There is a long tradition according to which to be a self (or person) entails a particular form of self-reflection or self-awareness as well as a level of care or concern for the self (Horkheimer 1947: 128; Adorno 1966: 218; Tugendhat 1986; Frankfurt 1988). So conceived, the self also includes what has been called “bourgeois interiority” (Gordon 2016: 4)—the vast domain of “private” thoughts, desires, hopes, and fears to which the self relates in a privileged way and over which it exercises a special authority. Further, it is also widely held that to be free or autonomous the individual must in an appropriate way “identify with” or “take responsibility for” this psychic life and the conduct issuing from it. Without this identification or appropriation, the individual can be described as alienated or not at home with himself or possibly even as a case of “disturbed” self-appropriation (Jaeggi 2014: 151). Finally, something like this conception of the self and its autonomy developed together with the rise of capitalism and plays an important role in its legitimation (Horkheimer 1947: 138; Siedentop 2014). The bourgeois individual is closely connected to the idea that each person exercises a special authority over himself—or even owns himself—and the idea that each person in virtue of this authority possesses special dignity and rights.

There is of course an equally long tradition of criticism of this understanding of the self and its powers and capacities. These criticisms range from claims that the capacity for self-knowledge is widely overestimated or that there is no such privileged authority to doubts that the domain of bourgeois interiority is a fiction or that the self (so conceived) does not exist (Metzinger 2011). Many have similarly criticized the ideal of individual autonomy and emphasized its ideological character. Subjectivity (or “interiorization”) has a deeply social and historical dimension and, as Adorno observed, “if you define interiorization as an absolute in contrast to that social dimension, and use it as the basis of an idea of pure human beings as such, you will have embarked on an irretrievable decline into ideology” (Adorno 2006: 186; see also 1966: 218). Significantly members of the Frankfurt school can frequently be found on both sides of this debate: they have relentlessly criticized the “absolutization” of the self and the more exorbitant claims made on its behalf while at the same time insisting upon the indispensable importance of a notion of the self (or “ego identity”) and the autonomous individual. As Adorno expressed it in *Negative Dialectics*: “To use the strength of the subject to break through the fallacy of constitutive subjectivity—this is what the author felt to be his task ever since he came to trust his own mental impulses” (Adorno 1966: xx). Nor is this position unique to him. All of the prominent members of the institute criticize the
excesses and ideological uses of the (bourgeois) individual and autonomy and yet continue to insist upon the value of these ideas and their importance for social criticism (Habermas 1979: 71–72; see also Honneth 2009; Jay 2016).

In view of this mixed or dual attitude among Frankfurt School theorists, it is significant that in recent years there has been resurgence of interest in the topic of individual autonomy that reflects this same ambivalence (Mackenzie and Stoljar 2000; Christman and Anderson 2005; Taylor 2005). At one level the more recent discussions operate with a more refined and nuanced terminology than that found in the earlier authors. But many of the same questions and concerns emerge there as well: Is the self or person a historical achievement or a deeper metaphysical reality? Does personal autonomy presuppose a “deep self” or is it better construed in relational or recognitional terms? Is autonomy to be understood procedurally or in connection with substantive values? What role might reference to individual autonomy play in social critique—for example, in explaining the phenomena of adaptive preference formation?

This chapter explores the views of three theorists who, despite some shared affinities, propose distinct accounts of the self and autonomy: Max Horkheimer, especially in Eclipse of Reason, traces the emergence of a procedural account of autonomy in the modern era that finally undermines the very idea of a self. Theodor Adorno, through a sustained reflection on Kant’s notion of freedom, proposes a substantive account, but one tied to a controversial metaphysics. Finally, Jürgen Habermas, in connection with his pragmatic turn, offers an account of the self and personal autonomy that is socially constituted and relational (or recognitional) in character. The implication of his recognitional account is then examined in connection with one form of ideology critique, namely an explanation of adaptive preference formation.

Horkheimer: “Self-Preservation without a Self”

Horkheimer’s writings reveal an interest in the individual self and autonomy throughout his entire career. His earlier writings trace the emergence of the “bourgeois individual” against the backdrop of changing social conditions largely within a Marxist framework—that is, in terms of the dynamic between developing human capacities and social structures (Horkheimer 1993: 118). Even the human interest in self-preservation (or what he also calls “egoism”) is not natural or fixed but is expressed in very different ways throughout history (Horkheimer 1993: 123–125). The philosophy of German idealism, which marks a high point with its emphasis on the autonomous individual, “arose above all from the contradictory circumstance that while the modern age proclaimed the spiritual and personal independence of man, the preconditions had not yet been realized for autonomy and rationally structured communal work within society” (154). Still, in his earlier essays, Horkheimer is relatively sanguine about the prospects for a rational social order in which the freedom and equality of each individual could be achieved (Horkheimer 1993: 40). However, beginning in the early 1940s and due in part to his increased collaboration with Adorno, Horkheimer advances a less optimistic picture that ties the interest in self-preservation to a specific conception of reason. As he expressed it in “The End of Reason,”

The destruction of rationalistic dogmatism through the self-critique of reason, carried out by the ever renewed nominalistic tendencies in philosophy, has now been ratified by historical reality. The substance of individuality itself, to which the idea of autonomy was bound, did not survive the process of industrialization. Reason has degenerated because it was the ideological projection of a false universality which now shows the autonomy of the subject to have been an illusion. The collapse of reason and the collapse of individuality are one and the same.

