The Alienation of Social Institutions:
A Social Ontological Account

Chris Ferguson

PhD Social and Political Thought

University of Sussex

September 2019

Declaration

I hereby declare that this thesis has not been and will not be, submitted in whole or in part to another University for the award of any other degree.

Signature: ....

University of Sussex

Chris Ferguson, PhD Social and Political Thought

The Alienation of Social Institutions: A Social Ontological Account

Summary

This thesis examines Marx's use of the concept of the alienation of social institutions. It examines Marx's use of the terms *Entäußerung* and *Entfremdung* in the context of his antecedents in Fichte, Hegel, Feuerbach, and other Young Hegelians such as Stirner and Bauer. I argue that the concept makes concrete claims about the becoming-independent of social institutions, and explore the basis on which Marx takes this to have normative import. With respect to the latter, I claim that this concept draws on a conception of freedom, and is thus importantly distinct from a more frequently-emphasised normative theme in Marx's work, that of perfectionism.

The concept of the alienation of social institutions is nevertheless somewhat opaque in Marx's writings, a product of the macro level of description at which Marx uses it. In order to adequately understand the descriptive claims this concept makes, we need a clearer idea of what social institutions are and what their relationship is to individuals. To this end, I engage in a critical reconstruction of John Searle's social ontology. Searle's work offers a useful, although occasionally unfortunately ambiguous, framework for describing social institutions and their alienation. I reconstruct the central claims of his account, clarifying several important aspects in order both to aid our understanding of social institutions and avoid some of the misleading (and conservative) implications that seem to follow from his account.

Finally, in order to understand the normative import of Marx's concept of alienation, I engage with the contemporary literature on freedom in order to reconstruct Marx's conception of freedom. I show how this conception of freedom underpins the normative import of his concept of the alienation of social institutions.

Contents

Acknowledgements

I am unlikely to have ever started, and most certainly could not have completed, this thesis without the kind support of my supervisor, Andrew Chitty. Andrew's help, from his patient and generous comments to his infectious enthusiasm and good cheer, has kept both this thesis and its author afloat at every stage of the process. I cannot thank him enough.

I would also like to thank the other faculty members of the Centre for Social and Political Thought, and those in the Department of Philosophy more broadly. In particular, thanks go to Gordon Finlayson and Darrow Schecter who, along with Andrew, inspired me during my MA to consider undertaking this project, and to Tony Booth, my second supervisor, for encouragement and advice.

This thesis took longer than expected to write, the benefit of which is that I have met so many brilliant and truly lovely fellow researchers. I would like to thank them all for their stimulating conversations and general comradery. While I cannot thank them all here, especial thanks go to Ane Engelstad, Robb Dunphy, Val Whittington, Richard Weir, Ezra Cohen, André Almeida, David Martinez Rojas, Umut Sahverdi, Johnbosco Nwogbo, Denis Chevrier-Bosseau, James Stockman, Anna Wimbledon, Neal Harris, and Jonny Lee.

Beyond Sussex, Sam, Bettie, Mark, Kai, Jimmy, Gian, Richard, Robyn, Adam and Kath have all been indispensable allies throughout. I love you all. Particular thanks go to Sam for proof-reading and giving crucially insightful and supportive comments on my final draft. Thanks also go to the members of Brighton City Grid for their companionship and for providing much-needed distraction.

This thesis is dedicated, first, to my family, in particular my parents, to whom I owe more gratitude than can be expressed here, and my brothers. Writing it has prevented me from visiting them all as often as I wish I could, especially in difficult times, and I hope they know how often they are in my thoughts and heart. I would also like to thank my second family, the Kirwans, whose generosity is rivalled only by their musical prowess.

Finally, it is also dedicated to my literal other half, Ebba, without whom none of this would have been possible (an eventuality she is entitled to prefer). Thank you for everything, over the last five years and the nine before that, and in particular for plotting our escape route!

Introduction

This thesis aims to explicate and reconstruct a concept which runs through much of Marx's work, most explicitly that up to and including the *German Ideology*, which I call the concept of *the alienation of social institutions*. I aim to present this concept as making concrete claims about the social world, and to clarify its normative significance, such that we might properly evaluate its contemporary relevance (the latter being a task I will not myself undertake).

This concept is connected to and underpins many of Marx's statements concerning *Entäußerung* and *Entfremdung*, both commonly translated into English as "alienation". Marx primarily uses these terms as predicates not of social institutions—which he calls "social formations" or "social relations"—, but of other features of human life: the human essence, labour, activity. Yet, I will argue, these formulations are frequently connected to the idea that social institutions are "alienated", and can be fully understood only given this connection. For Marx, social institutions are expressions or manifestations—externalisations—of human activity, and when they are alienated social institutions become in some sense independent of the agents whose activity constitutes them. This results in these social institutions restricting agents' freedom.

The term "alienation" is ambiguous between an end state and the process through which something arrives at that state: a social institution can be described as an alienated one, or we can talk of its alienation, the process through which it becomes alienated. I will be interested in both aspects and will use the term in both senses, with the context making clear which is meant. My initial focus will be on describing the process of alienation; at the end of chapter 4, I will suggest that a social institution that is alienated is one that has a particular *mode of existence*.

In chapters 1 and 2, I discuss Marx's uses of *Entäußerung* and *Entfremdung* in the context of his intellectual predecessors. I outline how Marx's use of these terms inherit a central idea from these writers, and how he uses these terms in importantly new ways. Chapter 1's focus is on the idea that certain entities, primarily religious phenomena and social institutions, are the product of agents' externalisation of some aspect of themselves. Chapter 2 focuses on the idea that these entities become independent, and examines the normative component of such claims.

Chapters 3 and 4 aim to provide the resources for reconstructing the concept of the alienation of social institutions so as to render it more easily understandable. In particular, this concept contains social-level terms that render its elements opaque: I suggest we need to translate these to the level of individual agents and their relations. To do so, I critically reconstruct John Searle's account of social ontology, which I then use to translate our central concept.

Marx's account associates the alienation of social institutions with unfreedom, but nowhere gives an explicit account of what he means by this term. In Chapter 5, I develop a conception of freedom that is able to give normative significance to that of the alienation of social institutions. I argue for a forward-looking conception of negative freedom according to which an individual is unfree if they are rendered unable to do what they want to do by an obstacle that is alterable by human agency, and discuss how this conception relates to social institutions' alienation.

As I have said, chapters 3 and 4 heavily feature the work of John Searle. This might seem like a strange resource to draw upon in understanding a Marxist concept, particularly given Searle's clear antipathy towards the Marxist tradition[1] and his account's conservative ideological implications (see chapter 4). Nevertheless, Searle's account can provide a helpful way into the questions we must address if we are to understand social institutions and their alienation. His incisive and clear (although sometimes ambiguous) discussion, I hope to show, can provide a useful platform from which to grapple with these questions.

However, this is true of many other authors working in this area. And, on reflection, almost any alternative account would have been preferable to use. This is because since writing these chapters it has become known that Searle has engaged in appalling practices of sexual harassment towards (at least) one former student and employee, behaviour for which his emeritus status at the University of California has been revoked.[2] In light of this behaviour, I have no desire to contribute to the reputation of Searle or his work, to the continued circulation of his name, nor the rehabilitation of his career. If I had more time, I would be tempted to rewrite these chapters so as to exorcise the references to him, in solidarity with those affected by his behaviour, as well as with those countless people affected by the institutional

---

[1] See Faigenbaum and Searle, *Conversations with John Searle*, 161.
[2] Weinberg, 'Searle Found to Have Violated Sexual Harassment Policies'.

sexism that enabled this behaviour to continue for so long.[3] As I do not have that time, I can only hope that my own contribution to the topics with which Searle has concerned himself can form a small part of an alternative literature to which others interested in these questions can refer.

---

[3] It seems Searle's behaviour was widely known for many years, his career being unaffected: Baker, 'UC Berkeley Was Warned About Its Star Professor Years Before Sexual Harassment Lawsuit'.

Part 1
Karl Marx and the alienation of social institutions

Chapter 1
Entäußerung and Entfremdung

The concept with which this thesis is concerned, which I call *the alienation of social institutions*, is one which, I argue, lies at the heart of Marx's early work, although rarely in explicit form. In particular, it underlies Marx's frequent discussions of the alienation—*Entäußerung* and *Entfremdung*—of, variously, human beings, their essence, or their labour. While my terminology extrapolates from Marx's own usage, I hope to show that Marx's discussions of alienation can be best understood only alongside the idea that social institutions themselves are alienated.

Marx's terms *Entäußerung* and *Entfremdung* are closely related, and I translate both simply as "alienation". Both come with significant baggage, some relevant to the concept of the alienation of social institutions and some not. I will not attempt to unpack all this baggage—to give a systematic account of the multiple meanings these terms had for Marx and the lineage of each—nor to claim that any particular bag is more valuable than the rest. Rather, I will provide a brief history of these terms' use insofar as this pertains to the concept that is my focus, picking through these bags selecting only what we need for our trip.

I begin with a brief history of these terms' use by Marx's predecessors, before examining—in this and the subsequent chapter—the extent to which Marx's use both follows these and attaches new meanings to them. This will require an examination of the connection between these terms and an account of the human "essence", and the role this plays in Marx's ability to criticise the social relations of the contemporary world.

1 Entäußerung and Entfremdung before Marx

1.1 Fichte

The first philosophical use of the word *Entäußerung* can be found in Fichte's *Attempt at a Critique of All Revelation*. This work analyses the relation between religion and human beings *qua* Kantian moral agents, or agents that are autonomous insofar as they act according to the law of practical reason (the "moral law") of which they themselves are author. The problem that Fichte saw as arising from this relation was the following. If "respect for the [self-legislated] moral law … must alone be the incentive for every purely moral action",[1] then this appears to be incompatible with the religious idea of acting according to God's will:

> reason obligates us to obey its law without referring back to a lawgiver above itself, so that it confuses itself and is simply destroyed and ceases to be reason if one assumes that there is still something other than itself which commands it.[2]

Fichte, who did not want to claim that religion was inherently immoral, resolved this tension by arguing that we respect God's will *because* it is "completely identical with that given to us by our own reason, according to which we are supposed to act".[3] As such:

> The determination of the will to obey the law of God in general can take place only through the law of practical reason and is to be presupposed as a lasting and permanent decision of the mind.[4]

> respect for God is based solely on his acknowledged agreement with this [the moral] law and hence on respect for the [moral] law itself.[5]

Fichte's argument, reminiscent of Plato's *Euthyphro* dilemma, is that autonomous Kantian morality and religion can be compatible only if one's respect for God derives from one's morality, rather than *vice versa*. This argument leaves Fichte with an explanatory problem: if we act according to God's will only because it is identical to a moral law that we respect independently and which we created, why bother with God and religion at all? As Fichte puts it:

---

[1] Fichte, *Attempt at a Critique of All Revelation*, 42.
[2] Fichte, 39.
[3] Fichte, 42.
[4] Fichte, 40.
[5] Fichte, 41.

it is clear in the first place that it does not matter at all even for the morality of our actions whether we consider ourselves obligated to something because our reason commands it or because God commands it. But from this, one cannot yet see at all what purpose the latter representation is supposed to serve, since its efficacy already presupposes the efficacy of the former, since the mind must already be determined to want to obey reason before the will to obey God is possible – since it appears, therefore, that the latter representation could determine us neither more universally nor more strongly than the one on which it depends and through which it first becomes possible.[6]

Fichte's solution is that our belief that the moral law is given to us by God is useful because it helps solve another problem associated with Kantian morality, namely that the moral law is not the only law that emanates from the agent. While the fact that the moral law is self-legislated may be sufficient to motivate us to act according to it in situations in which it is the *only* law we have given ourselves, other similarly self-originating laws—for example desires, which Fichte terms "natural laws"— exist, and these often come into conflict with the moral law. Given that natural laws have the same origin in the agent as does the moral law,

if the dignity of the law is determined solely according to that of the legislative subject, this natural law could appear to be of the same rank and worth as the law of reason.[7]

Thus "particular cases of the application of the law can be conceived in which reason alone would not have sufficient power to determine the will".[8] This is why, Fichte believes, a belief in God as lawgiver is necessary:

But if the duty that occurs in this case appears to us as commanded by God, or if the law of reason appears thoroughly and in all its applications as God's law, which amounts to the same thing, then it appears in a being regarding whom it is not a matter of our discretion whether we will respect him or refuse him proper respect. By every conscious disobedience to this being, not simply do we make an exception to the rule, but we directly deny reason in general. We are sinning not merely against a rule derived from reason but against its first commandment.[9]

It is in this context that Fichte introduces the term *Entäußerung*:

The idea of God, as lawgiver through the moral law in us, is thus based on an alienation [*Entäußerung*] of what is ours, on translating something subjective into a being outside us; and this alienation

---

[6] Fichte, 39.
[7] Fichte, 40.
[8] Fichte, 40.
[9] Fichte, 41.

> [*Entäußerung*] is the real principle of religion, insofar as it is to be used for determining the will.[10]

From the outset, then, *Entäußerung* described a process in which agents treat something which is internal to them or exists because of its relationship to them— the subjective existence of the moral law—as something which exists "outside us" or independent of this relationship. In addition, it suggested that in so treating something, it acquires an increased ability to affect our will or behaviour, i.e. to motivate us to act in accordance with the moral law despite our desires.

This complex meaning is expressed in the etymology of *Entäußerung*. *Äußerung* is primarily used to mean "remark", "statement" or linguistic "expression",[11] however it can be used more broadly to mean "expression" generally, "manifestation", or "sign".[12] *Ent-* is an inseparable prefix that modifies the attached verb, indicating the *beginning* of something, the *removal* of something, or its *reversal*.[13] Together, then, the word means the expression or manifestation of something in such a way that whatever is expressed changes, entering a new state.[14] In Fichte's account, agents express or manifest the moral law that they have given themselves in such a way that it takes on different characteristics or becomes something else, the commandment of God.

