Autonomy and authenticity are usually seen as largely individual matters. It is often thought that, in trying to understand what it means genuinely to lead one’s own life, we start by trying to get a handle on what individuals themselves want, and then examine the ways in which social pressures, political domination, material deprivation and so on interfere with them in ways that threaten or distort the self. This is the picture we encounter in Locke and Hobbes, right on through to contemporary liberalism. And, with some exceptions, it remains the dominant view.

It is decidedly not Jürgen Habermas’s view. He takes autonomy, agency, identity, authenticity and the self to be fundamentally intersubjective phenomena. Moreover, even in comparison with relational or social approaches to autonomy, he conceptualizes this social dimension differently and in ways that cut deeper than the now widely accepted idea that our identity is shaped by our surroundings or that many of our deepest convictions can be realized only together with others. As he remarked in recent autobiographical reflections, his early childhood experiences with surgery on his cleft palate, and the frustrating, humiliating difficulties in making himself understood through his speech impediment, “opened [his] eyes to the intersubjective constitution of the human mind and the social core of our subjectivity, as well as to the fragility of communicative forms of life and the fact that socialized individuals are in need of special protection” (BNR: 17). If, like Habermas, one is convinced that our fundamental vulnerability lies in the possibility of being excluded from the protective and enabling forms of social cooperation and mutuality (or of having them break down), it will
seem very strange to take the perspective of an independent individual as the self-evident starting point for reflection on autonomy and the self.

At the same time, Habermas has a deep distrust of groupthink or collectivism and stands squarely in the Enlightenment tradition that defends individual freedoms against domination.¹ Again, his personal experience is relevant, specifically, of seeing how Nazi ideologues and post-war German conservatives have sought support for their policies in philosophical theories that treat the individual as a mere “effect” of social or cultural forces, or as an expression of tradition, Spirit, language or Being.

This dual emphasis on the social (or “intersubjective”) and the individual makes for a rich conception of autonomy: Habermas is a staunch defender of individual autonomy, authenticity and self-determination, but not on the basis of the standard liberal empiricist understanding of the self-sovereign self. To appreciate the complexity without getting lost in the complications, it will be useful to start out with a few distinctions, to which I now turn.

Fives senses of autonomy in the wake of Habermas’s rejection of philosophy of consciousness

Habermas’s views on the self and autonomy are not as widely discussed as many other aspects of his philosophy, in part because there is no specific terminology he uses for the cluster of related positions he takes.² One can, however, distinguish five distinct theoretical contexts in which Habermas discusses autonomous agency and authentic identity: deliberative democracy, moral philosophy, free will, social theory and personal identity. One way to get a first approximation of these different dimensions is by contrasting them with states of persons who lack these dimensions of autonomy. To lack political autonomy is to be subjected to illegitimate domination by others, specifically by not being integrated in an appropriate way in processes of collective self-determination. To lack moral autonomy is to be incapable of letting intersubjectively shared reasons determine one’s will. To lack accountable agency is to behave as the result of compelling forces rather than to act for reasons. To lack personal autonomy is to be unable to engage in critical reflection about what to do with one’s life. And to lack authentic identity is to have one’s claim to recognition vis-à-vis others get no uptake.

Each of these senses of autonomy contribute to the overall conception of autonomy in Habermas’s work, but the relations between them are complex. My aim here is to show how they all illustrate several
consistent Habermasian themes – intersubjectivity, performativity and historicity – all of which are key components of Habermas’s project of developing an account of autonomous agency that moves behind the limits of what he calls the “philosophy of consciousness”.

Moving beyond “philosophy of consciousness” (or, as he sometimes says, taking a “postmetaphysical” approach) is not a matter of dismissing subjectivity and consciousness as legitimate topics of philosophical investigation. What Habermas is rejecting, rather, is a broad philosophical approach that first emerged prominently with Descartes, an approach that considers the key philosophical questions, such as how we ought to relate to others, what really exists and how we can know anything, to be best answered by focusing on how one’s mind and will grasp the world. The paradigm example of this is Descartes’s discovery that the one thing that he could not doubt was the idea that there existed, at a very minimum, some subject who was doing the doubting (Cogito ergo sum: I think, therefore I am). Habermas has written extensively on why this approach leads to various philosophical dead ends, how the so-called “linguistic turn” undermined the fundamental premises of the approach, and what the implications are of rejecting the philosophy of consciousness. For our purposes, the key idea is this. In analysing our practical dealings with the world, philosophy of consciousness (and most mainstream analytic philosophy today) tends to focus on understanding how an individual’s mental states (beliefs and desires) generate effects according to general causal mechanisms and do so in a way that can be fully grasped from the point of view of an outside observer. According to Habermas, this way of framing the task of understanding autonomous agents and their actions is still stuck in “philosophy of consciousness” in overlooking three key aspects of autonomous agency: intersubjectivity, performativity and historicity.

