter my objection that compatibilism is at best a truth about agents rather than a truth for them, Clarke proposes a variation on Frankfurt-style counter-examples that removes the distinction between the participant and observer perspectives. This time, the person who decides to do A knows about the mechanism that would
intervene in case a different decision is made and would ensure that A is performed: ‘Suppose that she wants to perform the deed anyway, and does so on her own, utterly unmotivated by her awareness of the would-be intervener (p. 53).’ Of course, this person would also be responsible for her action and the consequences—let’s assume that A leads to the negligent violation of another’s property rights. But one implication of the modified thought experiment makes it particularly clear that the person could have performed her action only on the assumption of being able to act otherwise: the person who stands accused for the consequences of her property offence can now introduce the intervening mechanism that she knew about in court as a mitigating circumstance: this external circumstance evidently restricted her freedom of action. But we can speak of restrictions in the freedom of action only on the assumption that there is freedom of will. External circumstances can restrict our range of free action [Handlungsspielraum] only if we can take it for granted that, within this range, one can act in this or that way.

(b) I share Clarke’s qualms about the incoherence of the notion of an unconditioned or intelligible will that intervenes, from the outside, in the world of nomologically determined phenomena. However, the allegedly exclusive choice between determinism and the assumption of unconditioned freedom is incomplete. If we conceive of free actions as rationally motivated actions that can be explained by reasons, we locate free will within a nexus of conditions. The language game of responsible agency operates with the presumption that freedom is conditioned. The actions we describe as intentional merely obey a different mode of causation than nomologically explicable events, but they are nothing otherworldly: they are not ‘anomalous’ or rule-less phenomena. The reasons taken up in rational motivation obey logical-semantic rules. That is why the reasons that serve to justify actions are, under appropriate circumstances, sufficient to explain them; in any case they are contributing parts of an explanation. One way or another, the social sciences also make use of this model of causation: social structures are not simply imposed blindly on behavior but rather develop their causal force via the agents’ interpretations.

(c) Finally, Clarke confronts my admittedly speculative proposal—for combining epistemic dualism with a unified image of the universe—with the counterproposal that insists on the image of a world that is described in pluralistic terms but is conceptualized in physicalist terms. Considered ontologically or in itself [an sich], the world is composed of elementary particles, but we can—depending on various interests and the available vocabulary—choose to describe the same objects in different ways and give, for example, micro- or macrophysical, biological, or folk-psychological descriptions: ‘There are no levels of reality in this picture, only levels of description . . . It is, in each case, simply a matter of there being a single thing variously described’ (Clarke, p. 55). I don’t see how the problem of overcoming epistemic dualism is supposed to be solved by the undisputed diversity of perspectives from which the same ordinary object—a chair, say—can be described as a bundle of particles or waves, as chemical bonds, as material of a specific sort, as a brown and heavy object, as Biedermeier furniture, as something manufactured in the workshop of a Viennese craftsman around 1840, as an exemplary expression of bourgeois lifestyle, and so on.

The hierarchical organization of perspectives as ‘levels of description’ suggest that neurological and folk-psychological descriptions of the same behavior can be related to each other conceptually and empirically in much the same way as in the case, for example, of the microphysical and macrophysical description of an object that expands when heated. As long as we are satisfied with the peaceful coexistence of various explanations and leave aside the question of whether a description can be reduced to
another, the question of ontological monism remains undecided. Because Clarke defends a stronger—indeed, physicalist—position, he seems to want to assimilate conceptually the semantic relations between intelligible symbolic expressions to empirical relations between observable mental states. A conceptual reduction of mental to bodily phenomena has not succeeded until now and an empirical reduction remains, as Clarke himself says, an article of faith: ‘I am not sure that anyone has provided, or will do so, a satisfactory naturalist account of normativity, but I see little reason to think that scientific explicable is incompatible with it’ (Clarke, p. 56).

3. Reply to John Searle

My combative colleague John Searle provides a description of the initial issue that I agree with. Because he helps himself to an intentionalist vocabulary in the process, however, he occludes the intersubjective dimension of free will from the outset. There is an internal connection, though, between the intuitive sense of freedom that accompanies all our actions and the reciprocal obligation, under certain circumstances, to provide others with an account. The neglect of this dimension becomes noticeable when Searle—like Clarke—treats epistemic dualism as merely a matter of varying levels of description.