1.2 Hegel

Hegel's use of *Entäußerung* is somewhat idiosyncratic, and examining his use of it in detail would complicate matters unnecessarily. Importantly, it is Hegel that first associates the term *Entäußerung* with that of *Entfremdung*. *Entfremdung* is most often translated into English as "alienation" or "estrangement", "*ent-*" having the same meaning as above, indicating change to a new state; "*fremd*" being most commonly translated as "alien" *qua* being "cut off" or "separate".

We can highlight two processes, each mirroring that described by Fichte, which Hegel describes using these terms. The first is equivalent to that described by Fichte but in a different context (and with different consequences). Here, Hegel describes a

---

[10] Fichte, 41.
[11] 'English Translation of "Äußerung" | Collins German-English Dictionary'.
[12] 'Äußerung - English translation in English - Langenscheidt dictionary German-English'.
[13] 'Inseparable Prefixes (Feste Vorsilben)'.
[14] See also Arthur, *Dialectics of Labour*, 147.

process whereby agents see themselves not as mere individuals but as part of the universal, as part of "culture". Thus:

> [individual] self-consciousness is merely a 'something', it has *actuality* only in so far as it alienates [*entfremdet*] itself from itself; by so doing, it gives itself the character of a universal, and this its universality is its authentication and actuality.[15]

> It is ... through culture that the individual acquires standing and actuality. His true *original nature* and substance is the alienation [*Entfremdung*] of himself as Spirit from his *natural* being. This alienation [*Entäußerung*] is, therefore, ... the *means*, or the *transition*, both of the [mere] *thought-form of substance* into *actuality*, and, conversely, of the *specific individuality* into *essentiality*.[16]

In these passages Hegel describes an historical process in which agents treat their "*natural* being" as "Spirit", their "individuality" as universal "essentiality". (This apparent historical process is equivalent to one described by Rousseau, in which individuals go beyond their particular wills to form a universal will.[17])

The second process Hegel describes using these terms is structurally similar, but with the poles almost exactly reversed. That is, instead of agents alienating their individuality in treating this as part of the universal (as above) or as the dictates of a supra-individual God (as in Fichte), it is instead a supra-individual, universal subject (God, the Logical Idea, or Spirit) which alienates itself, manifesting itself in more concrete forms, i.e. in human beings. Here, Hegel's use of *Entäußerung* draws on another source, namely Luther's use of it to translate *kenosis* in a passage that describes God's self-emptying in Christ.[18] Thus, Hegel writes that:

> in the actualization of its [Spirit's] Notion, in being Spirit, it passes over into *being-for-another*, its self-identity becomes an *actual*, self-*sacrificing* absolute Being; it becomes a *self*, but a mortal, perishable self. ... [An] alienated [*entfremdeten*] self[19]

> this beyond, through the alienation [*Entäußerung*] of the eternal Being, has entered the actual world, the actuality is an uncomprehended, sensuous actuality.[20]

---

[15] Hegel, *Phenomenology of Spirit*, 297 (§488).
[16] Hegel, 298 (§489), translation amended.
[17] Rousseau, *The Social Contract*.
[18] Chitty, 'Hegel and Marx', 490.
[19] Hegel, *Phenomenology of Spirit*, 325 (§532).
[20] Hegel, 326 (§534), translation amended.

Thus, in this second process "The immediate in-itself of Spirit … gives itself the shape of [individual] self-consciousness".[21] This process is clearly distinct from the first, for it describes a metaphysical phenomenon rather than a historical process in which agents *treat* something as independent: Spirit does not simply *treat* something internal to itself as external, but rather actually *becomes* something else.

Hegel, then, describes two processes of alienation, each travelling in opposite directions:

> Of this Spirit, which has abandoned the form of Substance and enters existence in the shape of self-consciousness, it may therefore be said—if we wish to employ relationships derived from natural generation—that it has an *actual* mother but an *implicit* father. For *actuality* or self-consciousness, and the *in-itself* as substance, are its two moments through whose reciprocal alienation [*Entäußerung*], each becoming the other, Spirit comes into existence as this their unity.[22]

The switch of subject involved in this second process is severely criticised by both Feuerbach and Marx, who name it "inversion".[23] In developing their own accounts of alienation, therefore, both draw only on the former process.

In the passages quoted above we can see a difference in emphasis evoked by each of the terms *Entäußerung* and *Entfremdung*. *Entäußerung* draws attention to the fact that it is some agent's own qualities which are externalised in something, whether this is individuals' qualities being externalised in Spirit or those of the "eternal being" being externalised in human beings. One translation therefore uses "self-emptying" for this term, in keeping with the Lutheran heritage noted above.[24] This emphasis fits with the etymology of the term, which we saw above derives from *Äußerung* meaning "expression" or "manifestation". In contrast, *Entfremdung* emphasises the separateness from the externalising agent of that which is externalised. For instance, two of its three uses above are followed by "from", suggesting this emphasis on separation. Again this fits with the etymology: *Entfremdung* more frequently emphasises the *alien*-ness—the *fremd*-ness—of that

---

[21] Hegel, 458 (§757).
[22] Hegel, 457 (§755), translation amended.
[23] See Feuerbach, 'Preliminary Theses on the Reform of Philosophy', 159, 161. For this form of criticism of Hegel by Marx, see Marx, 'Critique of Hegel's Doctrine of the State', 58–65; for his use of 'inversion' to summarise this criticism, see Marx, 'Economic and Philosophical Manuscripts', 396.
[24] Pinkard, 'Translator's Note', xlii.

which is manifested. This difference in emphasis is taken up, although not always systematically, in Feuerbach and Marx's writings.

1.3 Feuerbach

Feuerbach's use of *Entäußerung* and *Entfremdung* can be understood as an adoption and extension of Fichte's original usage and the equivalent first process described by Hegel. Indeed, Feuerbach's primary interest is precisely that in which Fichte coined the former term, namely the relationship between religion and human agents, and an explanation for the existence of religion. (We can also perhaps see the influence of Hegel's discussion of the Unhappy Consciousness, in which he similarly deals with religion.[25]) Thus he uses *Entäußerung* in describing this process:

> The personality of God is thus the means by which man converts the qualities of his own essence into the qualities of another being,—of a being external to himself. The personality of God is nothing else than the alienated [*entäußerte*, *vergegenständlichte*] personality of man.[26]

> In religion the essence of man is regarded as separate from man. The activity, the grace of God is the alienated [*entäußerte*] spontaneity of man, Free Will made objective.[27]

And similarly, *Entfremdung*:

> Why then dost thou alienate [*entfremdest*] man's consciousness from him, and make it the self-consciousness of a being distinct from man, of that which is an object to him?[28]

> But as religion alienates [*entfremdet*] our own essence from us, and represents it as not ours, so the water of baptism is regarded as quite other than common water; for it has not a physical but a hyperphysical power and significance[29]

Thus, these terms again describe a process in which agents treat something which depends on or is internal to them as something which has an independent existence "outside" them. As we can see, Feuerbach names several different phenomena as being that which is so treated. However, most frequently this is said to be human

---

[25] Hegel, *Phenomenology of Spirit*, 119–38.
[26] Feuerbach, *The Essence of Christianity*, 186, translation amended.
[27] Feuerbach, 196 translation amended.
[28] Feuerbach, 189, translation amended.
[29] Feuerbach, 194.

beings' "essence" [*Wesen*]. I will explicate the meaning of this term in Feuerbach's writing in section 3 below.

Feuerbach follows Hegel in using both of the terms *Entäußerung* and *Entfremdung* to refer to this process as a whole, and in often using each term to emphasise or draw attention to two distinct aspects or stages of this process. *Entäußerung*'s emphasis that it is *the agents' own* qualities which undergo the process of alienation can be seen in the first two passages quoted above. For this reason, George Eliot, who originally translated Feuerbach's *The Essence of Christianity* into English, often translates *Entäußerung* as "projection". Similarly, *Entfremdung*'s emphasis that that which undergoes the process of alienation comes to be seen in a new form, as separate, distinct and "other" to the agents to whom it originally belongs or in whom it originally (or "really") exists can be seen in the second two passages. (Indeed, *Entfremdung* is also used occasionally to describe one thing as being or becoming separated from another, without being associated with the process of alienation described above.[30])

So far, we have seen the similarities between Feuerbach and Fichte's accounts. Both describe a process in which something that depends upon an agent—the subjective moral law in Fichte, the individual's essence (and various other things) for Feuerbach—are treated as or believed to be properties or predicates of a God which is independent of them. However, there are important differences between Feuerbach's use of these terms and Fichte's use of *Entäußerung*.

First, whereas Fichte's Kantian-inspired investigation of religion was conducted using a framework that focused on individual agents, Feuerbach's investigation, as we will see in section 3, eschewed this framework in favour of one whose point of departure was the human species or society as a whole. For Feuerbach, the process of alienation that results in religion involves the alienation of something not by individuals but by societies and species.

Second, the explanatory component of Feuerbach's account is considerably more varied than Fichte's. Feuerbach's explanation of why agents alienate something of themselves unto God includes some statements that are reminiscent of Fichte's explanation:

---

[30] See, for example, Feuerbach, 114, 146.

> God is a deeply moving object, enrapturing to the imagination; whereas the idea of humanity has little power over the feelings, because humanity is only an abstraction; and the reality which presents itself to us in distinction from this abstraction is the multitude of separate, limited individuals.[31]

However, Feuerbach suggests many other explanations, for instance that religion is a form of wish fulfilment,[32] or the creation of an "audience for his griefs".[33]

Further, Feuerbach perhaps most famously explains the alienation of human beings' essence in God as a necessary step on the way to self-knowledge:

> Religion is man's earliest and also indirect form of self-knowledge.… Man first of all sees his essence as if *out of* himself, before he finds it in himself. His own essence is in the first instance contemplated by him as that of another being.[34]

While, then, Feuerbach's use of these terms implies a less specific explanation at the level of the overall purpose that this process serves, this is replaced by the suggestion that this process results from a sort of *mistake*, a "negation of human sense and the human understanding" encoded in "the web of contradictions and delusions called theology" such that religion is based on "absurdity"[35] and represents a "dogma of incomprehensibility".[36] Elsewhere, Feuerbach characterised this mistake by describing Christianity as "nothing more than a fixed idea".[37] Of course, as we have seen, this mistake is sometimes said to have positive consequences (i.e. agents' attainment of self-knowledge).

Third, Feuerbach extends his use of *Entäußerung* and *Entfremdung* beyond explicitly religious targets. In particular, he uses them to describe what he saw as the same process occurring in Hegel's work:

> To abstract means to posit the *essence* of nature *outside nature*, the *essence* of man *outside man*, the *essence* of thought *outside the act of thinking*. The Hegelian philosophy has alienated [*entfremdet*] man *from himself* in so far as its whole system is based on these acts of abstraction.[38]

---

[31] Feuerbach, 127–28.
[32] Feuerbach, 101.
[33] Feuerbach, 101–2.
[34] Feuerbach, 11, translation amended.
[35] Feuerbach, ix.
[36] Feuerbach, 181.
[37] Feuerbach, quoted in Lawrence, 'Foreword'.
[38] Feuerbach, 'Preliminary Theses on the Reform of Philosophy', 157.

As such, Feuerbach's use of *Entäußerung* and *Entfremdung* extends the use of these terms to another phenomenon that he sees as following the same process as he saw as occurring in religion:

> The *essence* of speculative philosophy is nothing other than the *rationalized, realized, actualized essence of God*. The speculative philosophy is the *true, consistent, rational* theology.[39]

Finally, Feuerbach's use of the terminology of alienation takes on a critical, normative aspect which did not exist in Fichte's use. For Feuerbach, to identify a phenomenon as following the process of alienation is to condemn it. For instance, the quotation at the beginning of our discussion of Feuerbach which described God as "the alienated [*entäußerte*] spontaneity of man, Free Will made objective" is preceded by the claim that "Nothing is more perverse than the attempt to reconcile miracle with freedom of inquiry and thought, or grace with freedom of will".[40] Elsewhere, Feuerbach—effectively responding to Fichte's explanation, implying that it defeats the purpose of the Kantian philosophy of autonomy which it is trying to save—writes that:

> revelation generates moral actions, which do not, however, proceed from moral actions;—moral actions, but no moral dispositions. Moral rules are indeed observed, but they are alienated [*entfremdet*] from the inward disposition, the heart, by being represented as the commandments of an external lawgiver, by being placed in the category of arbitrary laws, police regulations.… But the belief in revelation not only injures the moral sense and taste,—the aesthetics of virtue; it poisons, nay it destroys, the divinest feeling in man—the sense of truth, the perception and sentiment of truth.[41]

Here, Feuerbach describes the God that is the end product of alienation affecting agents' behaviour much as Fichte did—motivating them to act in accordance with the commandments of God, in spite of their desires—but with a critical twist: because agents act motivated by God's commandment instead of their own "moral dispositions", it "injures … the aesthetics of virtue". (Here Feuerbach is applying a common Christian criticism of Judaism, expressed in Kant's *Religion within the Limits of Reason Alone*, to religion generally, including to Christianity itself.[42])

---

[39] Feuerbach, 'Principles of the Philosophy of the Future', 178.
[40] Feuerbach, *The Essence of Christianity*, 196.
[41] Feuerbach, 173.
[42] Kant, *Religion within the Boundaries of Mere Reason and Other Writings*, 95.

Of course, as we saw above, Feuerbach also occasionally paints this process in a more positive light, as a necessary step on the way to self-knowledge. I will examine this theme below: for now, let us focus on the fact that Feuerbach's use of these terms often carries with it a critical, normative dimension.