**Intersubjectivity**

Intersubjectivity is perhaps the most central component of overcoming philosophy of consciousness and especially the guiding assumption that there is something located in the individual – motivation, reflection, desires, conscience, subjective experience and so on – that is prior to and ultimately independent of the social world it affects. This “subject–object dichotomy” is precisely what Habermas targets, as presupposing that there is an inside and an outside: an inside source that we can know from a “first-person” perspective and an outside world of effects, including other subjects, that we can study from the third-person perspective of the outside observer. Habermas’s alternative takes intersubjectivity to
be basic, in the sense that agency, critical reflection, motivation and so on are what they are in virtue of being elements of our interactions with others. For Habermas, the starting point for philosophical investigation of autonomous agency in particular is to be located in communicative interactions, democratic institutions, practices of giving and asking for reasons, and the informal background “lifeworld”. Consider a case I take up in the discussion of free will: when one acts for reasons, a full explanation of what one is doing must make reference to the cultural and linguistic background in virtue of which certain noises count as giving reasons. But this cannot be adequately grasped, according to Habermas, on the model of causal interactions among physical events, since reasons, objections, counterarguments and the rest are phenomena that, although perfectly real, come into view only within the “space of reasons”, to use the Wilfrid Sellars phrase that Habermas has taken up recently, especially in his debates with Robert Brandom. Thus, just as a “three-point basket” only exists within the meaningful space opened up by the game of basketball (and, in the US, only since 1979), “autonomy” is something that comes into view only once we take on certain attitudes in our interactions.

**Performativity**

Talk of attitudes is widespread, of course, within philosophy of consciousness, as elements in how an individual subject represents the world. In Habermas’s postmetaphysical approach, attitudes play a markedly different role, as *performatively undertaken presuppositions*. On this view, taking something to be a three-point basket, for example, is not best understood in terms of my pretending something because I predict that you will pretend it too. Rather, my taking something to be a three-point basket is an aspect of the engaged perspective of participants in the game of basketball. To be able to see a three-point basket, one has to be an “insider” to the game of basketball, even if one is participating only as a spectator. This perspective or attitude is not so much a way of *thinking* about the world as a way of *doing* something in the world, hence something “performatively undertaken”. The idea that certain phenomena have their reality as part of social practices conflicts with the influential physicalistic assumption that the only things that are real are those things that can be perceived from the detached (“third-personal”) perspective of the natural sciences. And, indeed, Habermas’s rejection of philosophy of consciousness is linked to his defence of an inclusive alternative to “hard”, eliminativist naturalism. For our purposes, it is important that this move to inclusive naturalism allows room
for situating ideals and counterfactual presuppositions as part of the reality that shapes agents and actions. In particular, even though we all know about one another that we are not fully rational or autonomous, it is nonetheless an unavoidable part of the practices within which autonomy is central that, in so far as we are engaged in those practices, we must presuppose one another’s autonomy. But this “actual role of performatively presupposed counterfactual assumptions” is something that can be grasped only once one moves from a philosophy of consciousness to the social pragmatism Habermas defends.

**Historicity**

Once our grasp of reality is understood not primarily as a matter of mental representation but of practical engagement, it becomes “de-transcendentalized” and thus something with a history. Habermas here stands in the tradition of Hegel and Marx (and, to some extent, Nietzsche and Foucault) in treating familiar topics of metaphysical philosophy, such as reason, morality or the mind, as real-world phenomena that emerge in particular social and historical contexts, rather than as timeless facts, relations or essences. This historicity occurs at two levels. First, these phenomena (including autonomy) are what they are now because of how they emerged, as part of complex real-world processes. In the case of autonomy, as we shall see, these developments are bound up with processes of modernization, involving changes in interpersonal relations, social institutions and material conditions. The second dimension of development occurs in the context of “life history”, that is, our individual biographical development, in which each of us acquires the know-how to be autonomous, or to reason, or to act morally. Since this individual process is deeply dependent on supportive social conditions, the historical character of an individual’s autonomy is further intertwined with autonomy’s intersubjectivity.

**Private autonomy, public autonomy and democratic self-governance**

The tension between individualistic and social understandings of autonomy is particularly clear in the context of political philosophy, and that makes political autonomy a good topic to take up first. On the one hand, there is the familiar sense of autonomy as a matter of being free from illegitimate interference in one’s choosing what to do. This is what Habermas calls “private autonomy”, which he sees as an indisputably
important principle. But it is complemented by “public autonomy”,
the form of self-governance that one exercises together with others in
authorizing laws and other forms of collective action. Ensuring public
autonomy is a matter of ensuring that those subject to collective deci-
sions, laws and so on can see themselves as its authors, as in the case
of democratically adopted laws. Importantly, this is not about getting
one’s way or not being interfered with. If I participate in a fair and
inclusive debate, but the vote at the end goes the other way, I may be
unhappy about my position not winning out, but my overall political
autonomy is still preserved, unless there are grounds for criticizing the
procedures. Traditionally, public and private autonomy have been seen
as opposed. Habermas see them as not only important and compatible
but as actually presupposing one another and emerging jointly. Neither
can be assumed to have primacy over the other; they are “equally basic”
[gleichursprünglich]. Public and private autonomy come into existence
and develop only by dovetailing. Without a legitimate political order
that guarantees individual rights and liberties, private autonomy simply
does not exist. To sum up a complex argument, autonomy is something
that emerges together with social practices, political institutions, legal
rights and so on, but this is not merely a point about circumstances that
are favourable for its development. It is not that autonomy becomes
difficult without the framework of rights or the institutions of demo-
cratic decision-making; it ceases to exist. Autonomy is thus a social
construction to the very core.