(Horkheimer 1978: 36)
Horkheimer's *Eclipse of Reason* offers his most sustained analysis of the paradoxical condition of the “supposedly autonomous individual” (Horkheimer 1947: 96). In this analysis—which roughly parallels that found in *Dialectic of Enlightenment*—the individual (as a self-conscious, purposive agent) gradually emerges over the course of natural history along with his capacity for reason. Reason itself develops in connection with attempts of human species to gain control over the forces of nature and is thus broadly in the service of the self-preservation of the species. Over a long historical period, competing interpretations of this capacity for reason and its relation to the individual vie with one another and shape our “modern” understanding. According to Horkheimer's story in the modern period, a notion of subjective (or instrumental) reason has gained an upper hand over the classical notion of objective reason (which claims to discern basic values inherent in the structure of reality). For subjective reason nothing has value or worth except that which is endowed with it by individual desire and so the idea of self-preservation is emptied of any content other than that of the desiring subject itself. Initially, according to Horkheimer, this understanding of reason and the individual, freed from the constraints of any traditional values, contributed to the growth of capitalism. (One is reminded of Marx's remark that “all that is solid melts into air.”) Over the long run, however, this understanding erodes any value or dignity that might have been attached to the self and the idea of reason in the service of self-preservation instead serves only a larger economic system. Horkheimer describes this as a crisis of reason manifested in the crisis of the individual:

> the individual once conceived of reason exclusively as an instrument of the self. Now he experiences the reverse of this self-deification. The machine has dropped the driver; it is racing blindly into space. At the moment of consummation, reason has become irrational and stultified. The theme of this time is self-preservation, while there is no self to preserve.
>
> (Horkheimer 1947: 128; Horkheimer and Adorno 1973: 48)

Despite its at times totalizing tone, Horkheimer's analysis anticipates many other criticisms of an “atomistic” individual that has lost all contact with objective values—one is reminded, for example, of Michael Sandel's much later critique of the Rawlsian “essentially unencumbered subject of possession.” The greater challenge confronting Horkheimer's position lies rather in his proposed solution. In the final chapter, Horkheimer suggests that the cure for this “disease of reason” is not to choose between objective and subjective reason, “but to foster a mutual critique and thus, if possible to prepare in the intellectual realm the reconciliation of the two in reality” (Horkheimer 1947: 174). He further suggests that the “untruth” does not lie in either concept of reason but rather in their “hypostatization” from one another.

> On the one hand, the social need of controlling nature has always conditioned the structure and form of man's thinking and thus given primacy to subjective reason. On the other hand, society could not completely repress the idea of something transcending the subjectivity of self-interest, to which the self could not help aspiring... The idea of self-preservation, the principle that is driving subjective reason to madness, is the very idea that can save objective reason from the same fate.
>
> (175)

More concretely, Horkheimer concludes that an interpretation of self-preservation that includes respect for individual life must be part of this understanding of objective reason. But it is difficult to see how such a reconciliation might be accomplished (see also Lohmann 1993).
Objective reason with its Platonic and/or religious ancestry suggests an independent order of values, whereas subjective reason suggests that something has value only to the extent that it is actually desired by a subject. At any rate, it is a project Horkheimer did not pursue further than these brief remarks.

Interestingly, there are some deep parallels to this juxtaposition of objective and subjective reason within contemporary discussions of individual autonomy. On the one hand, procedural accounts maintain that an individual is free or autonomous if she is acting from what she most desires or in accordance with her “decisive commitments” (Frankfurt 1988). On the other hand, others argue that a person cannot be autonomous if the desires that inform her motivational set conflict with certain objective values or fail to track objective reality (Wolf 1990; Taylor 2005). Although attempts have been made to buttress procedural accounts against some of its strongest criticisms—can a person autonomously will her own subordination?—it is far from clear that the attempt to “reconcile” procedural and substantive accounts offers a coherent goal.