1.4 The other Young Hegelians: Bruno Bauer, Max Stirner, and Moses Hess

Feuerbach's way of talking about religion as alienated—incorporating both description of the process outlined above and its critical, normative implications—became widespread amongst the so-called Young Hegelians, a group which included writers such as Bruno Bauer, Max Stirner, and Moses Hess (as well as Feuerbach, Marx and Engels themselves), all of whom described religion or Christianity in these terms.[43] These writers, however, carried Feuerbach's critique of religion over into a critique of the political[44] and economic[45] spheres. For instance, Bauer writes of the "alienation with respect to him [the individual] of the totality of the state",[46] while Stirner invokes Feuerbach (in the process of criticising him) when he describes the kind of liberalism he saw as underpinning the modern state—the "*state-religion*"—as follows:

> The **human** *religion* is only the last metamorphosis of the Christian religion. For liberalism is a religion because it separates my essence from me and sets it above me, because it exalts 'man' to the same extent as any other religion does its God or idol, because it makes what is mine into something otherworldly, because in general it makes some of what is mine, out of my qualities and my property, something alien[47]

While itself having historical roots—as mentioned above, one of Hegel's applications captures an idea employed by Rousseau in the realm of politics, while Hegel himself connected his discussion of religion and the Unhappy Consciousness with the Roman

---

[43] Hess, *The Holy History of Mankind and Other Writings*, 105, 127; Stirner, *The Ego and Its Own*, 307; Bruno Bauer, quoted in Moggach, *The Philosophy and Politics of Bruno Bauer*, 65, 77, 122.

[44] Rosen, *Bruno Bauer and Karl Marx*, chap. 9; Rosen, 'The Influence of Bruno Bauer on Marx's Concept of Alienation', 63; Hess, *The Holy History of Mankind and Other Writings*, 103; Stirner, *The Ego and Its Own*, 68, 77, 97, 175, 273.

[45] Moggach, *The Philosophy and Politics of Bruno Bauer*, 167; Hess, *The Holy History of Mankind and Other Writings*, 80; Stirner, *The Ego and Its Own*, 279.

[46] Bauer, quoted in Brudney, *Marx's Attempt to Leave Philosophy*, 120.

[47] Stirner, *The Ego and Its Own*, 158.

empire[48]—, these Young Hegelian writers' application of Feuerbach's critique of religion to political and economic matters is crucial to the development of Marx's ideas. I will not here examine these in detail, but will turn to them in chapter 2 when discussing Marx and Engels' criticisms of them. However, it is worth highlighting that just as Feuerbach's terminology had a clear edge of criticism, these other Young Hegelian writers employed this terminology not so much to explain but to criticise the social institutions which they labelled *entäußerte* and *entfremdet*. As Bruno Bauer put it, for them "philosophy is criticism of the existing reality".[49]

2 *Entäußerung* and *Entfremdung* in Marx's work of 1843-4

Marx, too, follows Feuerbach in using the terms *Entäußerung* and *Entfremdung* to describe a process resulting in religion. Thus in 1844 he writes:

> in religion the spontaneous activity of the human imagination, the human brain and the human heart detaches itself from the individual and reappears as the alien [*fremde*] activity of a god or of a devil[50]

And he follows Feuerbach too in applying this idea to Hegel's work:

> This movement in its abstract form as dialectic is therefore regarded as *truly human life*. And since it is still an abstraction, an alienation [*Entfremdung*] of human life, it is regarded as a *divine process*, but as the divine process of man. It is man's abstract, pure, absolute being (as distinct from himself), which itself passes through this process.[51]

Indeed, he writes that:

> [one of] Feuerbach's great achievement[s] is: (1) To have shown that philosophy [i.e. Hegelian philosophy] is nothing more than religion brought into thought and developed in thought, and that it is equally to be condemned as another form and mode of existence of the alienation [*Entfremdung*] of man's essence.[52]

However, he also follows Bauer, Stirner and Hess in advocating a shift in the object of investigation away from religion and towards social, political, and economic institutions. For instance, in 1843 Marx wrote to Arnold Ruge that:

---

[48] Hegel, *Phenomenology of Spirit*, 290–94.
[49] Quoted in Rosen, *Bruno Bauer and Karl Marx*, 124.
[50] Marx, 'Economic and Philosophical Manuscripts', 326.
[51] Marx, 396, translation amended.
[52] Marx, 381, translation amended.

> Feuerbach's aphorisms seem to me incorrect only in one respect, that he refers too much to nature and too little to politics.[53]

And later that year wrote that:

> It is the immediate *task of philosophy*, which is in the service of history, to unmask self-alienation [*Selbstentfremdung*] in its *unholy forms* once the *holy form* of human self-alienation [*Selbstentfremdung*] has been unmasked. Thus the criticism of heaven turns into the criticism of earth, the *criticism of religion* into the *criticism of law* and the *criticism of theology* into the *criticism of politics*.[54]

Like the other Young Hegelians described above, before 1845 Marx employed the language of *Entäußerung* and *Entfremdung* to carry out this unmasking. For instance, he employs *Entfremdung* in describing the existence of the political state:

> Hitherto, the political constitution has always functioned as the religious sphere, the religion of the life of the people, the heaven of its universality as opposed to the earthly existence of its actual reality.… Political life in the modern sense is the scholasticism of the life of the people. The monarchy is the perfect expression of this alienation [*Entfremdung*]. The republic is the negation of that alienation, but within its own sphere.[55]

In his 1844 writings Marx's focus turned largely to economic questions. One topic he addresses is the nature of value and money, and in discussing this he again employs *Entäußerung* and *Entfremdung*:

> Money is the universal and self-constituted value of all things. It has therefore deprived the entire world – both the world of man and of nature – of its specific value. Money is the alienated [*entfremdete*] essence of man's work and existence; this alien [*fremde*] essence dominates him and he worships it.[56]

> The inversion and confusion of all human and natural qualities, the bringing together of impossibilities, the divine power of money lies in its *nature* as the alienated and alienating [*entfremdeten, entäußernden*] *species-essence* of man which alienates itself by selling itself. It is the alienated [*entäußerte*] *capacity* of *mankind*.[57]

Similarly, he describes the contemporary division of labour in these terms:

> The *division of labour* is the economic expression of the *social nature of labour* within alienation [*Entfremdung*]. Or rather, since *labour* is only

---

[53] Marx, 'Letter to Arnold Ruge, May 1843', 400.

[54] Marx, 'A Contribution to the Critique of Hegel's Philosophy of Right. Introduction', 244–45, translation amended.

[55] Marx, 'Critique of Hegel's Doctrine of the State', 89, translation amended.

[56] Marx, 'On the Jewish Question', 239, translation amended. See also Marx, 'Economic and Philosophical Manuscripts', 361.

[57] Marx, 'Economic and Philosophical Manuscripts', 377, translation amended.

> an expression of human activity within alienation [*Entäußerung*], an expression of life as alienation of life [*Lebensentäußerung*], the *division of labour* is nothing more than the *alienated* [*entfremdete, entäußerte*] positing of human activity as a *real species-activity* or as *activity of man as a species-being*.[58]

> The consideration of the *division of labour* and *exchange* is of the highest interest, because they are the *perceptibly alienated* [*entäußerten*] expressions of human *activity* and *essential powers* as *species*-activity and *species*-powers.[59]

It is clear from these passages that a central theme in Marx's discussion is that it is human beings' *essence* that is alienated in these social institutions. In this sense, Marx's account of the alienation that results in social institutions mirrors Feuerbach's account of the alienation that results in religion quite closely. The concept of the human essence with which both Marx and Feuerbach operate, as well as the important differences between their respective accounts, will be taken up in section 3. For now, let us take stock of the use to which Marx puts these terms in these passages.

As used by Fichte through to Marx, the terms *Entäußerung* and *Entfremdung* refer to a *process*, in which an agent or agents externalise something (the moral law, their free will, their intellectual faculties, or their "essence") through creating some entity (God, the Logical Idea, or a political or economic institution) which they treat as independent of them. For Feuerbach and the other Young Hegelians such as Stirner and Bauer, the use of these terms to refer to this process further implies criticism of that which is created. I will discuss the sense in which this is the case also for Marx in the following chapter, but it is clear that in Marx's pen these terms have a similarly critical sense: as we saw above in the case of money, the product of this process not only "dominates" human beings, but has "deprived the entire world – both the world of man and of nature – of its specific value".[60]

We have also seen that in both Hegel's and Feuerbach's writings, the two terms often emphasise or draw our focus towards one aspect of this process. This, too, carries over into Marx's use. Consider, for instance, the uses of these terms in the two passages above concerning money. On the one hand, it is the "capacity of mankind" that is said to be *entäußerte* in money, this term drawing attention to the fact that

---

[58] Marx, 369, translation amended.
[59] Marx, 374.
[60] Marx, 'On the Jewish Question', 239, translation amended. See also Marx, 'Economic and Philosophical Manuscripts', 361.

it is this capacity that is expressed or externalised in money. On the other hand, money is said to be the *entfremdete* essence of man's work and existence, this drawing attention to the fact that this essence is now "alien [*fremde*]" (indeed, so alien that it "dominates him"). Chris Arthur is therefore right to say that:

> Probably Marx uses *Entausserung* when he has in mind that man *loses* something of himself through alienation, and *Entfremdung* to mark its appearance as something *other* than himself.[61]

Nevertheless, it is important to note that both terms are primarily used to refer to different aspects of an overall *process*, rather than to these phenomena in isolation.[62] And Marx just as often uses either term to refer to the process I have described as a whole, rather than to a particular part of it. This variable usage, along with the various other meanings that each of these terms have in Marx's writing, has no doubt confused many readers, for it seems that the same terms are often used with strikingly different meanings even in contexts in which he is talking about broadly the same thing. For instance, in the context of religious alienation, one and the same term might describe the agent's externalisation of themselves in God, the becoming-independent of God, the final state in which agents see themselves as governed by the will of an independent God, and so on, or to describe the overall process which includes all of these stages. It is only by acknowledging that these terms' primary referent is a process that this variable usage can be understood.

3 Marx and Feuerbach

Above I have primarily highlighted the commonalities between Marx's discussion of the alienation of the human essence in social institutions and Feuerbach's discussion of the alienation of the human essence in religion. However, the relationship between Feuerbach and Marx's accounts is more complex than I have presented it so far, and to understand Marx's account we need to highlight the important differences between the phenomena with which these writers are each primarily concerned, i.e. between religion and social institutions. This will lead us to an examination of their respective notions of the human essence, i.e. of that which they each take to be externalised in religion and social institutions respectively. It will then require us,

---

[61] Arthur, *Dialectics of Labour*, 148.

[62] Marx does, however, occasionally use them to refer to externalisation (for *Entäußerung*) or separation (for *Entfremdung*) shorn of their involvement in this process.

in chapter 2, to explore what it means for the externalised human essence to become estranged or alien from those agents whose essence it is.

3.1 The difference between God and social institutions

For Marx there is a fundamental difference between God and social institutions, namely that while God does not exist, social institutions do. Given this difference, the process of alienation that creates or results in these different phenomena—while, as we have seen, parallel in significant respects—are necessarily quite different. As he puts it:

> Religious alienation [*Entfremdung*] as such takes place only in the sphere *of consciousness*, of man's inner life, but economic alienation [*Entfremdung*] is that of *real life*[63]

Marx criticises Hegel at length for not taking this difference into account when discussing social institutions. For instance, he writes that:

> Hegel commits a double error.
>   The first appears most clearly in the *Phenomenology*, which is the birthplace of Hegelian philosophy. When, for example, Hegel conceives wealth, the power of the state, etc., as entities alienated [*entfremdete*] from the being of man, he conceives them only in their thought form . . . They are entities of thought, and therefore simply an alienation [*Entfremdung*] of *pure*, i.e. abstract, philosophical thought. Therefore the entire movement ends with absolute knowledge. What these objects are alienated [*entfremdet*] from and what they confront with their claim to reality is none other than abstract thought.[64]

This difference is crucial to understanding the differences between Feuerbach and Marx's respective discussions of alienation.

Before turning to these differences, however, it is worth examining their respective accounts of what it is that is alienated in God and in social institutions, i.e. their respective accounts of the human essence. These accounts are remarkably similar in a number of respects, although Marx's account draws importantly different conclusions.

---

[63] Marx, 'Economic and Philosophical Manuscripts', 349.
[64] Marx, 384.

3.2 The shared essence of Marx and Feuerbach

As we have seen, both Feuerbach and Marx (in 1843-4 at least) employ a concept of "essence" [*Wesen*] frequently in their respective discussions of the process of alienation that results in religion and social institutions. If, then, we are to understand more precisely the process that these authors are describing, it would help to understand their respective conceptions of the human essence.

Literature on Marx predominantly discusses his account of the human essence when reconstructing his normative evaluation of the social institutions of the contemporary world, rather than when asking what he takes to be externalised in these. I will turn to the former question in chapter 2: it is the latter question with which we are here concerned.