Here, again, we see the implications of rejecting the individualistic
metaphysics of philosophy of consciousness. I will mention two. First,
the emphasis on the intersubjectivity of political autonomy underscores
the importance of maintaining those social, cultural and material condi-
tions that make it possible. As Habermas regularly emphasizes, when
social trends and public policies undermine those conditions (roughly,
what he terms “lifeworld resources”), they cut deep into the very possi-
bility of autonomy. Second, as a social construction, political autonomy
is thus always caught up in ongoing processes of historical transfor-
mation. For example, as members of modern societies come to view
one another as entitled to welfare rights, the substance of autonomy
shifts as well. Looking forward, one can then ask, as Habermas does,
what changes in these conditions are necessary for ensuring that both
autonomy and democracy can develop in ways that make them able to
handle the new challenges served up by a globalizing and increasingly
complex world (PDM: 344).
Moral autonomy, reason and the will

Even this brief treatment of democratic self-governance highlights Habermas’s emphasis on the interdependence of individual and collective self-determination. Moral autonomy is distinctive, however, in being the context in which the egoistic desire to do what one feels like is expected to submit most unconditionally to the demands of the universalistic point of view. And, indeed, on Habermas’s approach, moral self-determination becomes indistinguishable from determination by reason. In order to remain focused on issues of autonomy and the self, I will limit myself to highlighting briefly this notion of autonomy as alignment with reason and then go on to discuss two of Habermas’s particularly interesting claims on the scope of morality and on moral motivation.

Habermas is keen to emphasize the Kantian point that moral autonomy is a matter of bringing what one wills in line with what one sees oneself as having reasons to do. The morally autonomous agent is self-determining because she is not subjected to any alien force, including desires or passions that we find ourselves having but wish we did not have. Moral autonomy, then, is a matter of desiring, choosing and acting in the way that one has reason to do. “In Kantian terms, the [morally] autonomous will is entirely imbued with practical reason” (BFN: 164). For the fully morally autonomous agent, it will be irrelevant that she strongly desires to do something that she knows to be wrong. But having one’s will determined by reason does not undermine one’s self-authorship, especially once it is clear that “listening to reason” is a matter of engaging, as a full and equal participant, in the ongoing process of giving and asking for reasons. Because of this, moral autonomy is not merely compatible with being responsive to relevant considerations raised by others. It requires it.

In linking moral self-determination with determination by reason, Habermas again illustrates how far he is from approaches that link autonomy to empirical influence and effective control. But this move also raises the concern that Habermas’s theory is far from empirical human lives as well. To address this concern, two aspects of Habermas’s approach to morally autonomous agency are worth highlighting.

**Ethical-existential autonomy**

Soon after first developing (with Karl-Otto Apel) his discourse ethics in the 1980s, Habermas came to revise the initial dichotomy between the moral will (guided by reason) and the strategic will (driven by
preferences and assisted by means-ends reasoning). He now distinguishes an “ethical” form of reasoning and autonomy in which neither reason nor will has the upper hand (JA: 12; Habermas 2003b: 1–15).

In this domain of significant personal choices, an individual ultimately appeals to some sense of what she just finds herself caring about and what sort of person she herself wants to be (rather than what all rational agents must want, as is the case with moral autonomy), and at least as long as moral norms are not being violated, it makes no sense to require agents to step back from their deep personal commitments and consider their merit from the moral point of view. At the same time, Habermas emphasizes, there is also room for progress in understanding what one really cares about; sometimes we discover that we were mistaken about what is important to us; not that we change our mind, but that we correct ourselves, often in response to our exchanges with others. This possibility for error opens up the possibility for rational insight, which Habermas, in typical intersubjectivist fashion, understands as a form of dialogue. In this “ethical-existential discourse” (as he calls it) one engages in an open-minded give-and-take about what one really cares about and finds important and tries to make sense of one’s personal commitments and values by fitting them into a larger sense of broader values, personal relationships, life history and so on. The capacity, then, to engage in this process in a fruitful way is what one could call personal autonomy, in a sense I return to below.

**The growth of autonomy**

It is also key to Habermas’s approach that autonomy in the domain of morality and ethics is a real-world phenomenon resulting from ongoing historical processes. Here again we see the dovetailing of individual development (“ontogeny”) and social evolution (“phylogeny”). As a matter in individual development, moral autonomy requires that an individual advance successfully through processes of cognitive development, psychological maturation, interactive competences and individual socialization (MCCA). At the phylogenetic level of the moral community, development also takes place, and in two quite different regards.

First, what morality demands of us is not something fixed but something that evolves over time. Despite Habermas’s emphasis on moral universalism, consensus and objectivity, he draws a clear distinction between claims to truth (which refer to a reality that transcends the reach of human knowledge) and claims to moral rightness (which refer to the set of norms for governing our interactions that is best justified in moral discourse). From inside the performative perspective of
participants trying to figure out jointly what norms are justified, moral debates feel like debates over truth, according to Habermas. But this does not change the fact that morality is ultimately a human social construction (TJ: ch. 6). The valid moral norms legislated and internalized by morally autonomous agents thus represent our current best efforts in the ongoing process of learning to solve the moral challenges continually posed by life and raised in discourse.