Adorno: Reading Kant on Freedom

Adorno shares Horkheimer’s broad thesis that subjective or instrumental reason, in the service of self-preservation, is at the same time the source of human domination—this was, after all, the theme of Dialectic of Enlightenment which they jointly authored. He also shares with Horkheimer—in a less guarded moment—the claim that the logic of instrumental rationality cannot be the whole story about reason: in a claim juxtaposing objective and subjective reason that could have been drawn from Eclipse of Reason, Adorno writes,

\[ \text{Ratio should not be anything less than self-preservation, namely that of the species, upon which the survival of each individual literally depends. Through self-preservation the species indeed gains the potential for self-reflection that could finally transcends the self-preservation to which it was reduced by being restricted simply to a means.} \]

(Adorno 1988: 273)

In a critique of Weber’s decisionism and commitment to value neutrality, Adorno claims that subjective reason, on pain of incoherence, must yield a rational end:

\[ \text{The preservation of humanity is inexorably inscribed with the meaning of rationality: it has its end in a reasonable organization of society, otherwise it would bring its own movement to an authoritarian standstill. Humanity is organized rationally solely to the extent that it preserves its societalized subjects according to their unfettered potentialities.} \]

(Adorno 1988: 272–273)

However, Adorno attempts to overcome the impasse into which Horkheimer was led by developing a more nuanced account of the individual self and autonomy. He does this in large part through a close reading of the German idealist tradition, especially Kant, and by incorporating psychoanalytic insights into the Kantian account. The result is a more complex reading that does not dismiss autonomy and subjectivity, but attempts to retrieve these notions “dialectically” from a dogmatic and ideological interpretation (Gordon 2016: 190). However, whether Adorno’s own proposal is ultimately convincing is still a matter of debate (see Habermas 2008; Jütten 2010; Freyenhagen 2013).
Adorno’s reading of Kant concentrates on the latter’s resolution to the “third antinomy” in the Critique of Pure Reason. This might seem surprising since much of Kant’s discussion of autonomy arguably does not importantly depend on any such resolution. As is well known, Kant proposes to settle the dispute between free will and causal determinism by introducing a distinction between two worlds—the phenomenal (or empirical) and the noumenal (or intelligible). Insofar as individuals are part of the empirical world, our conduct is subject to the principle of causality but as rational or intelligible beings individuals belong to the noumenal world and our actions are free. Kant maintained that this distinction between two worlds secured our autonomous and moral agency from the intrusion of the empirical sciences. However, along with many other critics, Adorno thinks that this resolution is purchased at too high a price. It implies, on the one hand, that all our action is caused and at least in principle subject to causal explanation while, on the other hand, the notion of intelligible agency introduces a unique notion of causality—what has more recently been called “agent causality”—that is difficult to combine with the empirical study of human behavior. “Man’s pure noumenality, devoid of any empirical substance and sought in nothing but his own rationality, does not permit us to make any rational judgment about why it worked in one case and failed in the next” (Adorno 1966: 295, 240).

Thus far, Adorno’s reading of Kant is not especially novel. But Adorno continues to press Kant on the inconsistencies created by his two-world metaphysics. For example, despite his solution to the third antinomy, Adorno points out that Kant continued to search for a crucial thought experiment (experimenta crucis) that would prove the worldly existence of our freedom after having argued that no such proof is possible (Adorno 1966: 223, 2006: 222). In one of Kant’s more notorious examples, he asks us to consider whether, when confronted with the choice between providing false testimony or the gallows, a person doesn’t know that it is at least possible for him to overcome all self-interested motives and act from pure practical reason alone (Kant 1956: 30). For Adorno, this appeal to a unique “fact of reason”—or our consciousness of the moral law—also reveals a deeper insight: the bourgeois affirmation of our freedom is always coupled with an insistence upon our responsibility and caution to use freedom “in accordance with the law.” It thus turns the Kantian notion of freedom into a form of rational compulsion. For Kant, acting freely and acting from the moral law ultimately coincide—this is what Allison calls the “reciprocity thesis” (Allison 2006: 136)—and for Adorno this means that freedom requires acting from laws that are universal in form and so precisely contrary to anything that is unique or specific to individuals as natural and historical beings. This juxtaposition of (rational) freedom and nature leads to the (aporetic) conclusion that

subjects are free in so far as they are aware of and identical with themselves; and then again, they are unfree in such [rational] identity in so far as they are subjected to, and will perpetuate, its compulsion. They are unfree as diffuse, nonidentical nature; and yet, as that nature they are free because their overpowering impulse will also rid them of [rational] identity’s coercive character.

(Adorno 1966: 299)