There is a common story about Marx's concept of the human essence, his relation to Feuerbach concerning this concept, and his changing views on both.[65] This is that prior to 1845 Marx employed a concept of the human essence inherited from Feuerbach, and that 1845 marks a turning point at which Marx abandoned this concept and criticised his former ally for holding this view. We can understand this story most clearly with reference to the sixth of Marx's 1845 *Theses on Feuerbach*, in which he is thought to have explicitly "broken" with this concept:

> Feuerbach resolves the religious essence into the human essence. But the human essence is no abstraction inherent in each single individual.
> In its actuality it is the ensemble of the social relations.
> Feuerbach, who does not enter upon a criticism of this actual essence, is consequently compelled:
> 1. To abstract from the historical process and to fix the religious sentiment as something by itself and to presuppose an abstract – isolated – human individual.
> 2. Essence, therefore, can be comprehended only as "genus", as an internal, dumb generality which naturally unites the many individuals.[66]

From this passage, and the role it is commonly given in the literature as evidencing Marx's break with his previous Feuerbachianism, we can see what this account takes Marx's pre-1845 notion of the human essence to be: it was as an "abstraction inherent in each single individual", "an internal, dumb generality which naturally unites the many individuals", which he here abandons in favour of some alternative

---

[65] For perhaps the most explicit rendition of this story, see Althusser, *For Marx*, 277–78.
[66] Marx, 'Concerning Feuerbach', 423, translation amended.

approach. (An alternative story, told by, amongst others, Norman Geras,[67] downplays the extent to which 1845 marks such a break: for these authors, it seems, Marx maintained a concept of the human essence as an "internal, dumb generality" throughout his post-1845 writings too.)

In order to clarify this story, let us be more specific about what Marx's pre-1845 notion of essence is said to be. I take its central claim to be that Feuerbach and Marx believed there to be a set of concrete needs, characteristics and capacities that all human beings have ("within them", so to speak) in every possible world. This account comes out most explicitly in David Leopold's account, which draws on Thomas Hurka's *Perfectionism*.[68]

The problem with this story, however, is that it fits with neither Feuerbach's nor pre-1845 Marx's explicit accounts of the human essence. That is not to say that these authors took there to be no characteristics that all human beings have in every possible world—as I will argue in section 3.4, Marx did think that some such characteristics could be identified at quite a general level—but rather that this is not entailed by their description of something as part of the human essence. Rather, both writers employ an interesting and somewhat idiosyncratic notion of essence. First, they describe certain characteristics that human beings possess here and now, including historically variable characteristics, as part of the human essence: in this sense—which I shall call the *contingentist* aspect—it seems that this notion of essence is specifically not "an abstraction", nor does it describe something possessed by every human being in every possible world. Second, they employ the concept to describe features of human life which are not "internal" or "inherent in each single individual". Rather, both writers describe supra-individual features of human beings, and in particular the social relations in which they live, as part of the human essence. I will now outline this peculiar notion of essence, before turning to how these writers employ it in their respective accounts of alienation.

---

[67] Geras, *Marx and Human Nature*.

[68] Leopold, *The Young Karl Marx*, chap. 4; Hurka, *Perfectionism*. The definition of essence as consisting in characteristics that human beings possess 'in all possible worlds' is Hurka's.

3.2.1 Feuerbach's 'essence' as contingent

Feuerbach's comments regarding the human *Wesen* can be divided into two rough groups. First are those passages in which he describes the predicates that the human "essence" consists in, primarily in describing these as alienated in the creation of God or religion (discussed in this subsection). Second are his references to the 'social' or 'species' nature of the human essence (discussed below).

Feuerbach's tendency to describe a wide range of predicates as the "human essence" can give the reader the impression of severe inconsistency. At various points, the claim that God is a manifestation of the "human essence" is cashed out as meaning that God signifies human beings' intellect[69]; their social institutions such as marriage and property[70]; their capacity for love[71]; their thoughts, feelings and passions[72]; that which they wish to be[73] and so on up to and including all of "the predicates … of nature and mankind: … in short, everything".[74] As Eugene Kamenka puts it:

> Unfortunately, Feuerbach did not always proceed with sufficient care. He tended to see the 'nature' of a thing in excessively simple terms, and his genetic-critical method rather encouraged him in this. It led him, all too often, to find the 'secret' of its nature in its *sine qua non* and then to make that *sine qua non* the whole of its 'nature': 'Religion *is* dependence', 'man *is* the sensory'. As a new *sine qua non* came up – for one and the same thing may have several conditions necessary for its existence – the second term in the identity changed.[75]

> Feuerbach presents each of these reductions so forcefully, with so much rhetoric, that he appears to regard each of them as the true *essence* in terms of which the whole of religion should be explained.[76]

This variety of "essential" predicates ought to give pause for thought to those who read him as holding a more standard conception of essence, for it suggests a conception which leaves little outside. As Feuerbach himself puts it, the central theme of his *Essence of Christianity* is:

---

[69] Feuerbach, 'Principles of the Philosophy of the Future', 182, 186.
[70] Feuerbach, *The Essence of Christianity*, 224.
[71] Feuerbach, 224.
[72] Feuerbach, 19.
[73] Feuerbach, 'Principles of the Philosophy of the Future', 221; Feuerbach, '"The Essence Of Christianity" in Relation to "The Ego And Its Own"', 83.
[74] Feuerbach, '"The Essence Of Christianity" in Relation to "The Ego And Its Own"', 81.
[75] Kamenka, *The Philosophy of Ludwig Feuerbach*, 96.
[76] Kamenka, 56.

> nothing other than the overcoming of the split into an essential and unessential I — the deification, that is, the presentation, the regarding of the whole man from head to foot.[77]

More troubling still for the standard account is the contingency of many of these "essential" predicates. Feuerbach sometimes traces the change and development of societies through their Gods: for example, the change from worshiping thunder directly to seeing it as a tool used by an anthropomorphic God is said to reflect changes in our relationship to nature, while the development of the "universalism" of Christianity is described as reflecting the breakdown of national boundaries characteristic of the Roman empire.[78] If our Gods manifest our essence, then it seems these changes must reflect changes in what this essence consists in. Elsewhere, Feuerbach portrays this contingency as so complete that:

> human nature presents an infinite abundance of different predicates, and for that very reason it presents an infinite abundance of different individuals. Each new man is a new predicate, a new phasis of humanity. As many as are the men, so many are the powers, the properties of humanity.[79]

Further, this contingency can be seen in Feuerbach's description of human beings' "essence" as a creation of human beings themselves: "Man is a product of man, of culture, of history".[80] For Feuerbach this is reflected in Christian doctrine by the inclusion in the sacrament of bread and wine which, as human-made, represent "the truth that Man is the true God and Saviour of man".[81] Such historically-produced predicates clearly cannot be features of all human beings in all possible worlds.

For Feuerbach, then, the predicates that constitute the human essence are contingent, and not necessarily characteristics of all human beings in all possible worlds. Rather than pointing to such universal characteristics, Feuerbach includes in the human *Wesen* those predicates which he thinks are particularly important in explaining those phenomena—most commonly, the dominant religious and philosophical beliefs—with which he is concerned.[82] This explanatory role requires

---

[77] Feuerbach, "'The Essence Of Christianity" in Relation to "The Ego And Its Own"', 83.
[78] Kamenka, *The Philosophy of Ludwig Feuerbach*, 44–45.
[79] Feuerbach, *The Essence of Christianity*, 19.
[80] Feuerbach, quoted in Kamenka, *The Philosophy of Ludwig Feuerbach*, 86.
[81] Feuerbach, *The Essence of Christianity*, 227.
[82] This, of course, seems to render Feuerbach's claim that God is the externalisation of the human essence somewhat tautological. My goal here is not to make Feuerbach's claims coherent, but to highlight those aspects of his account which had a particular influence on Marx.

only the empirical claim that human beings here and now have particular characteristics. Thus Feuerbach saw himself as championing an empiricism or "sensualism" which examined human beings as they exist, here and now. To know human beings in their "Individuality" required for Feuerbach a methodology that focuses not on "absolute essence" but "sensible essence".[83]

Feuerbach's advocacy for empiricism connects to his emphasis on *determinacy*, on "think[ing] the *concrete* not in an abstract but a concrete way, which acknowledges the real *in its reality*",[84] and which focused on "this person, this thing, that is, the particular".[85] Arguments of this kind are most commonly cited in the secondary literature as targeted against Hegel's starting points in the *Logic* and the *Phenomenology of Spirit*, starting as they do with abstracted concepts of "being", "here", "this" and "now"[86]: for instance, Feuerbach writes of "being" that:

> Being is *not a general concept that can be separated from things. It is one with that which is*. It is thinkable only as mediated, that is, only through the predicates which constitute the essence of a thing.… The being of the fish is its being in water, and from this being you cannot separate its essence.… The notion of *being resulting from a removal of* all *essential qualities* from things is only *your notion* of being – a fabricated, invented being, a being *without the essence* of Being.[87]

However, this form of argument is for Feuerbach equally applicable to abstracted conceptions of the "human essence":

> humanity is only an abstraction; and the reality which presents itself to us in distinction from this abstraction is the multitude of separate, limited individuals.[88]

Indeed, as we shall now see, Feuerbach primary focuses not on these separate individuals, but on these individuals in their relations to their "object",[89] i.e. to others and to the world: just as the fish cannot be separated in analysis from its being in water, "Feuerbach [who has a habit of speaking of himself in the third person] only treats the essence of man in society".[90]

---

[83] Feuerbach, '"The Essence Of Christianity" in Relation to "The Ego And Its Own"', 83–84.
[84] Feuerbach, 'Principles of the Philosophy of the Future', 222.
[85] Feuerbach, 225.
[86] Wartofsky, *Feuerbach*, 189; Johnston, *Between Transcendence and Nihilism*, 133; Feuerbach, 'Principles of the Philosophy of the Future', 214–16, 233–35, 238–39.
[87] Feuerbach, 'Principles of the Philosophy of the Future', 214–15.
[88] Feuerbach, *The Essence of Christianity*, 127–28.
[89] Feuerbach, 'Principles of the Philosophy of the Future', 181–82.
[90] Feuerbach, '"The Essence Of Christianity" in Relation to "The Ego And Its Own"', 91.

Despite Feuerbach's stress on the contingent character of the human essence, he does occasionally talk as if certain essential features are more stable or fundamental. In particular, in the introduction to his *Essence of Christianity*, he writes:

> What, then, *is* the essence of man, of which he is conscious, or what constitutes the specific distinction, the proper humanity of man? Reason, Will, Affection. To a complete man belong the power of thought, the power of will, the power of affection. The power of thought is the light of the intellect, the power of will is energy of character, the power of affection is love. Reason, love, force of will, are perfections—the perfections of the human being—nay, more, they are absolute perfections of being. To will, to love, and to think, are the highest powers, the absolute essence of man as man, and the basis of his existence.[91]

Here it seems that, however essential are the other predicates he attributes to human beings, these three "powers"—Reason, Will, and affection—are somehow *more* essential: they are the "perfections of man", human beings' "*highest* powers" and "*absolute* essence". Nevertheless, even here this claim fits uncomfortably with the standard account. As Feuerbach goes on to say:

> The divine trinity in man, above the individual man, is the unity of reason, love, will. Reason, Will, Love, are not powers which man possesses, for he is nothing without them, he is what he is only by them; they are the constituent elements of his essence, which he neither has nor makes, the animating, determining, governing powers—divine, absolute powers—to which he can oppose no resistance.[92]

Thus, these "powers" *transcend* individual human beings, are not possessed by them but are somehow implicated in their becoming what they are. Clearly, then, they are not understood as "internal" to individual human beings. These claims, however, are rather obscure, and it is difficult to know what to make of them. Nevertheless, some light can be shed on them by examining the second theme prominent in Feuerbach's discussion.

3.2.2 Feuerbach's 'essence' as social

The second theme in Feuerbach's discussion is that the human essence is *social*, that he is primarily concerned with the essence not of individuals but of the "species".

---

[91] Feuerbach, *The Essence of Christianity*, 2–3, translation amended.
[92] Feuerbach, 3.

It is often claimed that Marx's term *Gattungswesen*, or species-being/essence, derives from Feuerbach.[93] This claim is misleading, however: Feuerbach uses this term only once in his writings,[94] and there in a quite different sense to that in which Marx uses it (see section 3.4 below). Nevertheless, it is true that Feuerbach often associates the human essence with the notion of the human species. We saw this association above, when Feuerbach posed the question "what is the essence of man of which he is conscious, or what is that which constitutes in him his *species*, his humanity proper?" This connection can be found throughout Feuerbach's writings.

This connection is crucial for Feuerbach's story about human beings' attainment of self-knowledge through religion, mentioned previously. For Feuerbach, human beings' alienation of their essence in religion allows them to see this essence for what it is and therefore to come to know their own essence. Elsewhere, this same idea is expressed not using the term *Wesen*, but in terms of knowledge of the human species:

> man's idea of God is the idea of the human individual of his own species, … God as the totality of all realities and perfections, is nothing other than the totality of the qualities of the species compendiously put together in him for the benefit of the limited individual, but actually dispersed among men and realizing themselves in the course of world history.[95]

Elsewhere, Feuerbach argues that it is human beings' ability to think about themselves as part of a species that enables them to "make the essential nature of other things and beings an object of thought" and thus to attain (a particular kind of) "consciousness". Following this passage, he again equates human beings' essence and their species:

> Man is himself at once I and thou; he can put himself in the place of another, for this reason, that to him his species, his essence, and not merely his individuality, is an object of thought to him.[96]

This association is more than a mere idiosyncratic turn of phrase. Feuerbach emphasises repeatedly that when he talks about the human *Wesen* and its constitutive characteristics, he has in mind not the essential characteristics of any particular individual, but of human beings as a species (although, of course,

---

[93] See, for instance, Hanfi, 'Introduction', 44.
[94] Feuerbach, *The Essence of Christianity*, 127: 'the essence of the species'.
[95] Feuerbach, 'Principles of the Philosophy of the Future', 189.
[96] Feuerbach, *The Essence of Christianity*, 2, translation amended.

understood in a contingentist sense as the members of the species that exist here and now, not in the abstract):

> The single man *in isolation* possesses in himself the *essence* of man neither as a *moral* nor as a *thinking* being. The *essence* of man is contained only in the community, in the *unity of man with man* – a unity, however, that rests on the *reality* of the *distinction* between "I" and "You"[97]

Thus Feuerbach's claims about the predicates of the human *Wesen* do not concern the predicates of any individuals, but of society as a whole or of the sum of all individuals and of their relations to one another, their social relations. Similarly Feuerbach writes that "the nature of man exists indeed only in the opposition of I and Thou, man and wife",[98] and that "when I want to know you as an individual, I must not limit myself to knowing you alone, but must extend my knowledge beyond you, to your wife as well. To know an individual is necessarily to know at least two individuals" (and "On two, follows three", and so on): to know an individual in "essence" requires a mode of investigation which "extends its thoughts and feelings to the species, i.e. to other individuals".[99] Elsewhere this "essence" of human beings is expanded beyond even these social relations to incorporate "external nature" as well.[100]

One might be tempted to interpret Feuerbach's claims about self-knowledge in a more standard, internal sense: that he really means that individuals require social relations in order to *discover* what their own "essence" consists in. For instance, he writes:

> The other is my *thou*, – the relation being reciprocal, – my *alter ego*, man objective to me, the revelation of my own nature, the eye seeing itself. In another I first have the consciousness of humanity; through him I first learn, I first feel, that I am a man

However, this clearly does not capture Feuerbach's full meaning. The passage continues:

> in my love for him it is first clear to me that he belongs to me and I to him, that we two cannot be without each other, that only community constitutes humanity[101]

---

97 Feuerbach, 'Principles of the Philosophy of the Future', 244.
98 Feuerbach, '"The Essence Of Christianity" in Relation to "The Ego And Its Own"', 87.
99 Feuerbach, 86.
100 Feuerbach, *The Essence of Christianity*, 222 fn. 1.
101 Feuerbach, 131.