Searle’s first critique is directed against the dualism of language games and the corresponding epistemic perspectives: surely, he suggests, it is the same processes that we describe sometimes as semantic content or mental states and other times as neurological processes. Even if that were objectively correct, we could be sure that we are dealing with descriptions of the same processes only if we could translate equivalent statements from the one language into the other (and not just establish empirical correlations between them). The thing is, however, that the language we employ for psychological processes and semantic matters cannot be reduced to physicalist or behaviorist language.

Consider the simple case of an aesthetic description of a picture and the physicalist description of ‘the same’ picture (its frame, the canvas, the paint, etc.). In what sense do the aesthetic and physicalist descriptions refer to the same thing? If we take a scene from Turner, on the banks of the Rhine, with Cologne’s cathedral in the background, the content of the image that is amenable to aesthetic evaluation does not of course have extension within the picture frame that can be identified in time and space. Rather, the amazing, early impressionistic qualities of the painting ‘extend’ via an historical time period and a semantic domain, within which it first becomes possible to compare and contrast it with the painting style of the contemporary Constable, on the one hand, and of the several-decades-younger Paris impressionists, on the other. The space of stylistic qualities clearly cannot be mapped onto the physically measured space bounded by the picture frame.

Matters are similar in the case of exchanging thoughts during a discussion, for example, about freedom of the will. Isn’t it silly to think that this exchange of arguments (which go back into the 17th century and have hardly been enriched with new arguments since the middle of the 19th century) is played out solely in the brains of the participants? The ongoing process of cultural transmission [Überlieferungsgeschehen] with which our discussion connects up has materialized over a long historical stretch of time in numerous books and journals in various languages, whereby the form and linkages among these symbolic expressions follow rules that are, in the broadest sense of the term, ‘grammatical.’ It is undisputed that no painterly nuance and no statement can come about without mental and bodily operations and that all mental processes are ‘realized’ via the biological
substratum of brain processes. What is disputed is whether the neurobiological description of brain processes will one day be sufficient to explain or partially predict the precise corresponding mental states along with their semantic content.

I decidedly resist the accusation that my doubts about the philosophical interpretation of an established research strategy has anything to do with an intention to ‘block’ the interesting research stemming from this strategy. I have no doubts that it will lead to valuable results. Obviously, mental processes have long been the subject of neurological research. My doubt is restricted to the viability of the philosophical argument that one will be able to explain mental phenomena on the basis of the law-governed interaction of its physical and biochemical substratum, in line with the familiar nomological model, until a condition has been met that, up to now, has been met by every successful reductionist explanation: in the present case, a unified vocabulary for mental operations and brain states.

It seems to me that Searle is the one with the strange conception of neuroscience, when he says, on the one hand, ‘that all of those processes are grounded in the more fundamental phenomena of physics,’ although he admits, on the other hand, that ‘it is extremely unlikely . . . that we will be able to find type–type correlations between the mental phenomena that interest us, such as memory and perception, and the phenomena of atomic physics’ (Searle, p. 72). If neurobiology were not to discover any law-like connections between physical and mental phenomena (the possibility of which Donald Davidson disputed, and for good reasons) but were still to remain committed to the hypothesis that the brain is a deterministic system that entirely determines conscious phenomena, we would have to postulate something like causality without natural laws. If there weren’t any nomological connections between body and mind, we would need an explanation of what it could mean for the one to give rise to the other ‘causally.’

There is another point at which I run into a similar unclarity. Mental causation is a trivial phenomenon for Searle, because he takes it for granted that the higher-order phenomena of consciousness (and culture) can be conceptualized as emergent components of the brain—‘emergent’ in the weak sense. An intention to act can ‘cause’ (in the sense of nomological causation) the corresponding arm movement (which can be described at the neurological level as a physical–chemical event) to the extent to which the emergence of this intention itself is explicable as a neuronal event: ‘Of course the whole system only works because the so-called higher levels are grounded at the lower levels, at the level of neurons, synapses, and all the rest of it’ (Searle, p. 75).

While this formulation still admits of several interpretations, Searle’s car engine example suggests the reading I’ve just given: complex properties (the functions of spark plugs and cylinders) emerge from constellations of basic components. Whereas the interaction of the basic components can be explained nomologically (the combustion of carbon molecules), those constellations become relevant only at the higher level of the car as a system. However, such a reduction of weakly emergent properties will similarly pave the way from neurons to consciousness (and to culture) only if what happens at the system’s level of the mind can be broken down into its neurological components and represented, with the help of a theoretical model, in a way that explains how what happens at the level of the system is generated by the nomological interaction of these elements themselves. But what would the language be in which this model could be represented, if ‘we know that mentalistic phenomena are . . . irreducible’?