The second point has to do with the context-dependent development of motivational resources and supports for moral agents. The basic claim is that my effective ability to act morally involves having access to various “lifeworld resources” that support us when sheer insight and willpower aren’t enough, and as a tough-minded sociologist, Habermas knows that it would be naïve and unrealistic to expect them to be enough. As “lifeworld” resources, these supports are part of the implicit background of our actions, and they are a remarkably diverse bunch. They include what Habermas calls “personality”: character traits such as self-restraint or considerateness, as well as good habits that have become “second nature”, all of which make it decidedly easier to be morally autonomous. The same can also be said for customs and taboos, which also relieve individuals of the burden of having to motivate themselves morally all the time. The comfortable ease of traditional, conventional mores comes at a price, however, for they function as they do only as long as they are uncontestable. With the shift to a modern “postconventional” social order, customs and habits lose their unquestioned authority and can motivate only if they are taken to be legitimate. In modern societies, enforceable law, justified in a way that fits with individuals’ political autonomy, comes to provide a primary mechanism for ensuring that people do the right thing. Of course, if you live in a society that has bad laws or were raised with bad habits, it will still be harder for you to act morally. This is why Habermas puts so much emphasis on the importance of developments not only at the level of laws but also at the informal level of culture and personality – on what he calls a “rationalization of the lifeworld”. In this sense, improvements in socio-psycho-political structures constitute improvements in one’s moral autonomy.

Accountable agency, free will and the space of reasons

We have already seen how Habermas’s rejection of philosophy of consciousness shapes his views on moral and political autonomy. But the implications are perhaps most striking in his recent interventions in free
will debates, particularly in his recent efforts to counter the increasingly widespread talk of neuroscience and psychology “proving” that free will is an “illusion”.12

In debates over free will, the central issue is whether what we do is ever truly up to us, given that everything that happens in the universe, including movements of our bodies or changes in brain states, is determined by prior states of the world in accordance with deterministic laws of nature. As it is typically framed, the question of autonomous agency is a matter of having a certain sort of causal efficacy, and we seem to be faced with a nasty dilemma: give up the idea that we cause our actions (as distinct from our brains causing our actions) or presuppose some sort of alternative form of causation in which agency operates outside the laws of nature. Since the latter amounts to something like voodoo, it is argued, autonomy and the conscious will turn out to be illusions.

As mentioned earlier with the example of three-pointers in basketball, Habermas’s proposed solution centres on what it means to do something as a participant within a social practice. Consider now chess. Everything that happens in a chess match could ultimately be explained in terms of microphysical causal chains. If we do that, then there are no pawns or knights or rooks, only bits of wood. Clearly, however, chess players are not suffering from an illusion or pretending in thinking that a bit of wood is a pawn. They’re not pretending it is a pawn; it is a pawn. Imagine someone grabbing the microphone at a chess tournament and dramatically announcing that the game of chess is a fraud, because “pawns don’t exist!” Such a person would be, as the saying goes, unclear on the concept. Like pawns, reasons for acting are perfectly real, even though (like pawns) they exist only as part of social practices in which we give and ask for reasons. But, assuming that free will is a matter of being able to respond to reasons, then there are few good reasons to doubt whether autonomous agency is a natural part of the (social) world.

That, in a nutshell, is Habermas’s view on free will, although there are many further details, complexities and implications that could be discussed.13 Here I shall focus on three aspects that are relevant for understanding Habermas’s core claim regarding the intersubjectivity of autonomy.

First, as he did with regard to democratic self-governance and moral autonomy, Habermas links free, autonomous agency to rationality and justification. This move places Habermas with Plato, Kant and others on one side of perhaps the most fundamental controversy regarding free action and autonomous agency. The other side is occupied by empiricists (from Hobbes and Locke to Russell and contemporary
neuroscientists), who connect autonomy and freedom with a causal power to make events happen. Habermas’s basic intuition is that this causal efficacy is not really the freedom we care about. Indeed, such a view seems unavoidably to make the causation of arbitrary bodily movement – “Lift your arm intentionally and spontaneously, at exactly the moment you feel like doing it” – into the defining case of free action. But these are, almost by definition, trivial cases, because they are supposed to be done for no reason at all. The freedom and agency we really care about, by contrast, are better typified by actions in which we believe that our choices matter to us, that we have reasons for them, and that they make sense to us. And that puts free agency squarely in the domain of aligning oneself with reasons.

Second, Habermas sees neuro-sceptics about free will as failing to appreciate its historical nature. From the standpoint of natural science, “acting for reasons” might seem like a respectable element in a scientific explanation; but it is clearly part of the natural history of our species. Like the existence of language centres in the human brain or “Sicilian Defences” in the game of chess, acting for reasons is something that has emerged in an ongoing process of historical development. Indeed, on Habermas’s view, the real challenge for empirical research regarding free will is to explain, in detail, the co-evolution of the space of reasons and the human brain’s capacity to engage with the cultural artifacts we refer to as reasons. Like the historical linguistics of, say, the Frisian language, understanding the evolutionary emergence of “acting freely for reasons” would involve exploring causal, developmental processes of a peculiar sort of entity, namely, something that can be studied only in so far as we have already acquired it. After all, just as you cannot study the evolution of Frisian without having mastered the language, you cannot study the evolution of the space of reasons unless you know how to make moves in the game of giving and asking for reasons. Thus a really serious science of free will would also have to account for itself: “The natural genealogy of the mind is a self-referential project; the human mind tries to capture itself [sich einholen] in comprehending itself as a product of nature” Habermas (2007b: 42). Indeed, by denying that reasons are part of the natural world, physicalistic forms of naturalism have difficulty viewing their own accomplishments as part of the real world.