Here Feuerbach's claim is that the "humanity" that I become conscious of through the other is not one that was "within" me before I became conscious of it: it is constituted in this relation; without each other neither of us would "be" (who or what we are).

This theme of the human essence being the essence of the species can shed light on Feuerbach's frequent claim that the human essence, while a contingent phenomenon situated in the here and now, is in a sense transcendent and independent *vis-à-vis* any particular individual. We saw this claim above in Feuerbach's suggestion that the predicates of the human essence are "divine, absolute powers that he is powerless to resist". Similarly, language—"the realisation of the species"—entails "the transcendence of the individual separateness of its members".[102] As such, predicates of the human essence are said to "have an intrinsic, independent reality; they force their recognition upon man by their very nature".[103]

A related claim often found in Feuerbach's work is that the social relations that constitute human beings' essence themselves make individuals what they are (and in this sense, perhaps, shape their "essence" in the more standard, internal sense). For instance, he writes:

> Precisely because he [the individual] lives in the contemplation of the whole, he also lives in the consciousness that he is himself no more than a part, and that he is what he is only by virtue of the conditions which constitute him a member of the whole, or a relative whole.[104]

And:

> That he is, he has to thank Nature; that he is man, he has to thank man[105]

Feuerbach's most frequent example is the influence of society on individuals' thought *via* language-use,[106] but he offers others:

> Wit, acumen, imagination, feeling as distinguished from sensation, reason as a subjective faculty—all these so-called powers of the soul are

---

[102] Feuerbach, quoted in Wartofsky, *Feuerbach*, 182.
[103] Feuerbach, *The Essence of Christianity*, 18.
[104] Feuerbach, 141.
[105] Feuerbach, 70.
[106] Feuerbach, 'Towards a Critique of Hegel's Philosophy', 63; see also Wartofsky, *Feuerbach*, 172.

> powers of humanity, not of man as an individual; they are products of culture, products of human society.[107]

This theme is important, in particular in relation to Feuerbach's claims about human beings' *self-creation* discussed above: insofar as the human essence *qua* the characteristics of the species and the social relations between all human beings itself shapes the characteristics of individuals and what individuals become, this in turn shapes the contingent human essence of the future.

Of course, the use of the term essence or *Wesen* to refer to the predicates characteristic of the human species as a whole appears strange to our eyes. In order to reduce this strangeness, I would like to make two observations. First, the term *Wesen* has at least one relevant connotation which the English word "essence" does not, namely that of "the suggestion of a whole system that goes to make up an institution",[108] as in the words *Postwesen* and *Schulwesen*. Seen in this light, the use of *Wesen* to describe the social relations through which individuals interact is perhaps not so strange. Second, one way of understanding Feuerbach's use of *Wesen* might relate to the concept's function, i.e. that for Feuerbach that which is described as the essence of the species plays the functional role that in previous theorists' work was played by a more conventional notion of essence. For example, if in previous theorists' work the concept of essence was used to describe that which motivated individuals to act,[109] then a theorist who believed that the structure of social relations played a significant role in motivating individuals to act might explain this by describing these social relations as humans' true "essence". Regardless of the validity of these two suggestions, it is clear that, however idiosyncratic, the term *Wesen* at least invoked these connotations for Feuerbach, a point of crucial importance as we consider the role of this term in Marx's early work.

---

[107] Feuerbach, *The Essence of Christianity*, 71.

[108] Kamenka, *The Philosophy of Ludwig Feuerbach*, 37; See also 'Wesen | Translate German to English'.

[109] For instance, the Christian belief which, according to Feuerbach, held that "God is that being who acts in me, with me, through me, upon me, for me, is the principle of my salvation, of my good intentions and actions". Feuerbach, quoted in Wartofsky, *Feuerbach*, 277.

3.3 Marx's sixth thesis: an alternative reading

If these themes are as prominent in Feuerbach's writings as I have presented them, one might wonder what sense we are to make of Marx's sixth thesis, discussed above. Here it is again:

> Feuerbach resolves the religious essence into the human essence. But the human essence is no abstraction inherent in each single individual. In its actuality it is the ensemble of the social relations.
> Feuerbach, who does not enter upon a criticism of this actual essence, is consequently compelled:
> 1. To abstract from the historical process and to fix the religious sentiment as something by itself and to presuppose an abstract – isolated – human individual.
> 2. Essence, therefore, can be comprehended only as "genus", as an internal, dumb generality which naturally unites the many individuals.

As discussed above, the standard reading of this thesis is that in it Marx breaks with Feuerbach's and his own previous notion of the human essence as an "abstraction inherent in each single individual", paving the way for some alternative conception.

Let us examine this standard reading more closely. For it, the first sentence summarises Feuerbach's project, including a statement that Feuerbach has a particular conception of "the human essence". The second sentence objects to this conception, and the third proposes an alternative. The fourth sentence makes another criticism of Feuerbach (that, presumably because of his conception of the human essence, he does not criticise the ensemble of social relations). Marx then outlines some numbered consequences of Feuerbach's mistaken conception.[110]

It should be clear from the above that Marx, according to this reading, must wildly misrepresent Feuerbach's conception of the human essence. As we have seen, Feuerbach's use of this concept described neither something "abstract" nor something "internal". As one scholar of Feuerbach, Larry Johnston, writes:

> it is difficult to conceive where in Feuerbach's work Marx could have acquired the notion that this essence is an 'abstraction inherent in each single individual.'[111]

---

[110] For examples of this kind of reading of the sixth thesis, see Margolis, 'Praxis and Meaning: Marx's Species Being and Aristotle's Political Animal', 333; Mészáros, *Marx's Theory of Alienation*, 85, 149; Hook, 'Marx and Feuerbach'.
[111] Johnston, *Between Transcendence and Nihilism*, 201.

Further, not only does the apparent criticism in the second sentence imply a reading of Feuerbach that is deeply at odds with Feuerbach's account, but the apparent proposed alternative in the third sentence—that the "human essence" is in fact "the ensemble of the social relations"—appears to be a perfectly plausible summary of Feuerbach's contingentist and social conception as described above.[112]

Did Marx, then, simply misread Feuerbach, attributing to him a conception of the human essence that he did not, in fact, hold? And does this suggest that Marx, in the years prior to 1845 in which he was a "Feuerbachian",[113] in fact had a conception of the human essence that was unwittingly at odds with Feuerbach's contingentist and social conception? It would seem that the standard reading of this thesis would commit us to these positions. However, I would like to propose an alternative reading of Marx's sixth thesis, according to which we are not committed to the view that Marx missed Feuerbach's interesting and idiosyncratic contingentist and social conception of the human essence.

This alternative reading is one that takes *all* of the first three sentences as summarising Feuerbach's position. It would then take only from the fourth sentence onwards to contain a criticism of Feuerbach, in the form of highlighting one difference between Feuerbach's conception and that which Marx favours—that Feuerbach "does not enter upon a criticism of this actual essence"—followed by two consequences which follow from this. As such, this reading would take the 6th thesis to be saying something like the following:

> Feuerbach resolves the religious essence into the human essence. But [*in contrast to the everyday conception of the "human essence", for Feuerbach*] the human essence is no abstraction inherent in each single individual.
> In its actuality[, *as Feuerbach saw*], it is the ensemble of the social relations.
> Feuerbach[*'s conception is inadequate, however, insofar as he*], … does not enter upon a criticism of this actual essence [and] is consequently compelled:
> 1. To abstract from the historical process and to fix the religious sentiment as something by itself and to presuppose an abstract – isolated – human individual.
> 2. Essence, therefore, can be comprehended only as "genus", as an internal, dumb generality which naturally unites the many individuals.

---

[112] See again Johnston, 201.
[113] Engels, 'Ludwig Feuerbach and the End of Classical German Philosophy', 364.

I would like to argue for the plausibility of this reading on two fronts. The first relates to Marx's unusual formulation in the third sentence, that the "human essence … [i]n its actuality … *is* the ensemble of social relations". The strangeness of this formulation has been commented on by several Marx scholars. For instance, Norman Geras claims that:

> the 'nature' of the human beings cannot just *be* the ensemble of social relations. Whether by their 'nature' is meant only the characteristics of their human nature or, rather, all those that compose their overall character, there is no obvious sense in which this set of characteristics can be the same thing as an ensemble of social relations.[114]

My first argument in favour of the alternative reading is that this unusual formulation in fact reflects Feuerbach's own, and as such serves as a poor alternative to it. For example, we saw above that for Feuerbach "The essence of man is contained only in the community, in the unity of man with man",[115] while elsewhere he states that "The state is the realized, developed, and explicit totality of the human essence".[116] Further, Marx in his early, "Feuerbachian" writings repeatedly uses such a formulation, suggesting that this sentence represents a continuation of a sentiment held during the height of Feuerbach's influence on Marx. For example, in 1843 Marx writes that "Man is *the world of man*, state, society"; in 1844 he writes that "*Human nature is the true community of men*" and that "the *community* … is the *true* community of man, *human* nature",[117] and elsewhere that "the essence of *man* is the *true community* of man".[118] Indeed, in 1844 Marx explicitly praised Feuerbach for having "founded true materialism and real science by making the social relation of 'man to man' the basic principle of his theory",[119] while in a letter to Feuerbach in the same year Marx wrote:

> In these writings you have provided — I don't know whether intentionally — a philosophical basis for socialism and the Communists have immediately understood them in this way. The unity of man with man, which is based on the real differences between men, the concept of the human species brought down from the heaven of abstraction to the real earth, what is this but the concept of *society!*[120]

---

[114] Geras, *Marx and Human Nature*, 35–36.
[115] Feuerbach, 'Principles of the Philosophy of the Future', 244.
[116] Feuerbach, 'Preliminary Theses on the Reform of Philosophy', 172, translation amended.
[117] Marx, 'Critical Notes on the Article "The King of Prussia and Social Reform. By a Prussian."', 418–19.
[118] Marx, 'Excerpts from James Mill's *Elements of Political Economy*', 265.
[119] Marx, 'Economic and Philosophical Manuscripts', 381.
[120] Engels, 'Ludwig Feuerbach and the End of Classical German Philosophy', 354.

My second argument in favour of this alternative reading relates to the criticism of Feuerbach's position contained in this passage. The two readings contain importantly different criticisms of Feuerbach. In the standard reading, the criticism apparently contained in the second sentence amounts to an accusation that Feuerbach cannot see that the "human essence" is "the ensemble of social relations" (which he obviously cannot therefore criticise), but instead takes it to be an "abstraction inherent in each single individual". On this reading, the numbered criticisms in the second half of the passage are presumably clarifications of the first, primary criticism: this mistaken conception of the human essence leads Feuerbach to abstract from the historical process and to posit "an abstract – isolated – human individual", thereby comprehending the human essence "as an internal, dumb generality which naturally unites the many individuals". On my alternative reading, however, the criticism is more subtle. Starting from the fourth sentence, here the initial criticism is *not* that Feuerbach cannot *see* the "real essence" (i.e. that he cannot see that "the ensemble of the social relations" is the "human essence"), but rather that, while he sees this, he does not *criticise* the ensemble of the social relations. Further, *because* ("consequently") he does not criticise these, he ends up (1) abstracting from the historically contingent ensemble of the social relations to see relatively stable, permanent ("fix[ed]") features of this social "essence" with reference to which he explains the "religious sentiment", and in addition is led to "abstract" (again, read "fix") the nature of human individuals (because seen as existing in relatively stable, fixed social relations). As such, (2) Feuerbach is accused of arriving at what is in all but name a conception of the human essence as abstract and inherent in individuals ("an internal, dumb generality which naturally unites the many individuals"), and therefore his conclusions are in tension with his stated methodology and his explicit conception of the human essence as outlined above.