Searle appears to have an alternative in mind that provides him with a way of dealing with the problem of translating from an intentionalist into a physicalist vocabulary. From
the outset, he seems to opt for an expanded notion of ‘the physical,’ which includes the mental as a property of higher levels of organization (of the physical itself). What is unclear to me is how this kind of philosophy-of-nature approach to such an encompassing concept of the physical is to be squared with a scientific realism that identifies ‘nature’ with all of the object domains of the established natural sciences. This question arises regardless whether one hopes to provide a nomological explanation of neurological events with the help of deterministic or statistical laws.

Searle thinks that, if we proceed from the premise that the brain is a ‘deterministic system,’ it might well turn out that the brain operates non-deterministically. In that case, the brain would represent a system that (if I understand this correctly) permits ‘real gaps’ to arise at the level of the emergent mind: ‘That is, each stage of the neurobiological process is not sufficient by itself to determine the next stage by way of causally sufficient conditions’ (Searle 2005, 230). I’m amazed by this revision of his earlier position, which still seems plausible to me. I don’t see how freedom of the will can be explained by chance. And I have just as much trouble seeing how ontological monism fits with the idea that chance gaps in quantum events are going to be filled in with spontaneous interventions (?) of free will.

A final criticism can be resolved in short order. For we don’t need to fight over the use of the expression ‘performative.’ For example, we cannot state that ‘the brain’ or the neuronal system ‘believes,’ ‘feels,’ or ‘intends’ something without committing a performative contradiction. For the implicit content that we presuppose to be true in the act of asserting is incompatible with the propositional content of that statement: whenever we make any sort of statement we presuppose that the truth claim for the asserted proposition is being raised by someone who takes the role of the first-person singular—and not by a brain. On the other hand, there is no disagreement that the detection of such performative contradictions deliver arguments with limited scope of application. As I said in my paper, they can’t defeat a determinist who claims that a reductionist research strategy will one day prove that free will is an illusion. Searle underestimates, however, the problem of the performative limits of self-objectification when he brushes aside my critique of the objectivist attitude of scientism (that is, its blindness for the epistemic relevance of the participants perspective) with the remark that this is probably a matter of my confusing epistemology and ontology.

4. Reply to Michael Quante

Michael Quante begins his incisive contribution by redirecting the course of the discussion. He takes the two theses that I see as sufficient for defining the scientistic position, and then adds a third proposition in order to provide a more comprehensive definition. Thus I feel prompted to start with a terminological remark. Scientific realism is inspired by the image of a materialistically composed and causally closed world. This generates the deterministic thesis: the world is structured in such a way that it is possible in principle, assuming sufficient knowledge, to explain or predict the occurrence of any given states or events, on the basis of deterministic or statistical laws. By contrast, the classical principle of sufficient reason [Satz vom Grunde], that all states and events in the world have a cause, represents a weaker claim. The explicability of things isn’t biased in favor of one particular model, that of the nomological pattern of explanation (for which modern physics serves as the standard). When it comes to actions—which, in connection with the problem of free
will, are at issue as explananda—rational and hermeneutic explanations, for example, also count (in combination, as needed, with assumptions on regarding circumscribed empirical constraints and ‘mechanisms’). If Quante wants to call this weaker view ‘determinism,’ then I am a ‘determinist.’ But in my paper, I employ that term in the specific and expanded sense of ‘determined on the basis of natural laws.’ This view depends on the objectivistic fallacy that leads to the reification of the observer perspective always trumps the participant’s perspective. I don’t understand how I thereby create an ‘incompleteness problem’ for myself. For, as Quante himself notes, the presupposition of free will conflicts only with the thesis that all behavior is determined by natural laws.

I think it is inconvenient to depart from established terminology by loosening the meaning of ‘scientism’ and extending it to all approaches that start out from the simple aim of looking for an explanation of anything that happens in the world (without a simultaneous commitment to a specific—namely the nomological—model of explanation). If we were to connect the term ‘scientism’ merely with the global principle of granting the established academic disciplines a monopoly in settling controversies in the realm of secular knowledge, then I would be a defender of scientism—and the Pope probably would be too. The further terminological proposal to understand the dualism of epistemic perspectives as a kind of ‘incompatibilitism’ also seems not particularly helpful. I just want to maintain that the two corresponding language games cannot be reduced to one another.