One of the more intriguing aspects of Adorno’s Kant interpretation is his attempt to integrate his critique of rational psychology (and so notion of a free will) with psychoanalytic theory: “Freud was the first to register the full implications of the Kantian critique of an ontology of the soul, or ‘rational psychology’” (Adorno 1967: 81). The interpretation of moral conscience in connection with the work of a superego is one obvious example. But more important is Adorno’s critique of Kant’s sharp opposition between reason and nature. For Adorno, the human capacity for reason—rational control over natural processes—is not
The opposite of nature but the product of human nature (or psyche) in its efforts to make nature (and itself) more accommodating (Adorno 1966: 289). The ego—that is, rationality or the reality principle—thus contains a bit of nature within it. But Adorno's point is more far reaching than this claim might suggest. Ego lies, so to speak, on the border or frontier between inner nature (human psyche) and the external world and so contains elements of both: "The concept of ego is dialectical, both psychic and extrapsychic, a quantum of libido and the representative of outside reality" (Adorno 1967: 86). The ego gains its autonomy by internalizing a bit of the world and by developing its own relation to the wider psyche. This is a central insight for Adorno's understanding of psychic life: the rational self is constituted through the interaction of (an ever changing) inner nature and the outer world. There is no rational ego—or pure rationality—that does not betray signs of the life of unconscious desires and marks of an external reality. Further, in an apparent departure from Freud, the rational ego can operate unconsciously such that the repression of instinctual life can itself be irrational (or out of proportion to what is required for individuals to live successfully in the world). Adorno illustrates these claims via a critique of contemporary discussions of rationalization and the shift in psychoanalytic literature from an individual free of repression to a "well-integrated" individual. Rationalizations may be a form of self-deception (so irrational) but still disclose truths about the current conditions of social life (and so rational). Similarly, in contrast to the "revisionists," the "well-integrated individual" may not be identical with the autonomous individual but rather one whose psychic life has been tailored to meet the requirements of an irrational society (Adorno 1967: 85–86). Adjustment or integration (Anpassung) would then signal the extinction of spontaneity and autonomy (Adorno 1988: 240). At any rate, the insights of psychoanalysis betray any simple contrast between a free and rational subjectivity (or interiority) and an external (and unfree) natural and social world.

Three departures from Kant's conception of freedom are especially significant for Adorno's own account. First, individual freedom is a social achievement in two senses.

But whether the subject is autonomous in reality, whether it is able to decide one way or the other... depends on the opposite of this subjectivity that has inflated itself into an absolute in this fashion. That is to say, it depends on objective reality. For it is this, the organization of the world, the nature of the world, that actuality determines the extent to which the subject achieves autonomy, and the extent to which it is vouchsafed or denied. (Adorno 2006: 222)

Autonomy is not a wholly private and individual affair; rather it depends on a society that overcomes conditions of want and coercion and that provides appropriate background institutions. In a further sense, subjectivity or "interiority" (Innerlichkeit) is conceptually or "constitutively" social and can't be characterized apart from this social dimension (Adorno 2006: 186). Individual autonomy (as a status) requires the recognition of others (see below).

Second, autonomy is not in opposition to or at the expense of instinctual life (or "inner nature"). To be sure, as an orthodox Freudian, Adorno doubts that a free society could do away with all repression. But the "unconscious ego" promotes an irrational repression of inner nature in the service of irrational social ends. In a free society, by contrast, there would be no need for "surplus repression" and the relation between ego (reality principle) and id (instinctual desires) could develop in a more reciprocal manner, freed from the irrational demands of an (unconscious) ego and the heightened "revenge" of instinctual life (Adorno 1967: 85; O'Connor 2013: 126f.). To be sure, however, there is continued debate on precisely how far Adorno's interpretation of Freud would let him proceed along these lines (Benjamin 1977; Whitebook 2006).
Finally, Adorno claims that practical reason is impotent and cannot alone provide motivation for an agent to act (Adorno 1966: 229). This implies in turn that freedom cannot be identified exclusively with practical reason and it leads Adorno to reject Kant's basic understanding of spontaneity. For Kant rational action requires a moment of radical spontaneity—the agent “incorporates” a desire into his maxim or makes it a reason for action—in order to move from the realm of “mere” passivity to genuine (rational agency) (Allison 1996: 118). Adorno by contrast explicitly identifies spontaneity with a moment of irrationality or unintelligibility: it is what he calls the “jolt” or “surge” of an archaic impulse that lies outside the rational ego or at least exists prior to its consolidation (Adorno 1966: 221–222, 2006: 228). He refers to this notion of spontaneity as the “addendum” or “additional factor” and it marks his most radical break with Kant (Adorno 1966: 226; Freyenhagen 2015: 255f.). Like Kant, this notion of spontaneity is necessary for free action but, unlike Kant, it is associated not with the noumenal realm but with the realm of archaic impulses and nature.

Despite this sharp departure from aspects of Kant's account, Adorno's treatment of Kant is not entirely critical. Indeed, in some of his public lectures he seems to embrace Kant's idea of autonomy without reservation: “The single genuine power standing against the principle of Auschwitz is autonomy, if I might use the Kantian expression: the power of reflection, of self-determination, of not cooperating” (Adorno 1983, 1988: 195). These comments are not limited to his public lectures, but can be found in the same texts in which he criticizes Kant's philosophy (Adorno 2006: 178). Thus, the central concern for Adorno seems to be not a complete dismissal of the Kantian view of autonomy but rather that the challenges confronting its realization will be too quickly dismissed. This is not the entire story as it ignores his problems with Kant's dualisms but it is still a point worth emphasizing since it distances Adorno's position from more radical calls for the “end of the subject.”