Third, and perhaps most important, Habermas views the reality of free will as bound up with the reality of our social practices of holding one another accountable, taking each other to be praiseworthy or blameworthy, and expecting that others view us capable of justifying (usually) what we do. Like the assumption that the person on the other
side of the chess board knows the rules of chess, our belief in one another’s free will is part of the game, in this case, “the language game of responsible agency” Habermas (2007b: page?). This presupposition may be defeated, for example, in cases of mental disorders or compulsion, but the process of identifying nonautonomy can be undertaken only by those who are part of the practice of distinguishing cases of acting responsibly from cases of being unable to appreciate what one is doing. Since there is no way to determine whether someone can respond to reasons except from the standpoint of other reason-users (just as there is no way to determine whether someone can speak Frisian without appealing to the judgement of Frisian-speakers), autonomous agency is something that emerges only in the mutual attribution and contestation of that status.

Habermas’s discussion of autonomous agency (in the sense of free will) is also important in bringing out links with his understanding of science, humanity and the natural world. But, among his discussions of various senses of autonomy, it is an outlier in having little to do with the notion of gaining critical distance on one’s reasons and challenging, as it were, the rules of the language game. Habermas’s key idea in the discussion of free will is that the contrast with determinism lies in the ability to respond to reasons rather than physical forces. The link between autonomy and the power of critique is more central in his discussions of other senses of autonomy.

**Personal autonomy, social integration and modernization**

Alongside political self-determination, moral autonomy and accountable agency, we can also distinguish a form of autonomy usually referred to as “personal autonomy”. This has to do with the capacity to lead one’s own life, to determine what to do with the freedom one has. This is what we strive for with our children, that they acquire the capacity, in the process of maturation, to navigate the choice-situations that give shape to their lives: whom to love, where to live, how to contribute productively to society and so on.

Most contemporary theories of autonomy attempt to specify general criteria for what autonomy is, in terms of timeless and universal standards of reasons-responsiveness, internal cohesiveness, and so on. Habermas’s approach differs from these in emphasizing the socio-historical development of personal autonomy as a real social phenomenon. Autonomy, in this sense, is to be specified in terms of competencies that one needs for finding one’s way in an increasingly
complex world. As a real-world social phenomenon, autonomy emerges as a result of contingent historical processes, both within the life history of the individual and within the development of societies, particularly in processes of modernization.

Habermas takes his cue from Émile Durkheim, who linked the obligation to be a “person” – that is, “an autonomous source of action” – to processes of differentiation in society, whereby individuals increasingly get different roles that they must fulfil, in coordination with others:

Thus the advance of the individual personality and that of the division of labor depend on one and the same cause. Thus also it is impossible to will the one without willing the other. Nowadays no one questions the obligatory nature of the rule that ordains that we should exist as a person, and this increasingly so.

(Durkheim 1984: 336)\textsuperscript{17}

Habermas ties this point about the contingency and historical specificity of the imperative to be autonomous to the development of social structures. Because personal autonomy initially has its value and function in the context of navigating choices, the more complex a society becomes and the more choices individuals must face, the richer and more developed their autonomy has to be in order to be able to fulfil that function successfully. The upshot is that in more complex societies, not only must one (as Durkheim argued) be autonomous; there is also an expansion in what being autonomous involves.

This process of modernization continues in the form of what Ulrich Beck (1992) and others have termed “individualization”.\textsuperscript{18} As traditional restrictions have loosened regarding choices of occupation, marriage partners, religion, lifestyle and so on, contemporary individuals must make dramatically more decisions than their grandparents. The expansion in possibilities for choice brings with it an expansion in the responsibilities for choosing well. If one grows up in a highly conventional social world where one’s choice of occupation, spouse, religious affiliation and so on is set at birth by one’s place in society, there is not much autonomy to exercise and not much to be blamed for in terms of career planning. As these matters all become the responsibility of individuals throughout society, everyone needs to become autonomous, in this sense of developing the skills to make complex choices, guided by a clear understanding of what one really cares about.

Here again we see familiar themes in Habermas’s approach: personal autonomy is a real phenomenon that has emerged historically as a feature of the intersubjective world. In this dynamic process of societal
evolution, individuals need to develop certain autonomy competencies, which in turn facilitate a further growth in social complexity, leading to yet greater demands on individuals to orient their lives autonomously and to navigate this increasingly complex social world.

Two further dimensions of Habermas’s analysis of personal autonomy are crucial to his overall view: the idea that autonomy involves the capacity for relating critically to the demands placed on one and the idea that, as in the case of moral autonomy, social structures (the “lifeworld”) provide enabling support for the exercise of personal autonomy.

From conventional accountability to critical autonomy
For Habermas, autonomy is critical autonomy, which he understands as a radicalization of the ability to explain one’s choices to others. What we have seen so far is that social transformations require one to make more decisions and that one is held accountable for making good decisions. This accountability is not just a matter of cleaning up your own mess when you make bad decisions, it is also a matter of being accountable to others for justifying your choices. Although Habermas acknowledges that people grant each other a degree of “private autonomy” (legally guaranteed leeway to do as they please), he emphasizes that cooperating with others and seeing ourselves as autonomous requires being able to justify one’s choices. In one sense, one can do so by appealing to conventions (as when one decides to get married or go to church just because “that’s what one does”) or to one’s preferences (as in, “I’ve always pictured myself as the breadwinner for a family”). For complicated reasons, Habermas believes that this conventional form of accountability is inadequate:

The independent performances that are here demanded from the subjects consist of something different from rational choices steered by one’s own preferences; what these subjects must perform is the kind of moral and existential self-reflection that is not possible without the one taking up the perspective of the other.  