This second, more subtle criticism of Feuerbach is mirrored in Marx's other criticisms of him in 1845, and as such appears to be the more plausible reading of this passage. For a start, within the *Theses on Feuerbach* Marx criticises Feuerbach by saying that "after the earthly family is discovered to be the secret of the holy family, the former must then itself be destroyed in theory and in practice".[121] Here the first half of the sentence is clearly a summary of Feuerbach's project (Feuerbach explicitly makes this reduction of the holy family to the earthly family[122]), and the

---

[121] Marx, 'Concerning Feuerbach', 422.
[122] Feuerbach, *The Essence of Christianity*, chap. 24.

criticism in the second half of the sentence mirrors that in my reading of the sixth thesis: the "real" basis of the holy family, which Feuerbach sees, must be criticised ("destroyed in theory and in practice"). This criticism returns in *The German Ideology*, written with Friedrich Engels in 1845. Here, Marx and Engels write that Feuerbach "knows no other 'human relations' 'of man to man' than love and friendship, and even then idealised. He gives no criticism of the present conditions of life".[123] This theme arises also in their objection to Feuerbach's description of himself as a "communist":

> Feuerbach's whole deduction with regard to the relation of men to one another is only aimed at proving that men need and *always have needed* each other. He wants to establish consciousness of this fact, that is to say, like the other theorists, he merely wants to produce a correct consciousness about an *existing* fact; whereas for the real Communist it is a question of overthrowing the existing state of things.[124]

Here there are two strands to Marx and Engels' criticism, both of which mirror those described above. First is that Feuerbach's "deduction" describes the social relations between individuals in too general a form, lacking sufficient historical detail: he only proves "that men need and always needed each other" and "knows no other 'human relations' 'of man to man' than love and friendship"; thus "As far as Feuerbach is a materialist he does not deal with history".[125] Second, Feuerbach is said to be content with *describing* this social-relational "essence", instead of subjecting it to criticism. The latter complaint is for Marx and Engels connected to the first, insofar as sight of the historical specificity and emergence of features of social relations encourages the view that these are changeable and thus criticisable.

There is, of course, a potential tension between Marx and Engels' criticism and my account of Feuerbach above, which included an emphasis on the contingency of the human essence. This tension can be reduced if we distinguish between what we might call *conceptual* and *practical* permanence. Whereas for Feuerbach the human essence, because contingent and social, can never be *conceptually* permanent, can never be said to be by definition and necessarily fixed, in his practical dealings with history Feuerbach focuses on those aspects of social relations which are relatively stable, that is, *practically* permanent. In particular, Feuerbach describes the tension

---

[123] Marx and Engels, 'The German Ideology', 41.
[124] Marx and Engels, 57–58.
[125] Marx and Engels, 41.

between individuals and their "essence" *qua* their society as a common cause of religious belief:

> Whoever, therefore, does not put the species in the place of the Divine, leaves a gap in the individual which will be filled again by the idea of God, i.e. the personified essence of the species.[126]

As such, he claims that "so long as we have just two, as man and wife, we still have religion. Two, difference, is the origin of religion".[127] He therefore portrays history as consisting of two stages: an early stage in our evolutionary history in which individuals' primary dependence was on the goods of nature, and a later stage—beginning still very long ago—in which individuals become interdependent, creating the tension between individuals and their social essence that is productive of religion.[128] Marx and Engels' critique can be seen in this light, as a criticism of Feuerbach's lack of historical complexity, and of implicitly assuming that a long history of religious tension between individuals and their social essence implies that this is a practically permanent feature of life, immune to criticism. In contrast, Marx and Engels in 1845 argue that this tension between individuals and their social essence is as contingent and changeable as are the other features of life which Feuerbach historicised:

> Here [in Feuerbach's *Principles*[129]] every exception is expressly conceived as an unhappy chance, as an abnormality which cannot be altered. Thus if millions of proletarians feel by no means contented with their living conditions, if their 'being' does not in the least correspond to their 'essence', then, according to the passage quoted, this is an unavoidable misfortune, which must be borne quietly. These millions of proletarians or communists, however, think quite differently and will prove this in time, when they bring their 'being' into harmony with their 'essence' in a practical way, by means of a revolution.[130]

3.4 Marx's concept of the human *Wesen*

As I have argued above, Marx well understood the contingentist and social aspects of Feuerbach's concept of the human essence, and Marx's own comments regarding the human essence can be seen as heavily influenced by these themes.

---

[126] Feuerbach, '"The Essence Of Christianity" in Relation to "The Ego And Its Own"', 87.
[127] Feuerbach, 87.
[128] See Kamenka, *The Philosophy of Ludwig Feuerbach*, 42–43.
[129] Feuerbach, 'Principles of the Philosophy of the Future', 214–15.
[130] Marx and Engels, 'The German Ideology', 58.

Insofar as the contingentist aspect is concerned, Feuerbach's focus on "sensible essence" in contrast with "absolute essence" is echoed in several of Marx's comments, for instance his claim to be concerned with "the real essence of the finite real, i.e. of what exists and is determined",[131] or his emphasis on understanding "the essence of man … in terms of its *real human existence*" as opposed to "as an imaginary detail".[132] It is, however, the *social* character of Feuerbach's concept that Marx adopts most fully. I will therefore talk primarily of Marx's *social* conception of the human essence.

We have already seen that Marx throughout 1843 and 1844 makes statements such as "the essence of man is the true community of man".[133] Similarly, he writes that:

> If, for example, the analysis of the family, civil society and the state etc. leads us to regard these modes of man's social existence as the realization and objectification of his essence, then the family etc. will appear as qualities inhering in a subject. In that event man will remain the essence of all these realities, but these realities will also appear as man's *actual* universality and, therefore, as common to all men.[134]

That these are more than mere turns of phrase can be seen in Marx's many other comments that make sense only assuming a Feuerbachian social conception of the human essence. For example, in his *notes on James Mill* Marx describes what production "as human beings" might look like, saying that in acting as "the *mediator* between you and the species" I would have "brought about the immediate expression of your life" and shown myself to be "an essential part of yourself" (in the process of which "I would have directly *confirmed* and *actualised* my authentic nature, my *human, communal* nature").[135] The idea that an individual's "essence" describes the social relations of which she is a part makes more sense of this claim than any other I can think of, insofar as my acting as part of a particular kind of social relation would (partly) bring "about the immediate expression of your life", i.e. your essence, i.e. these social relations themselves. Similarly, in his *Critique of Hegel's Doctrine of the State* Marx writes:

> *every* [social] function is representative. For example, a cobbler is my representative in so far as he satisfies a social need, just as every definite form of social activity, because it is a species activity, represents

---

[131] Marx, 'Critique of Hegel's Doctrine of the State', 80.

[132] Marx, 98.

[133] Marx, 'Excerpts from James Mill's *Elements of Political Economy*', 265.

[134] Marx, 'Critique of Hegel's Doctrine of the State', 99, translation amended.

[135] Marx, 'Excerpts from James Mill's *Elements of Political Economy*', 278, translation amended.

> only the species. That is to say, it represents a determination of my own essence just as every man is representative of other men. In this sense he is a representative not by virtue of another thing which he represents but by virtue of what he *is* and *does*.[136]

Again, the most sense that can be made of this passage is an interpretation that assumes a Feuerbachian conception of the human essence: how can the activity of a cobbler represent "a determination of my own essence" *because* their activity "represents only the species" unless this entire "species" is viewed as my "essence"?

That is not to say, however, that Marx's discussion of the human *Wesen* is identical with Feuerbach's. Marx's discussion has a somewhat different focus or emphasis: his primary interest is not on the fact that the human essence is a predicate of the species as a whole, but rather on those features of human beings that makes this the case. For Marx the human species can be described in terms of having a holistic essence *because* of some feature of human beings, and it is this feature that is of central importance to him.[137] This feature is the fact that human beings are *species-beings* (an idea seemingly drawing from Hegel's notion that human beings are essentially *Geist* or Spirit[138]).

Part of what makes humans species-beings for Marx is that they are each what they are—they have the specific qualities and needs that they have—only by virtue of their interrelations with others. This idea is connected to the secondary theme found in Feuerbach's work described above that what individual human beings are is affected by the social relations in which they live. For instance, Marx writes that:

> Only through developed industry, i.e. through the mediation of private property, does the ontological essence of human passion come into being[139]

And that:

> the *senses* of social man are *different* from those of non-social man. Only through the objectively unfolded wealth of human nature can the wealth of subjective *human* sensitivity – a musical ear, an eye for the

---

136 Marx, 'Critique of Hegel's Doctrine of the State', 189-190, translation amended.
137 Cf. Ollman, *Alienation*, 63. There is, of course, a tension here insofar as that feature by virtue of which human beings can be described collectively as having a social essence can itself be described as individuals' "essence" in a more traditional sense. The perfectionist account of Marx picks up on this. I will comment on this tension in chapter 2 section 3.
138 Chitty, 'Hegel and Marx', 478–86.
139 Marx, 'Economic and Philosophical Manuscripts', 375.

beauty of form, in short, *senses* capable of human gratification – be either cultivated or created.[140]

This social cultivation of needs goes beyond simply the development of more sophisticated needs for the goods produced by others, however: it includes the creation of a need for other human beings, by which Marx presumably means the need for certain kinds of social interaction with others. For instance, Marx writes that the extent to which this has occurred can be seen in the form of heterosexual relationships:

> This relationship also demonstrates the extent to which man's needs have become human needs, hence the extent to which the other, as a human being, has become a need for him, the extent to which in his most individual existence he is at the same time a communal being.[141]

Given this social shaping of what individuals are, Marx, along with Feuerbach, sees human beings as self-creating:

> the *whole of what is called world history* [as] … nothing more than the creation of man through human labour, and the development of nature for man, he ["socialist man"] therefore has palpable and incontrovertible proof of his self-mediated *birth,* of his *process of emergence*.[142]

A second aspect of the fact that humans are species-beings is that, partly because of the social needs described above, they can and do *act* in particular ways.

> The *actual, active* relation of man to himself as a species-being, or the *actualisation* of himself as a real species-being, i.e. as a human being, is only possible if he really employs all his *species-powers* – which again is only possible through the cooperation of mankind and as a result of history – and treats them as objects[143]

It is in this sense that, as we saw above, the family, civil society and the state etc. were described as "modes of man's social existence as the realization and objectification of his essence".[144] Of course, Marx does not think that the social institutions of the contemporary world are *adequate* manifestations of individuals' character as species-beings, as we will see in chapter 2. Yet he insists that they are nevertheless partial manifestations of this fact. For instance, in *On The Jewish Question* Marx complains that in contemporary society it is only in the state that

---

[140] Marx, 353.
[141] Marx, 347.
[142] Marx, 357.
[143] Marx, 386, translation amended.
[144] Marx, 'Critique of Hegel's Doctrine of the State', 99.

each human being "behaves … as a species-being, in community with other men", whereas civil society is instead "the sphere of egoism and the *bellum omnium contra omnes*",[145] i.e. a sphere in which individuals' character as species-beings is not adequately manifested.

That human beings "objectify" their character as species-beings through activity in this way itself has a consequence: it enables individuals to come to a kind of self-knowledge, to come to know that they are species-beings. The human being, therefore:

> is a being for himself and hence a *species-being*, as which he must confirm and activate [*betätigen*] himself both in his being [i.e. in his activity] and in his knowing.[146]

For instance, he writes that the kind of production for one another, even in the capitalist form of private property, gives producers insight into this:

> The other man is also the owner of private property, but of *another* thing which I lack but which I neither can nor will dispense with, which I *need* to complete my own existence and to realize my own essence.
>   The bond which unites the two owners is the *specific nature of the object* which constitutes their private property. The longing for these two objects, i.e. the need for them, shows each owner, makes him conscious of the fact, that he stands in another *essential* relation to the objects than that of private property, that he is not the particular being as he imagines, but a *total* being and as a *total* being his needs stand in an *inner* relation to the products of the labour of others – for the felt need for a thing is the most obvious, irrefutable proof that that thing is part of *my* essence, that its being is for me and that its *property* is my property, the particular quality peculiar to my essence.[147]

In this sense:

> The practical creation of an *objective world*, the *fashioning* of inorganic nature, is proof that man is a conscious species-being, i.e. a being which treats the species as its own essential being or itself as a species-being[148]

These passages connect to another theme central to Marx's thought in 1844, namely that individuals can come to know their character as species-beings not only through experiencing the social activity by means of which individuals produce for one another, but also through experience of the physical objects produced thereby. This

---

[145] Marx, 'On the Jewish Question', 220.
[146] Marx, 'Economic and Philosophical Manuscripts', 391, translation amended.
[147] Marx, 'Excerpts from James Mill's *Elements of Political Economy*', 267.
[148] Marx, 'Economic and Philosophical Manuscripts', 328.

is not a theme on which we can reflect here, although it will be mentioned again in chapter 2.

By seeing that their own activity objectifies their own socially-created and other-regarding needs, then, individuals can come to know that they are species-beings: the individual can "look … upon himself as the present, living species".[149] This connects Marx's account to the Hegelian theme of *Geist*'s coming to self-consciousness through its alienation or externalisation—*Entäußerung*—of itself in something other.[150] We encountered this theme above, where I described Feuerbach's ambivalent attitude towards Christianity: while on the one hand he criticised it as "perverse" and antithetical to autonomy and morality, he also saw it as a necessary step in our attainment of self-knowledge. We will see in chapter 2 that much the same ambivalence can be found in Marx's assessment of capitalism.