What is the alternative understanding of freedom (autonomy) and agency that Adorno wants to oppose to Kant's solution to the third antinomy? Interestingly, Adorno does not pursue a strategy that has appealed to many neo-Kantians. This involves a rejection of the “two world” interpretation of Kant in favor of the idea of “two standpoints” on one world (which however can only be accessed from one or another standpoint) (Allais 2015: 8). A strong textual basis for this reading is found in Kant's remark that insofar as we regard ourselves as agents we must think of ourselves as free (Kant 1996: 114; Allison 2012). Adorno dismisses this passage with the brief remark that it turns freedom into a mere “fiction.” He associates it with Vaihinger's interpretation of Kant as a “philosophy of the ‘as if’” and suggests that this reduces the problem of freedom to a mere pragmatic interest (Adorno 2006: 244–245). Further, Adorno repeats his claims that this account of freedom (which again connects the notion of freedom to the idea of acting under universal law) renders it a form of coercion: Kant's claim that we have no choice but to act under the idea of freedom (as universal law) reveals the “clenched hoof” that finally subordinates Kant's notion of human freedom to the idea of a universal causality (Adorno 2006: 246).

In contrast to Kant's two-world metaphysics, Adorno appeals to a competing Romantic conception of nature (see Jütten 2010). Rather than the idea of natura naturata (constituted nature, as a closed causal realm), he invokes the Romantic notion of natura naturans (constituting nature) in which human spontaneity and (human) nature are not in deep conflict but work in tandem (Adorno 2006: 245). The problem of freedom and causality (Kant's third antinomy) is dismissed as a “pseudo-problem” that in fact arises from the real experience of freedom and coercion in a social world that has taken on a “quasi-natural” form. This “second nature” however is the result of an instrumental reason (or “identity thinking”) that has obscured a deeper remembrance of nature in the self and turns the world (including other agents and ourselves) into objects. Only through a reversal of this “derailed natural history” can a proper understanding of human freedom (and its harmony with nature) be achieved.
Habermas: Kantian Pragmatism, Recognitional Autonomy, and Adaptive Preferences

Habermas’s writings reflect the same ambivalence regarding the self and individual autonomy that can be found in Horkheimer and Adorno. He is critical of the ideological role these ideas have played in capitalist society but he also recognizes that a notion of the autonomous individual is indispensable for the normative foundations of critical theory (Habermas 1979: 72–73). Deeply influenced by the thought of Durkheim (as well as Hegel), Habermas acknowledges that the self is a “social construction”—that is, the product of a long and contingent historical process—but nonetheless insists that it is more than a “mere fiction”—it is an idea central to (modern) social practices that cannot simply be set aside (Habermas 1992: ch. 7). Like Adorno, he resists embracing a “false positivity” or premature acceptance of “bourgeois interiority”—a pure subjectivity untainted by society and granted excessive powers and capacities—but insists that an intersubjective or recognition account of autonomy must still be preserved.

Both Adorno and Habermas attempt to move beyond Kant’s transcendental idealism with its reliance upon a “constitutive subjectivity” but they do so in very different ways. Adorno’s negative dialectics exposes the limits of Kantian dualism—and so at least implicitly points to a unity of nature and freedom that is prior to the rise of instrumental reason (Freyenhagen 2015: 259). Habermas, by contrast, offers a pragmatist reading of Kant that aligns more closely with the epistemic or “standpoint” interpretations of Kant. With respect to the concept of freedom (or autonomy), Habermas also recasts the third antinomy in terms of two standpoints or perspectives. As agents, we experience nature as a limiting or constraining condition on our actions (Habermas 2007: 19, 2008: 189) but this encounter with nature (accessible from the participant’s perspective) is not the same as the “theoretical” or “observer’s perspective” that assumes a domain of closed causal laws. This latter conception of nature cannot simply replace the former nor can the former be reduced to it. Thus, the “resolution” of the third antinomy does not involve an ontological claim about two worlds but an epistemic claim about the mutual irreducibility of two perspectives.

In what he labels “perspectival dualism” (Habermas 2008: 165) or “linguistic dualism” (208), Habermas locates these two perspectives in the “grammatical structure” of language or, more specifically, in the practice of linguistic communication. In a further step, Habermas associates the participant’s perspective with the mundane or everyday experience of social actors (or what Sellar’s called the “manifest image” of the world). On Habermas’s view, although scientific insights can modify aspects of this manifest image they cannot ultimately threaten or undermine the participant’s perspective—at least not without radically altering the deep-structural “grammar” of our linguistic practices. Finally, Habermas attempts to show that the observer’s perspective presupposes the participant’s perspective since science cannot eliminate the need for argumentative practices rooted in everyday language use (Habermas 2008: ch. 1; Baynes 2016: 36–37). So formulated, this amounts to a version of Kant’s claim that we can act—that is, understand ourselves as rational and accountable agents—only under the idea of freedom combined with Kant’s thesis concerning the primacy of practical over theoretical reason (Gardner 2006; Allison 2012: 98). For
Habermas, accountable agency is an idealization built into our communicative practices that can’t be abandoned while still regarding ourselves as capable of engaging in reasoned arguments (Habermas 2007: 23f).