(PMT: 199; see also TCA 2:91)

The basic intuition is that there is a kind of expansion of one’s perspective required in coordinating one’s action communicatively with others, since everyone then has to be attuned to how their verbal behaviour will be understood. This role-taking ability is already required for conventional interaction. As the horizon of who can challenge the view is extended beyond one’s parochial community and beyond the present
time, the requisite interactive competence becomes one that incorporates a reflexive and critical attitude. Individuals must understand their choices as able to stand up to challenges from others. Habermas argues that the demand for postconventional personality structures (“ego-identity”) coincides with a demand for certain competencies for interaction.19

The growth of autonomy
This is a lot to expect from people, as Habermas is fully aware. And this leads to the second point: autonomy is not something we can pull off by ourselves. Our ability to be personally autonomous in Habermas’s sense depends on how we are raised, on the culture that frames our choices, and the institutional guarantees that facilitate choosing and leading an autonomous life. With regard to all these developments, however, societal pressure and demands will simply overwhelm individuals who are not able to respond effectively. As we saw in the case of moral motivation, the development of postconventional personality structures and full autonomy-competence also requires what Habermas refers to as a supportive “lifeworld that meets us halfway” [entgegenkommende Lebenswelt] (e.g. MCCA: 207). In this sense, what is required is not just any lifeworld but rather a rational lifeworld, where “rational” is understood in terms of what it is that enables agents to navigate these learning processes and maintain what they have accomplished.20 This a further reason why, as a social critic, he is very concerned about cultural trends, social transformations and (especially in his work in the 1970s) the psycho-social developments that affect childhood socialization.

In sum, we can say that there are three ways in which the autonomy of someone can be affected by societal transformations: processes of individual development (education, upbringing, material resources), demands from social institutions that burden capacities for self-steering and social processes that encourage (or discourage) critical reflection (often by providing contexts for critical reflection). As a result, “respecting the autonomy” of others becomes decidedly not a matter of getting out of the way. Here again we see Habermas’s conception of the intersubjectivity of autonomy.

Self-realization, identity and authentic selfhood

In turning now to discussions of authenticity, it is time to say something about the widespread objection that Habermas’s emphasis on the
intersubjectivity and agreement goes wrong in rendering autonomy a matter of agreement and fitting in with others, whereas it ought to be about listening to the beat of a different drummer, and swimming against the tide. Where, it might well be asked, is the unique, individual self in all this? In responding, Habermas needs to find a way to accommodate our intuitions about autonomy and authenticity being emphatically about individuality and particularity. The basic (modern) aspiration at issue here is a desire to be recognized for who one is; not for some rational, idealized version of who one ought to be, and certainly not for being what others want one to be.

It will come as no surprise that Habermas insists, even here, on an intersubjectivist account. For, although he shares the intuition about the importance of particularly, he does not share the commonly held idea that authentic individuality is somehow about uniqueness, about being different from everyone else who exists or has existed. In this section, I start out from Habermas’s rejection of uniqueness, and then reconstruct three steps in his alternative view: (i) that meaningful and authentic life choices must be criticizable, (ii) that denying the truth about one’s life history undermines the basis for recognition for who one is, and (iii) that the most fundamental mode of identity we claim from each other is that we claim to have a point of view in dialogue.21

So what exactly is wrong with thinking of authentic, individual identity in terms of being uniquely distinct from others? Indeed, according to Durkheim (in the same work quoted above), what we expect of an autonomous individual is that there is “something within him that is his and his alone, that makes him an individual” (1984: 335). There are several problems with defining oneself in terms of uniqueness. Aside from making one’s own identity hostage to the fickle choices of others (to imitate one, for example), being a bundle of properties that happened to depart from others is clearly not what we are after. After all, as Charles Taylor writes, “Defining myself means finding what is significant in my difference from others. I may be the only person with exactly 3,732 hairs on my head, or be exactly the same height as some tree on the Siberian plain, but so what?” (1992: 36). These uniquely distinguishing facts lack any personal significance for me, they are not aspects of who I am that I take ownership of in the relevant sense. Moreover, it is not really clear what it could even mean to build my sense of who I am around either of those facts. And in the absence of it being at least intelligible how that can be something that has significance for my self-understanding, it cannot play that role.
The intersubjective language of value-clarification

This brings me now to the first positive claim in Habermas’s intersubjective account of authentic identity and autonomous selfhood: that understanding oneself in one’s own individuality is not a matter of trying to ascertain the descriptive facts; rather, it is a matter of making sense of the significance of one’s actions, feelings, thoughts, desires, experiences and so on. But if there is to be the requisite difference between making sense and merely having the feeling that one is making sense, there must be a possibility for going wrong here. That is, when working out who one is, one is actually always doing two things: one is making the case for a particular understanding of what one feels, does, thinks and so on, but one is also trying to figure out whether that account makes sense. But intelligibility is not something I can just decide. Striving for authentic individuality (which is no guarantee for attaining it) requires striving to make sense of oneself in terms that are intelligible, and this necessarily links one to public language.  