Finally, that human beings can attain this kind of self-knowledge in turn manifests in new needs and hence, potentially, in their activity. For instance, human beings who are species-beings in the stronger sense that they have attained self-knowledge of their character as species-beings will, Marx thinks, develop a need to produce for others in a particular way, that is, to produce for them motivated not by instrumental reasons regarding what they will get in exchange, but by the mere reason that these others are also species-beings and are a part of their own social essence in the Feuerbachian sense. Thus, where this need manifests in productive activity for others:

> In your use or enjoyment of my product I would have the *immediate* satisfaction and knowledge that in my labour I had gratified a *human* need, i.e. that I had objectified *human nature* and hence had procured an object corresponding to the needs of another *human being*.[151]

> In so far as man, and hence also his feelings, etc., are *human*, the affirmation of the object by another is also his own gratification.[152]

Part of this need seems to derive from a need for something like "recognition" by these others (an idea coming from Hegel[153]):

---

[149] Marx, 327.
[150] Hegel, *Phenomenology of Spirit*, chap. 6; Schacht, *Alienation*, chap. 2.
[151] Marx, 'Excerpts from James Mill's *Elements of Political Economy*', 277.
[152] Marx, 'Economic and Philosophical Manuscripts', 375.
[153] Chitty, 'Recognition and Property in Hegel and the Early Marx'.

I would have acted for you as the *mediator* between you and the species, thus I would be acknowledged [*erkennt*] by you as the complement of your own being, as an essential part of yourself.[154]

3.5 Marx's version of externalisation

Given this emphasis on human beings' essential interconnection and the particular kinds of *activity* that follow from this, we can already see the distinctive way in which Marx takes human beings to externalise their essence in the creation of social institutions. That is, social institutions are "externalisations" of human beings' "essence" in the sense that their social activity, itself the result of their essential character as species-beings, *constitutes* these social institutions. As Marx puts it:

> Since the essence of *man* is the *true community* of man, men, by activating their own essence, produce, create this human community, this social being which is no abstract, universal power standing over against the solitary individual, but is the essence of every individual, his own activity, his own life, his own spirit, his own wealth.[155]

Thus:

> In the practical, real world, self-alienation [*Selbstentfremdung*] can manifest itself only in the practical, real relationship to other men. The medium through which alienation [*Entfremdung*] progresses is itself a practical one.[156]

Marx therefore describes social institutions as each "a particular alienation [*Entfremdung*] of man and each is centred upon one particular area of alienated [*entfremdeten*] essential activity",[157] i.e. as constituted by a particular area of human beings' social activity.

It is because social institutions are in this sense constituted by the activity of agents *qua* species-beings that Marx describes social institutions as their "species-life" and "species-activity":

> The process of *exchange* both of human activities in the course of production and of *human products* is equal to the species-activity and the species-spirit whose real, conscious and authentic existence consists in *social* activity and *social* enjoyment…. Therefore, this *true community* does not come into being as the product of reflection but it

---

[154] Marx, 'Excerpts from James Mill's *Elements of Political Economy*', 277.
[155] Marx, 265.
[156] Marx, 'Economic and Philosophical Manuscripts', 331.
[157] Marx, 362.

arises out of the *need* and the *egoism* of individuals, i.e. it arises directly from their own activity.[158]

Here the mode of externalisation or alienation of the human essence that Marx describes in the creation of social institutions is explicitly contrasted with the kind of externalisation that Feuerbach describes in the alienation that creates religion: whereas the latter is "the product of reflection", the former "arises directly from their own activity". Similarly, in his *Critique of Hegel's Doctrine of the State* Marx writes that "the state cannot be regarded as a simple reality, it must be viewed as an activity, as a differentiated activity".[159]

---

[158] Marx, 'Excerpts from James Mill's *Elements of Political Economy*', 265.
[159] Marx, 'Critique of Hegel's Doctrine of the State', 71.

Chapter 2
The critical dimension of Marx's concept of alienation

We saw in chapter 1 that for Marx in 1843-4 social institutions are created through human beings' alienation or externalisation of their essence as species-beings, in the sense that the activity enabled and required by the fact that human beings are species-beings constitutes these social institutions; social institutions are the externalisations of human beings' essence as species-beings by way of their species-activity. This, then, captures the first stage of the process of alienation described above, that most frequently referred to with the term *Entäußerung*. Nevertheless, as we have seen, the concept of alienation that Marx inherits from Feuerbach (and Fichte, Hegel, and the other Young Hegelians) implies more than just a claim about the creation or ontology of some entity, but also a claim about its becoming in some sense alien to or independent of those who created it, and—in Feuerbach and the other Young Hegelians' application of this idea—that this becoming-alien is *criticisable*. These themes are central to Marx's account too.

We must, then, examine what this becoming-alien of social institutions consists in, and the sense in which this might be ground for criticism. Before addressing these questions, however, it is worth addressing an important theme in the secondary literature on Marx, which emphasises the *perfectionist* strand in his thought as the basis for his criticism of alienated social institutions, and in particular of those characteristic of capitalism.

1 Marx's perfectionist criticism of capitalism

In Chapter 1 section 3.4 I sketched Marx's conception of humans as species-beings, with particular emphasis on its implications for understanding what it means for human beings to externalise or *entäußert* their essence in social institutions. While this involved focusing on the descriptive aspects of this idea, many passages also clearly had normative significance. We saw this first in the claim that humans' character as species-beings gives rise to certain *needs*, for this description clearly

has normative significance: needs are things that those who have them are better off having satisfied, and which to that extent *ought* to be satisfied. We saw it also in Marx's description of the social relationships that humans as species-beings came to need: the extent to which an agent has developed certain kinds of needs for others, we were told, "demonstrates the extent to which man's needs have become human needs". Further, this claim is preceded by the claim that:

> It is possible to judge from this relationship the entire level of development of mankind. It follows from the character of this relationship how far *man* as a *species-being*, as *man*, has become himself and grasped himself[1]

Not only does this passage invoke the normative notion of needs, but seems to claim that the acquisition of certain needs is itself a positive development. This was reflected too in Marx's complaint that in the social institutions of the contemporary world agents' character as species-beings was manifested incompletely, only in certain aspects of their life (those constitutive of the political sphere) and not in others (those constitutive of civil society). Here, humans' manifestation of their character as species-beings is seen not only as a descriptive fact, but as something that they *ought* to do. This theme carries over particularly clearly into Marx's critique of capitalist social relations.

We have already seen that for Marx the social relations in which human beings produce for one another are a product of their being species-beings in a descriptive sense: they have complex needs that are satisfied only by the work of others, and this is manifested in social relations of production and exchange. Thus, in the social relations of capitalism:

> the *human heart* can be heard throbbing behind the façade of property,
> in all of them we witness man's dependence upon man. Whatever the
> nature of this dependence it is *human*[2]

However, despite this these social relations are said to be ill-suited for species-beings in several ways. First, capitalist social relations prevent many individuals from developing the kind of complex, developed senses made possible by living with one's fellow species-beings:

> The man who is burdened with worries and needs has no *sense* for the
> finest of plays; the dealer in minerals sees only the commercial value,

---

[1] Marx, 'Economic and Philosophical Manuscripts', 347.
[2] Marx, 'Critique of Hegel's Doctrine of the State', 170.

and not the beauty and peculiar nature of the minerals; he lacks a mineralogical sense; thus the objectification of the human essence, in a theoretical as well as a practical respect, is necessary both in order to make man's *senses human* and to create an appropriate *human sense* for the whole of the wealth of humanity and of nature.[3]

Second, they prevent agents from acquiring the kind of self-knowledge necessary for them to act as species-beings in the stronger sense. Instead of seeing the social relations surrounding them as manifestations of their own and others' shared social essence, individuals see them simply as facts of the world with which they must each—as individuals—contend. As such, they view themselves not as species-beings but only as individuals:

> In the same way as alienated [*entfremdete*] labour reduces spontaneous and free activity to a means, it makes man's species-life a means of his physical existence.
> Consciousness, which man has from his species, is transformed through alienation [*Entfremdung*] so that species-life becomes a means for him.
> (3) alienated [*entfremdete*] labour therefore turns *man's species-being* – both nature and his intellectual species-powers – into a being *alien* to him and a *means* of his *individual existence*. It alienates [*entfremdet*] man from his own body, from nature as it exists outside him, from his spiritual essence, his *human* essence.[4]

As mentioned previously, it seems that for Marx part of the process of attaining self-knowledge comes about through the worker's direct relationship with the physical object they have produced, seeing reflected in it that it was created for others as part of the producer's own social essence. Part, then, of his explanation of why workers in capitalism are prevented from attaining such self-knowledge is that this physical product is taken from them:

> In tearing away the object of his production from man, estranged [*entfremdete*] labour therefore tears away from him his *species-life*, his true species-objectivity, and transforms his advantage over animals into the disadvantage that his inorganic body, nature, is taken from him.[5]

Finally, even were this self-knowledge to somehow develop, our need to produce *for others* in a particular way, as fellow species-beings, and to be recognised by them through this production, is limited in capitalist production and exchange:

---

[3] Marx, 'Economic and Philosophical Manuscripts', 353–54.
[4] Marx, 329–30.
[5] Marx, 329.

> Hence our exchange cannot be the mediating movement which confirms that my product is for you because it is an *objectification* of your own nature, of your need. For our products are not united for each other by the bond of *human nature*. Exchange can only set in *motion*, it can do no more than confirm the *character* each of us bears in relation to his own product and hence to the product of the other. Each of us sees in his product only his *own* objectified self-interest, hence in the product of others the objectification of a *different*, alien self-interest, independent of oneself.[6]

Marx, then, has many reasons for disliking capitalism, reasons rooted in his conception of humans as species-beings. And this conception underlies much of his picture of what a future communist society might look like: in his portrayals of communist society, the various ways in which contemporary social institutions are ill-suited to humans' essence as species-beings and their potential knowledge of themselves as species-beings are said to be overcome.[7]

This normative aspect of Marx's portrayal of humans as essentially species-beings, and the idea that he evaluated social institutions according to the extent to which they allow for the development and satisfaction—the perfection—of individuals' essential needs in this sense, is what I call the *perfectionist* strand in his writings. This perfectionist strand forms the central plank in most secondary accounts of Marx's early criticisms of capitalism.[8]

However, it seems that this perfectionist strand cannot account for the full strength of Marx's criticism of capitalist social relations or the modern state. Consider, for instance, the last lines of this characteristic passage describing market exchange:

> Men, not as abstractions, but as real, living, particular individuals *are* this community. *As* they are, so it is too. To say therefore that *man* is alienated [*entfremdet*] from himself is identical with the statement that the *society* of this alienated man is the caricature of a *true community*, of his true species existence, that therefore his activity is a torment to him, his own creation confronts him as an alien power, his wealth appears as poverty, the essential bond joining him to other men appears inessential, in fact separation from other men appears to be his true existence, his life appears as the sacrifice of his life, the realization of his essence appears as the de-realization of his life, his production is the production of nothing, his power over objects appears as the power of

---

[6] Marx, 'Excerpts from James Mill's *Elements of Political Economy*', 275.
[7] See in particular Marx, 277–78.
[8] See, for instance, Leopold, *The Young Karl Marx*, chap. 4; Wood, *Karl Marx*, chap. 2; Geras, *Marx and Human Nature*, 70.

objects over him; in short, he, the lord of his creation, appears as the servant of that creation.[9]

Here we can see the perfectionist criticism that the social institutions of market exchange do not manifest activity in accordance with humans' essence as species-beings such that individuals do not attain self-knowledge and "the essential bond joining him to other men appears inessential". However, this perfectionism lies alongside a quite different set of sentiments: that "his own creation confronts him as an alien power" and "his power over objects appears as the power of objects over him" such that he "appears as the servant of that creation".

2 The becoming-alien of social institutions

This different set of sentiments can be found throughout Marx's work, and seem to constitute a significantly stronger kind of criticism than that the social institutions of the modern world only partially manifest or satisfy our essence as species-beings. That is, Marx frequently excoriates private property and capitalism not only for not being good enough, for not adequately reflecting and enabling agents to know their social essence, but for being actively harmful and, in particular, for restricting human beings' *freedom*. He describes the social institutions of capitalism as "the relation of *dominance* and *slavery*"[10] and "*forced* labour",[11] which people conform to only "because they *dominate* him",[12] and from which they need to be "emancipated".[13] It seems impossible to justify these characterisations on the basis that capitalism only moderately and incompletely satisfies or manifests humans' essence as species-beings.

These claims are often connected to the suggestion that agents are in some sense separated, "estranged" (to use a common translation of *Entfremdung*), or cut-off from something of themselves. In particular, Marx repeatedly talks of agents being separated from themselves or from their own essence.[14] Similarly, Marx often talks

---

[9] Marx, 'Excerpts from James Mill's *Elements of Political Economy*', 265–66, translation amended.
[10] Marx, 277.
[11] Marx, 278; Marx, 'Economic and Philosophical Manuscripts', 326.
[12] Marx, 'On the Jewish Question', 239.
[13] Marx, 'Economic and Philosophical Manuscripts', 333.
[14] For alienation from oneself, see Marx, 'On the Jewish Question', 240; Marx, 'Economic and Philosophical Manuscripts', 326. For alienation from one's essence or species-essence,

of agents becoming separated from social institutions themselves. For instance, he writes that

> the *mediating function* or movement, human, social activity, by means of which the products of man mutually complement each other, is *alienated* [*entfremdet*] and becomes the property of a *material thing* external to man, viz. money. If a man himself alienates [*entäußert*] this mediating function he remains active only as a lost, dehumanized creature. The relation between things, human dealings with them, become the operations of a being beyond and above man. Through this *alien* [*fremden*] *mediator* man gazes at his will, his activity, his relation to others as at a power independent of them and of himself - instead of man himself being the mediator for man. His slavery reaches a climax…. Hence this *mediator* is the lost, alienated [*entfremdete*] *essence* of private property, private property *alienated* [*entäußerte*] and external to itself; it is the *alienated* [*entäußerte*] *mediation* of human production with human production, the *alienated* [*entäußerte*] species-activity of man.[15]

Thus, not only is the human essence said to be alienated in the production of social institutions, but these social institutions *themselves* are said to be alienated.