At the core of this understanding of the participant’s perspective is a basic notion of autonomy or communicative freedom. Communicative freedom refers to the capacity of agents to accept or reject the various claims raised in basic types of speech acts (or illocutionary acts). This notion of freedom is prior to other notions of autonomy and is contained in the idea that, as language users, individuals occupy a normative standing or status. Competent speakers can make assertions, promises, requests, and the like because they attribute to one another (and themselves) the normative power to challenge or accept the various sorts of claims raised in those speech acts. Communicative freedom refers to this normative power and successful or “felicitous” communication depends on its mutual ascription. This notion of communicative freedom is neither a wholly “subjective” capacity nor an indication of our membership in a noumenal realm. It is rather intersubjective (or recognitional) from the outset and it presupposes a network of social practices and institutions that Habermas (with a nod to Husserl) refers to as the “lifeworld.” In their interpretive activity, actors draw upon the norms and practices of this lifeworld and appeal to them in their exchange of reasons. In a rather vivid metaphor, Habermas describes the complementary relation between the communicative action of accountable agents and the sociocultural lifeworld as a “circular process”:

While the segment of the lifeworld relevant to the situation encounters the actor as a problem which he has to solve as something standing as it were in front of him, he is supported in the rear by the background of his lifeworld. Coping with situations is a circular process in which the actor is two things at the same time: the initiator of actions that can be attributed to him and the product of traditions in which he stands as well as of group solidarities to which he belongs and processes of socialization and learning to which he is subjected.

(Habermas 1987: 135)

However, he also emphasizes that this description of a “circular process” should be accepted with caution: actors are not products of the lifeworld in a way that renders them passive nor should the lifeworld be regarded as a self-perpetuating process—a macro-subject—that has a life of its own. Rather, it is individuals (and groups) who reproduce the lifeworld through their communicative action and the lifeworld is “saddled upon” the interpretive accomplishments of its members (Habermas 1979: 121, 1987: 145). Or, to express the same point in somewhat different terminology, reference to the situatedness of the agent (and her reasons)—a “product” of the lifeworld—must not be at the expense of her status as a rational and accountable agent, but should rather be seen as a central condition of that agency. Habermas’s model of social agency seeks to avoid the extremes of treating agents as “cultural dopes” who passively reproduce their lifeworld, on the one hand, and rational agents who always act (insofar as they are rational) to maximize their own self-interest, on the other hand. Neither of these extremes captures adequately the idea that actors are deeply embedded in their lifeworld and accountable agents who reproduce it through their communicative action (Habermas 1992: 43).

Habermas recognizes that there is a diversity of forms of autonomy, including moral, political, legal, personal (or individual) (Anderson 2011: 92). Importantly, however, these other conceptions are derived from or built upon the basic notion of agential autonomy; nor is there a more fundamental notion of individual self-relation or freedom on which the notion of agential autonomy depends, in a way that might suggest a person must first be a self or
The Self and Individual Autonomy

possess individual authority before he or she can possess the normative power associated with speech acts (see, for example, Habermas’s important exchange with Dieter Henrich in Dews 1999). In Between Facts and Norms, Habermas describes the “co-originality” of public and private autonomy (which correspond roughly to the notions of political freedom and civil liberty), but in his account, these two notions of autonomy themselves presuppose the notion of agential autonomy outlined above (Habermas 1996: 4). Similarly, the idea of moral autonomy—or the capacity to act from moral norms—presupposes both the more minimal notion of agential autonomy and the idea of membership in a moral community (compare Darwall). Finally, even individual (or personal) autonomy—or what Habermas calls “ethical self-understanding”—presupposes a relevant public to whom one can in principle give an account or justification for one’s decisions and commitments. Habermas describes the rise of this form of individual autonomy in connection with nineteenth-century confessional literature (such as Rousseau’s Letters) in which claims to an authentic life are still dependent on the recognition of others:

No one can dispose over his identity as property. The guarantee under discussion must not be conceived according to the model of a promise through which an autonomous speaker binds his will… The self of ethical self-understanding is dependent upon recognition by addressees because it generates itself as a response to the demands of another in the first place. Because others attribute accountability to me, I gradually make myself into the one who I have become in living together with others.

(Habermas 1992, 170)

In sum, these further notions of autonomy are also deeply recognitional in character and dependent upon the basic idea of communicative freedom and agential autonomy.

One of the primary reasons for the Frankfurt School’s interest in autonomy was to better understand its ideological function and its role in the critique of that ideology: viewed negatively, appeals to individual autonomy appear to be “false” and in the service of wider social imperatives; on the other hand, it also seems that some positive notion of individual freedom or autonomy must be presupposed by those who engage in social criticism (Horkheimer and Adorno 1973: 198f.; Habermas 1979). An analogous tension can be found in the recent literature on individual autonomy and its use for social critique (see Mackenzie and Stoljar 2000; Veltman and Piper 2014; Oshana 2015). For example, some feminists have appealed to a notion of autonomy to help explain how the oppressed might be complicit in their own oppression. Under the label of “adaptive preferences” the claim is that some individuals acquire preferences that undermine their own autonomy and work against their own interests (Stoljar 2014). Others respond that such appeals to an “autonomy deficit” display a form of disrespect toward the oppressed (Khader 2011). Habermas’s recognitional account of autonomy might offer a way to navigate some of this challenging terrain.