So, one’s sense of “who-one-is-and-wants-to-be” is criticizable. It can fail to make sense. When such unintelligibility threatens, disputes can arise. These disputes are resolved, according to Habermas, neither through the agent’s resolute self-assertion nor by the community’s majority vote but rather in an open and dialogical process. As is the case for “discourse” generally, participants in ethical-existential discourse aim to figure something out jointly and in a manner that is maximally open to relevant considerations. Habermas describes this as follows:  

In this context the critique of self-deceptions and of symptoms of a compulsive or alienated mode of life takes its yardstick from the idea of a conscious and coherent mode of life, where the authenticity of a life-project can be understood as a higher-level validity claim analogous with the claim to truthfulness of expressive speech acts. (IO: 27)  

As he clarifies in a footnote, higher-level or “more complex” validity claims such as theories are not true or false in the way that individual sentences are true or false.  

The notion of incorrectness in the domain of authentic identity is thus unusual in two ways. First, authenticity must be assessed on the basis of how acts fit into an overall life, which explains why isolated out-of-character acts do not show someone to be inauthentic. Second, like the question of whether someone is being truthful about her expression of feelings, claims to authenticity can really only be made good on by actually living one’s life in a way that supports it. You are not the only one who can raise critical questions
about whether your actions do, in fact, support your claim to really care about something; but ultimately, you are the only one who can ensure that the evidence obtains, by living your life authentically.

The inauthenticity of selective self-examination

What is particularly important in assessing authenticity is an agent’s willingness to face the facts about himself. No matter how clear, appealing and coherent one’s sense of who one is, if it can be maintained only by suppressing key facts about what one has done or deceiving oneself about one’s deep motives, then one lacks authenticity.24 Partly, this is for the same reason that “massaging the data” undermines findings in science. But furthermore, someone who presents a selective picture of his past or refuses to take responsibility for past actions – something Habermas famously accused Martin Heidegger of doing regarding his Nazi past25 – is presenting a self for recognition to which he cannot actually lay legitimate claim. It is not simply the self-portrait itself that is flawed; it is the mode of self-portrayal. A faulty portrait can be filled in or corrected by others; but no one else can take over the task of presenting oneself, and no amount of information presented by others can rectify the failure of authenticity found in a distorting self-presentation.

In analysing this deeply performative dimension of authenticity, Habermas takes his cue from Kierkgaard’s image of voluntarily and courageously facing judgement by an omniscient God. He takes this “vertical axis of the prayer” and uses the example of Rousseau’s Confessions to “tip this into the horizontal axis of interhuman communication” (PMT: 167). Kierkgaard’s model of radically honest self-presentation thus gets transformed into Rousseau’s model of self-portrayal to the unrestricted audience of the book’s readers. (The contemporary parallel would be a highly personal blog with unrestricted access.) Once this shift is made, “the individual can no longer redeem the emphatic claim to individuality solely through the reconstructive appropriation of his life history; now the positions taken by others decide whether this reconstruction succeeds” (PMT: 167). This dramatically intersubjectivist account of authenticity is remarkably radical. Most of us are accustomed to picking rather carefully those to whom we tell all, and few of us actually want to hear everyone’s tell-all tales. Habermas is not advocating wholesale autobiographical exhibitionism but rather articulating the ideal to which a fully authentic individual can be held. It is a stance or attitude rather than behaviour that is mutually expected here. And it is not just about getting the facts right. As Habermas says about the case of Rousseau:
These confessions belong to a different genre than the descriptions that a historian could give of Rousseau’s life. They are not measured against the truth of historical statements, but against the authenticity of the presentation of self. They are exposed, as Rousseau knows, to accusations of mauvaise foi [Sartre’s term for “bad faith”] and of self-deception, not simply to that of being untrue. (PMT: 168)

In other words, there is a performative dimension to Rousseau’s confessions and, by extension, all claims to authenticity of the form, “This is who I am.”

The performativity of claims to self-authenticity
And this performative dimension – a dimension that is tied to the act of asserting a claim to being who one is – brings me to the third, and perhaps most elusive, piece of Habermas’s intersubjectivist account of personal identity and authentic selfhood, the idea of vouching for oneself and being recognized by others for so doing. On Habermas’s interpretation, what we do in making these claims is to vouch for ourselves, and this is something fundamentally different from raising “validity claims” to truth or even rightness (PMT: 190). Vouching is a matter of issuing to others a guarantee that one can make good on (or fail to make good on) by living up to one’s claim. And this is something that one must do oneself. Alongside this existential significance of vouching for oneself, there is nevertheless something intersubjective about it, for one can only undertake it vis-à-vis others. Vouching for oneself is always an appeal for recognition from others (which is not to deny that their failure to recognise one may turn out to reflect their narrow-mindedness). We can aspire, in private, to live up to certain goals, but we can vouch for ourselves only to others.

What is particularly striking about this notion of securing confirmation from others with regard to one’s vouching for oneself (for which Habermas invents the term “Selbstvergewisserung”), is that Habermas rejects the idea that this involves others agreeing with one on one’s choices or self-understanding. Rather, what matters is that one is recognized as willing and able to try: “the meaning of this guarantee has been completely grasped by the addressee as soon as he knows that the other is vouching for his ability to be himself” (PMT: 169). The assertion I make performatively to all is that I am the one who is ultimately accountable for leading my life. Paradoxically, however, that very reality is not an assertion of an independent
metaphysical claim; rather, its reality is partly constituted in the space of intersubjective recognition.