Quante correctly characterizes my objection to compatibilism of the Frankfurtian variety. His metacritique leaves me unconvinced, however, for his suggestion that we differentiate between the observer perspectives of scientists and of philosophers isn’t likely to save this reading of compatibilism either. ‘Determinism,’ as I use the term in my paper, refers to a conception that has a deep impact on our self-understanding as responsible agents, whether the one who defends it is a neurobiologist or a philosopher. Unless, of course, the philosopher offers—in contrast to the determinist position—a comprehensive explanation of the universe that is compatible with our ordinary self-understanding. This leads to the question that challenges Quante and me in equal measure, as to whether there is an alternative to scientistic naturalism. Without in any way putting into doubt the authority of the natural sciences, an acceptable alternative would have to bridge the gap between the ‘nature’ objectified by these natural sciences and the ‘culture’ we are familiar with by ‘inhabiting’ it.

Neither the humanities nor the social sciences achieve explanations of actions and interactions on a theoretical basis that allows for strict conditional predictions. This is a feature of all rational or interpretive explanations, because they move within a conceptual frame that leaves intact the propositional attitudes of persons and the openness of their practical deliberations. The same holds for another type of explanation, namely those from developmental psychology, when the theoretical frame preserves the conceptual linkage to either the rational reconstruction of cognitive competencies (Piaget) or to the biographical self-understanding of adults (Freud). Of course, all of these explanatory schemes remain limited to the cultural realm. The alternative we are looking for would consist in a type of monism that includes culture as a region of the natural world without squeezing it into a reductionist pattern of explanation.

Quante confronts me with the interesting question regarding the conceptual consistency of the weak naturalism that I propose as an alternative to both hard naturalism and a metaphysical philosophy of nature. Since we are dealing with a speculative area here, we can at best assess the prospects of possible theoretical strategies.
The second of the options that Quante offers me coincides with the Idealist approach just mentioned and is unacceptable because it flouts scientific knowledge. Quante rightly favors the third option, an approach based on developmental history. This would have to provide, along the same lines as the first option, a unitary perspective from which the evolutionary transition from nature to culture can be explained in a way that would promise to integrate all relevant empirical knowledge. I don’t deny the paradoxical flair of the move towards detranscendentalizing, via natural history, the most general features of symbolically constituted forms of life (as we reconstruct them ‘from the inside’).

The objectivistic fallacy of scientism has its roots in the reification [Verselbständigung] of the perspective of the observer. Philosophical reflection can also get entangled in the snares of objectivism. The counterpart to scientism is to be found in an Idealist philosophy of nature that reflects anew, from the first-person perspective, upon the objectivated entirety of Geist (‘mind’ or ‘Spirit’) as it returns back to itself from nature. Schelling and Hegel assimilate natural evolution to a process of self-reflection, in the course of which the human species recaptures its own genesis. If one knows, however, that one has to avoid both of these metaphysical temptations and instead seek to link up with empirical knowledge from evolutionary biology, perhaps one can avoid the dilemma that Quante poses. A non-Idealist natural history of consciousness and culture will have to reflect on the respective merits of what we get to know from both, the observer and the participant’s perspective. Empirical knowledge must still retain its veto power, even when we filter it through the lens of a critically appropriated history of science. I combine this with a pragmatist reading of the nomological approaches in natural science. This view allows us to understand scientific claims in realist terms, without requiring us to identify the ‘nature’ of a future natural history of culture with the ‘nature’ of natural science. The price one supposedly has to pay for this move—of interpreting the natural sciences in pragmatist terms, without sacrificing the claim to realism—will no longer seem to cost one anything, as soon as one leaves scientism behind.

NOTES

1. This concern is not, as John Searle claims, a distraction from the real theoretical issue. If we locate the presupposition of free will properly—not in the introspective reassurance of first-person, intentional consciousness but rather in the language game of responsible agency—the criminal law discourse provides a particularly apt example to analyze, because it renders precise the component aspects of the attribution of responsibility. The intentionalist model of arbitrary arm movements (or ‘Buridan’s ass’ choices between equally preferred options) confuses matters, because it artificially separates the act of will from the rational motivation for these bodily movements.

2. See the outstanding discussion by Bieri (2001, chapters 6 and 7).

REFERENCES


Jürgen Habermas, Ringstraße 8b, 82319 Starnberg, Germany.