Jon Elster initiated the recent discussion concerning adaptive preferences with the suggestion that such preferences circumvent or curtail an agent’s autonomy. According to Elster, an adaptive preference is a preference that has been unconsciously formed or altered as a result of an (unchosen) reduction in the set of feasible options available to the agent (Elster 1983: 117). It is something that happens to the agent, rather than something the agent does. The fox is frustrated by his inability to reach the grapes and his preference is modified in order to reduce his level of frustration. However, it is not modified on the basis of a choice or through an exercise of his deliberative capacities. Others have suggested that Elster’s account is too thin and that a more substantive account of individual autonomy is required in order to identify the objectionable feature of adaptive preferences. Natalie Stoljar, for example, has
argued that even if a woman consciously and reflectively embraces certain oppressive social norms, there may still be a failure or deficit in her autonomy. A person living in a society with oppressive social norms might have so deeply internalized these norms that she would continue to endorse them even if she were critically to reflect upon them in accordance with a procedural account of autonomy. On her view, we need a more substantive account of autonomy—one that incorporates objective values—in order to address the worry about adaptive preferences. Finally, in a further response, some proceduralists have claimed that substantive accounts like Stoljar's mischaracterize the conditions for autonomy and/or set the bar too high (Christman 2009).

In the context of this recent discussion, Habermas's account of individual autonomy is best viewed as an example of a normative competence approach. What characterizes such an approach is not whether a particular preference conforms to a set of substantive values or even whether it would be endorsed in critical reflection, but whether or not it is the outcome of an exercise of a normative competence (Benson 2000). In Habermas's case, the basic idea is that for a preference to be (properly) attributable to an agent it must arise from an exercise of her communicative competence—where this competence refers at the most basic level to the agent's capacity to respond with reasons to claims raised in speech acts. More generally, it refers to an agent's capacity to offer justifications for her conduct and preferences to a relevant audience. Such a normative competence supposes that the agent has a basic (normative) status and authority to give reasons and to respond to reasons offered by another (normatively competent) agent. What the possession (and appropriate ascription) of this competence requires in a given situation is a complex question and depends on a wide variety of factors. It will include as well an answer to questions concerning the relevant audience to whom the person is accountable and the appropriate domains or aspects of her life for which she owes an account. The notion of communicative competence should also not be construed in an overly "intellectualist" manner, as it will include many affective and emotional skills and dispositions as well (see Anderson and Honneth 2005). To anticipate and as a first approximation, preferences held by the agent that did not arise through the exercise of her communicative competence (with an important qualification below) are not properly attributable to her as an agent—they are not genuinely her own—and so are adaptive in Elster's initial sense.

Quite independent of Habermas's work, Andrea Westlund has proposed a strikingly similar account that she describes as a competence (or disposition) for answerability. Individual (or personal) autonomy requires (an exercise of) a disposition or readiness to give reasons (be answerable) to others. A person is autonomous with respect to a given preference then if the preference is formed through an exercise of her normative competence for answerability rather than in a manner that thwarts or circumvents that competence. Three features of her normative competence approach are especially worth noting for the similarities with Habermas: first, the competence for answerability is still conceived quite formally or procedurally in the sense that there will be a wide range of reasons that could be offered that are compatible with the exercise of this competence and preferences that cannot be judged to be non-autonomous directly on the basis of their content. Nonetheless, as others have pointed out, it remains at least a "weakly substantive" account in the sense that it relies upon some substantive normative assumptions and (more controversially) not all preferences could be compatible with an exercise of this competence (Benson 2011: 4).

Second, as Paul Benson and others have emphasized, a competence for answerability requires an adequate sense of self-worth (Benson 2000; Anderson and Honneth 2009). If a person does not view herself as worthy or entitled to have a voice or to have an opinion—to be, in other words, an authoritative source of valid claims—she would fail to have the relevant normative competence for answerability. Similarly, to see oneself as answerable to
others also supposes that one has a sufficient or adequate sense of one’s own self-worth. Autonomy thus requires not only possessing the authority to give and respond to reasons but an adequate appreciation or recognition of this authority with respect to oneself. One must, in other words, have some recognition of oneself as a “self-authenticating source of valid claims” (Rawls); one must have, and to some extent recognize oneself as having, a voice that can command an (equal) authority just as much as it is (equally) answerable to another. Finally, as Westlund points out, individual autonomy does not only depend causally on others for its development, but is also “constitutively” social: it is a normative power that by definition requires the recognition or “uptake” of another (Westlund 2009: 44). A person can only be answerable or have the authority to make a demand if that capacity and authority is recognized by another who also has that status and capacity. Her relational account thus again closely parallels Habermas’s.