In sum, Habermas’s theory of the intersubjectivity of authentic individuality has three elements. The first element relates to the intersubjectivity of value-clarification and the publicly evaluative language employed there. The second element articulates the inauthenticity of suppressing facts about one’s past or surrounding oneself with flatterers.26 And the third element has to do with the performative character of claims to authenticity, as a matter of vouching for oneself to others. Ultimately, these points are related: if we want to understand what it means to get beyond lazy self-characterization and conventional labels and to really answer the question of who one really is, we need some way of capturing the radical depth with which this question can be asked. And engaging as co-participants in ethical-existential discourse requires recognizing one another as willing to take responsibility for how one lives one’s life.

Conclusion

The foregoing overview is intended to provide a sense of the richness of Habermas’s wide-ranging discussions of various usages of “autonomy” as well as a sense of the consistent themes that cut across these individual discussions. There are, of course, numerous points on which Habermas could be criticized or, at the very least, where the argument could be more fully developed. One task yet to be done, however, is to explain exactly what the interconnections are between these different usages of autonomy. What, for example, is the relationship between political and personal autonomy, particularly with regard to private autonomy? To what extent does moral autonomy presuppose free will, and vice versa? And what is the precise relationship between authenticity and ethical-existential autonomy? These are important interconnections to work out, but they will have to be left for another day. What is clear, in any case, is that Habermas sees the deep interconnections as coming properly into view only once one moves away from the traditional oppositions between the individual and the social – typically based on a philosophy of consciousness – and makes the turn to an intersubjectivist approach.27
Notes

1. I recall Habermas, just before the start of a weekly seminar meeting in 1988, remarking on his just-completed first official visit to the DDR: “There really is such a thing as a pathological society” [“Es gibt doch so etwas wie eine pathologische Gesellschaft”].

2. Indeed, although the title of Peter Dew’s collection of Habermas interviews, Autonomy and Solidarity: Interviews (1992a), captures two centrally motivating themes in Habermas’s work, the word “autonomy” is to be found on only two pages. Important discussions of Habermas’s views on autonomy can be found in Kevin Olson and William Rehg’s in this volume, Chapters 7 and 6, and in Allen (2008); Baynes (1992); Warren (1995); Forst (2002); Cooke (1992).

3. See especially PDM, PMT, TJ and Chapter 3 in this volume.

4. See Searle (1956) and Brandt (1994). See also Barbara Fultner’s discussion in Chapter 3 in this volume.


7. For a full discussion of Habermas’s account of political autonomy, see Chapter 7 in this volume.

8. See also Searle’s discussion of how the increasing complexities associated with globalization call for a cosmopolitan expansion of democratic institutions, in Chapter 10 of this volume.

9. For a discussion, see William Rehg’s contribution to this volume, also for a discussion of Habermas’s view that that moral insight is realized through participation in moral discourse.

10. On the lifeworld, see Chapter 4 in this volume.

11. See Chapter 8 in this volume, Habermas (1988b) and BFN. See also Heath (2008); Heath & Anderson (2010).

12. Habermas’s arguments – and references to his opponents – can be found in “Freedom and Determinism” (BNR: 151–80); and “The Language Game of Responsible Agency and the Problem of Free Will” (2007b).

13. For a full discussion, see Anderson (2007).

14. This argument, which has its roots in critical theory’s longstanding commitment to the reflexivity of science, is developed in the rather speculative final section of that essay.


16. This is a central theme in recent work in the tradition of Wilfrid Sellars (esp. Empiricism and the Philosophy of Mind). For further elaborations, see Brandt (1994); Anderson (2008); Lance & Heath White (2007); Habermas, “From Kant to Hegel: On Robert Brandt’s Pragmatic Philosophy of Language” (TJ: 131–73).

17. See also Habermas’s discussion of Durkheim in TCA 2: 43–111.


19. Thomas McCarthy nicely summarizes the contrast between conventional and critical reason-giving in terms of Habermas’s notion of discourse: “Accountable agents are able to provide publicly defensible … accounts of their actions and
beliefs, that is, to satisfy interaction partners that their beliefs and actions are backed by good reasons. Autonomous agents are able to do so also at a critical-reflective level of discourse” (1994: 44).

20. For an excellent discussion of these issues, see Offe (1992).

21. The most important text by Habermas on this topic is the rich and complex essay, “Individuation through Socialization” (PMT: 193–200), which also contains an extensive discussion of George Herbert Mead’s theory of subjectivity. Other important texts are “A Genealogical Analysis of the Cognitive Content of Morality” (IO: 3–46); “Employments of Practical Reason” (JA: 1–17); and “Historical Consciousness and Post-Traditional Identity: The Federal Republic’s Orientation to the West” (1989a: 249–68). For excellent discussions of these issues, see Cooke (1991, 1994a).

22. This idea that meaning is not something “at our disposal” is a central theme in Habermas, one that is strongly influenced by Ludwig Wittgenstein’s argument for the impossibility of a “private language”.

23. IO: 273, fn. 37. See also the final section of Rehg’s contribution in this volume.

24. There are very complex issues of the continuity of personal identity over time that could be raised here, as has been done especially by Derek Parfit (1986). Habermas does not discuss them, however.

25. See Max Pensky’s discussion in Chapter 1.

26. For a discussion of when it might not be actually appropriate to require others to account for themselves, see my discussion of the “ethics of accounting practices” in Anderson (1996: ch. 7.2).

27. I would like to thank Thomas Fossen, Barbara Fultner, Antti Kauppinen, Lillian O’Brien, Tom Bates and Markus Schlosser for feedback on earlier drafts.