How does this normative competence account bear on the discussion of adaptive preferences? Roughly, a preference is adaptive if it does not arise from an exercise of a person’s communicative competence or normative competence for answerability. Obviously, this initial formulation needs further clarification and some important qualifications. As Westlund also notes, not every (autonomous) preference actually held by an agent arises from a deliberate or conscious exercise of this competence. A great deal of “deliberation” about one’s preferences can occur at a fairly unreflective and perhaps even unconscious level. Second, and more importantly, since realistically (and quite generally) not all preferences are the direct result of any exercise of such a normative competence, this proposal must be amended to include a modified version of Christman’s authenticity condition (Christman 2009: 155): if the agent were to reflect on the preference in question, she would endorse it—that is, she would regard it as a preference that could have arisen through an exercise of her normative competence (in a world sufficiently similar to her own). However, as Westlund again points out, there is a real danger here: it makes little sense to speak of a person possessing a normative competence—and even less to speak of her as autonomous—if she never exercises the competence (Westlund 2009: 35). It also would be problematic to describe preferences ascribed to a person as “autonomous” if the ascription relied solely on a counterfactual claim about what a person would endorse absent any significant exercise of the competence. So, while for any given preference it may be plausible to consider it autonomous if it satisfies this counterfactual condition, it would not be plausible to consider an agent autonomous if she did not actually exercise her competence with respect to some range of her preferences that she valued and cared about. That is, there must be a real limit with respect to when this counterfactual condition is invoked in the absence of the actual exercise of the person’s normative competence. Third, and finally, the exercise of a competence can be quite local and domain specific: the fact that someone does not regard himself as entitled to a voice in one context, or is not so regarded by others, does not necessarily mean that he lacks a normative competence for answerability (Mackenzie 2008: 523). This of course implies that empirical ascriptions (even self-ascriptions) of a competence are extremely difficult and should be done with great caution and “epistemic humility.” It does not mean, however, that a social theorist is never warranted in proposing such hypotheses to explain some social phenomenon or that she is always displaying disrespect toward an individual in making such a judgment.

At this point Habermas’s account of communicative competence offers some additional resources. He locates both the interpretation and value of autonomy in history—and specifically within processes of modernization—and he does not oppose autonomy and socialization (in, say, the fiction of a “deep self”) but treats certain forms of socialization as necessary conditions for individual autonomy. Both of these aspects of his account are captured in his contrast between “normatively ascribed agreements” and “communicatively achieved
understanding” (Habermas 1985: 340). The more specific details of this view can’t be pursued here, but what matters for this conception is that with modernization comes a broadening scope or range in which it is expected that the individual will assume responsibility for himself and his commitments—in religious faith and values, in love and intimate relations, in work and career choices, and in solidarities and allegiances (Habermas 1990: 199).

Of course, this historical dimension of Habermas’s account brings complications with it as well. On the one hand, it both allows for and requires much greater historical perspective and cross-cultural variation when speaking of autonomy and its value. On the other hand, it calls for much more nuanced ascriptions of autonomy as is evident in his own distinction between a thin and thick notion of communicative freedom (Baynes 2016: 117). It allows, for example, for both a more minimal notion of agency (that presumably must be presupposed even in “normatively ascribed agreements” if we are to be speaking of agency at all) and the much more demanding ideal of personal autonomy if we are talking about communicatively achieved understanding across a broad range of an individual’s practical identity and in more than highly episodic and circumscribed manner.

How does the normative competence account fare with respect to the worry about deeply entrenched preferences and tightly scripted social norms mentioned above? Isn’t it possible that a person might be “in the grip of a norm,” even in exercising her normative competence for answerability, such that a preference she has might still be “adaptive” (where that means, not properly attributable to her, or not truly her own)? I confess that I find such a scenario difficult to imagine. It would mean, for example, that the deferential wife would not be able to regard her preferences as ones she could have acquired through an exercise of her normative competence—where this means she could not see those preferences as ones she would be able to give good reasons for in an exchange with her relevant peers or interlocutors. But others may have different intuitions at this point. Of course, much will depend once again on how thinly or thickly we understand the competence for answerability or the ability to exchange reasons—and this, I have suggested, can and should vary with context.

To possess the normative competence for answerability does not mean being able to give a (good or reasonable) answer to just anyone who happens to ask for one (Westlund 2009). Habermas’s view also allows for the fact that ascriptions of responsibility should be considered in connection with a wider assessment of social policy and goals—greater expectations for accountability must be matched with resources for its realistic exercise (Anderson 2008). These observations underscore, once again, that actual judgments about a person’s failure with respect to the exercise of a normative competence—and so about whether his preference is adaptive—are judgments that must be made with extreme caution. At the same time, it suggests that a recognition of autonomy can perform a valuable role in the type of social criticism that is a distinctive feature of the Frankfurt School tradition.

References
THE SELF AND INDIVIDUAL AUTONOMY


**Further Reading**

Habermas, J. (2008) *Between Naturalism and Religion*. Malden, MA: Polity Press. (Contains Habermas’s recent essays on autonomy (or freedom) including critique of Adorno.)