As we saw in chapter 1, Marx—following Feuerbach—employs a rather curious notion of "essence" according to which the existing social relations themselves are said to be the human "essence". Thus, the two claims—that the human essence becomes separated from or alien to agents, and that social institutions do so—are synonymous. It seems, indeed, that the latter formulation can make sense of what Marx means by the former, which by itself, operating with a more traditional, internal sense of essence, seems to make little sense: what could it possibly mean for my essence to become independent of and alien to me, on a more traditional understanding of essence?

It seems that Marx takes the idea that in capitalism the social institutions in which agents live are separated from them, have become alien or independent, to in some sense explain the association between alienation and unfreedom, introduced above. For instance, he writes that in capitalism a (male) worker:

---

see Marx, 276, 329–30; Marx, 'Excerpts from James Mill's *Elements of Political Economy*', 269.

[15] Marx, 'Excerpts from James Mill's *Elements of Political Economy*', 260–61, translation amended.

can act practically and practically produce objects only by making his products and his activity subordinate to an alien substance and giving them the significance of an alien substance – money.[16]

Workers' activity is therefore "determined by social configurations alien to the worker"[17] and by:

social needs alien to him and which act upon him with compulsive force. He must submit to this force from egoistic need, from necessity; for him the needs of society mean only the satisfaction of his personal wants while for society he is only the slave that satisfies its needs[18]

Of course, as we have seen, these social institutions are not only said to *be* agents' social "essence", but are also constituted by their *activity*, and this perhaps makes the relation between the becoming-alien of social institutions and a conception of freedom clearer. Thus, Marx describes "the act of alienation [*Entfremdung*] of practical human activity" as:

the relationship of the worker to his own activity as something which is alien and does not belong to him, activity as passivity, power as impotence, procreation as emasculation, the worker's *own* physical and mental energy, his personal life, – for what is life but activity? – as an activity directed against himself, which is independent of him and does not belong to him.[19]

This idea that the process of alienation involves agents becoming separated from their own social institutional essence fits well with the process of alienation described by Marx's predecessors described in chapter 1. As we saw there, for Fichte alienation described a process whereby a moral agent not only externalises something—their conception of the moral law—but comes to see this as something independent of and separated from themselves, i.e. as God. Similarly, Feuerbach presented the creation of God not simply as the projection of the characteristics of the human species, but as a process in which human beings mistakenly saw this God as something distinct and separate from them, as something alien. Nevertheless, as we have seen, the processes of externalisation involved in the creation of God and of social institutions are importantly different, and to understand Marx's claim concerning the separation of these from agents requires further discussion. If social institutions are constituted by our activity, what could it mean to say that they are at the same time separated from or alien to us?

---

[16] Marx, 'On the Jewish Question', 241.
[17] Marx, 'Economic and Philosophical Manuscripts', 268.
[18] Marx, 269.
[19] Marx, 327, translation amended.

One possibility is that Marx's claim about the alienness of social institutions *vis-à-vis* individuals is the same as Feuerbach's claim about the alienness of human beings' essence when externalised in God. That is, we might think that the phenomenon highlighted by Marx is that the individuals whose activity constitute social institutions *mistakenly believe* that these are independent of this constituting activity. This interpretation might be read into the passage quoted above in which "the *mediating function* or movement, human, social activity … is *alienated*": there, Marx describes "man" as "gaz[ing] at his will, his activity, his relation to others as at a power independent of them" such that this relation becomes "a *veritable God*" and a "cult". This might suggest a broadly Feuerbachian account of the sense in which alienated social institutions are alien: individuals mistakenly believe these to be independent of them.

However, this does not adequately explain many of Marx's comments on the matter. We can see this, first, in two passages that discuss both the alienation that results in religion and in social institutions:

> Just as in religion the spontaneous activity of the human imagination, the human brain and the human heart detaches itself from the individual and reappears as the alien activity of a god or of a devil, so the activity of the worker is not his own spontaneous activity. It belongs to another, it is a loss of his self.[20]

> But it is clear from an analysis of this concept that, although private property appears as the basis and cause of alienated labour, it is in fact its consequence, just as the gods were *originally* not the cause but the effect of the confusion in men's minds. Later, however, this relationship becomes reciprocal.[21]

In these passages, it first appears that the two processes are being equated (and we have already seen their similarities). However, on closer inspection both also reveal the difference Marx sees between the two processes. In the first passage, what is equivalent to the mistaken belief that results in religion is the fact that the worker's activity "is not his own" but rather "belongs to another". It is thus not a mistaken belief about capitalist social relations' constitution, but rather the fact that the worker's activity is *in fact* not "his own". This difference is even clearer in the second passage. In the alienation that results in religion, the alienness of the human essence was "the effect of the confusion in men's minds". But in the alienation that results in

---

[20] Marx, 326–27.
[21] Marx, 332.

private property, this alienness is a result not of any confusion but of "alienated labour", i.e. of a particular kind of *activity*. It is a mistaken way of acting, rather than a mistaken way of thinking.

This idea comes out also in Marx's comments on the work of the political economists such as Smith and Ricardo. For Marx, these authors' mistake is not that they, along with the agents involved, mistakenly believe the social institutions of capitalism to be independent. Rather, their description of them in terms that suggest this independence have "merely formulated the laws of estranged labour"[22]: they correctly describe the *actual* independence of these institutions *vis-à-vis* workers.

Thus, for Marx the alienness of social institutions does not consist in agents' false beliefs that these are alien or independent of their activity. Rather, it consists in their being—in some sense that we will need to explain—*actually* so alien and independent. As he puts it:

> Alienation [*Entfremdung*] appears not only in the fact that the means of *my* life belong to *another* and that *my* desire is the inaccessible possession of *another*, but also in the fact that all things are *other* than themselves, that my activity is *other* than itself, and that finally – and this goes for the capitalists too – an *inhuman* power rules over everything.[23]

As this passage perhaps makes clear, the idea that social institutions are *actually* alien or independent of individuals more adequately explains the connection between alienation and unfreedom: it is because "an *inhuman* power rules over everything" that agents are unfree, a connection that would be difficult to explain if the alienness of social institutions consisted merely in false beliefs.

Finally, as we have seen, the description of social institutions as alien seems to point to a particular *kind* of social activity or social relation, i.e. those in which agents' activity "belongs to another" and so on. As such, it seems, the alienness of social institutions can be alternatively described as the fact that they have a particular *mode of existence*, that they are constituted by social activity of a particular kind. This emerges particularly clearly in the following passage (some of which we considered above):

> as long as man does not recognize [*erkenne*] himself as man, and hence give the world a human organization, this *community* appears in the guise of

---

[22] Marx, 332.
[23] Marx, 366.

alienation [*Entfremdung*]. For its *subject*, man, is a being alienated [*entfremdetes*] from himself. Men, not as abstractions, but as real, living, particular individuals *are* this community. As they are, so it is too. To say therefore that *man* is alienated [*entfremdet*] from himself is identical with the statement that the *society* of this estranged [*entfremdeten*] man is the caricature of a *true community*, of his true species-existence[24]

Here, the alienness of social institutions is reframed as not having a "human organization" and being "the caricature of a *true community*". This idea is expressed elsewhere as that

> The *division of labour* is the economic expression of the *social nature of labour* within alienation [*Entfremdung*]. Or rather, since *labour* is only an expression of human activity within alienation [*Entäußerung*], an expression of life as alienation of life [*Lebensentäußerung*], the *division of labour* is nothing more than the *alienated* [*entfremdete, entäußerte*] positing of human activity as a *real species-activity* or as *activity of man as a species-being*.[25]

Here, the implication seems to be that within the division of labour individuals act "within alienation" in the sense that they act within social relations of a particular kind: the division of labour is the alienated *form* of the externalisation of agents' essence as species-beings, i.e. an alienated form of social institution. Similarly, elsewhere Marx writes that the capitalist division of labour is "*human activity as species activity in this its alienated* [*entfremdete und entäußerte*] *form*".[26]

Nevertheless, it is important to note that in both quotations above mistaken beliefs still seem to be involved in the alienation of social institutions. In the first, the kind of activity that constitutes alienated social institutions is said to result from "man's" failure to "recognize himself as man", while, in the second, mistaken beliefs are invoked in the formulation that the division of labour is an "alienated *positing*". Yet elsewhere there seems to be the suggestion that mistaken beliefs *result* from the alienation of social institutions: as we saw above, one negative consequence of the capitalist mode of production was said to be that it did not allow agents to attain self-knowledge of themselves as species-beings. In this sense, it seems that to fully understand the more frequently-emphasised perfectionist strand in Marx's work, and in order to *explain* the phenomena criticised by means of this perfectionist framework—in particular, to understand how it is that the contemporary world is

---

[24] Marx, 'Excerpts from James Mill's *Elements of Political Economy*', 265–66.
[25] Marx, 'Economic and Philosophical Manuscripts', 369.
[26] Marx, 369.

such that agents cannot know that they are and thus act as species-beings—we must first understand the idea that these social institutions can be alienated and thus become alien to them, including understanding the mistaken beliefs that somehow result in this alienation.

There seems, then, to be in Marx's account a complex two-way relationship between the alienation of social institutions and agents' having of mistaken beliefs. I will not attempt to reconstruct this complex relationship. As we will see, mistaken beliefs play a central role in the account of the alienation of social institutions that I will present. However, the key point here is that while such mistaken beliefs might be crucial in this account, the alienation of social institutions does not *consist in* mistaken beliefs, as does the alienation that results in God in Feuerbach's account.

In chapter 1, I described the concept of alienation in Marx's predecessors as describing a *process*, in which agents externalise some aspect of themselves which then becomes independent of them in such a way that it affects their activity. I have argued that in Marx's discussion of the externalisation that results in social institutions and of the becoming-alien of these, we see a parallel process being described, albeit one with important differences related to the kind of externalisation involved and the sense in which these become alien or independent. This process is what I refer to as *the alienation of social institutions*: it is the process in which a social institution is created (through externalisation of agents' "essence" through their social activity) and itself becomes alien or independent of these agents. This description, of course, remains rather abstract, and leaves many questions open: it is the goal of subsequent chapters to answer these questions in order to arrive at a more satisfactory account of this process. For now, let us formalise the account I have given so far. The concept of the alienation of social institutions describes a process in which:

(1) Group of agents X externalise their social essence through their social activity, thereby constituting social institution Y

(2) Social institution Y becomes independent of group of agents X

(3) That (2) is the case enables social institution Y to influence group of agents X's behaviour in a way in which it otherwise would not. This constitutes a restriction of their freedom.

3 The German Ideology

While, as I argued in chapter 1, the sixth of Marx's 1845 *Theses on Feuerbach* does not mark the exact break with Feuerbach that many have taken it to be, it nevertheless contains a significant criticism of Feuerbach. This was that while Feuerbach's explicit concept of the human essence was *not* of an "abstraction inherent in each single individual", instead describing the characteristics of actual human beings including the social relations in which they live, his lack of criticism of these social relations and his insufficiently historical account of these meant that he nevertheless thought of humans as relatively stable, as having practically permanent characteristics, and thus as amounting in all but name to the kind of notion of essence which his official account eschewed. From the account I have given above of Marx's portrayal of humans as species-beings, we might wonder if a similar criticism does not also apply to Marx's pre-1845 work: while Marx's explicit account of essence follows Feuerbach in being both a contingentist and a social account, he nevertheless draws from this conception a series of conclusions that end up looking rather like a more traditional conception of essence: human beings, as species-beings, seem to *essentially* share an (historically variable) contingent and social essence with others, and this sharing has consequences for how they do and ought to behave. It is perhaps this tension, along with the influence of Max Stirner's vociferous attack on Feuerbach as a traditional essentialist in his 1844 *The Ego and It's Own*[27] (an attack which, given the account I have given above, seems misleading, and one which Feuerbach objected to strongly[28]), that lead Marx to abandon the terminology of the human "essence" in *The German Ideology*.

Alongside this abandonment of the terminology of the human essence and species-being, this work also largely eschews the terminology of *Entäußerung* and *Entfremdung*. For example, *Entäußerung* is completely absent, while *Entfremdung* appears almost exclusively in quotes from or in direct reference to Stirner,[29] whom this work was devoted to criticising, along with Bauer and Feuerbach. However, the *concept* of the alienation of social institutions that I have been explicating throughout these chapters persists in *The German Ideology*. In particular, three central claims described above can be found: that i) social institutions are constituted

---

[27] Stirner, *The Ego and Its Own*, 46 and *passim*.
[28] Feuerbach, '"The Essence Of Christianity" in Relation to "The Ego And Its Own"'.
[29] Marx and Engels, 'The German Ideology', 231, 245–46, 281–82, 350. p.230 is an exception in which 'entfremdet' is used in Marx and Engels' own voice.

by our social activity; ii) in the contemporary world these social institutions have become independent or alien, in such a way that they restrict agents' freedom; and iii) the sense in which these have become independent or alien is importantly different to that in which God or religion does.

3.1 The constitution of social institutions

Marx and Engels' account of social relations in *The German Ideology* is even more explicit than is Marx's early work that these are constituted by the *activity* of agents. For instance, they write that:

> The social structure and the state are continually evolving out of the life-process of definite individuals, however, of these individuals, not as they may appear in their own or other people's imagination, but as they *actually* are, i.e., as they act, produce materially, and hence as they work under definite material limits, presuppositions and conditions independent of their will.[30]

This is expressed alternatively as social institutions being "produced by [the] … self-activity" of individuals,[31] while elsewhere Marx and Engels, translating Stirner's central question into what they see as the correct question, say (in a passage clearly expressing all three stages of the process of the alienation of social institutions described above):

> How is it that in this process of private interests acquiring independent [*verselbständigen*] existence as class interests the personal behaviour [*Verhalten*] of the individual is bound to be objectified [*versachlichen*], alienated [*entfremden*], and at the same time exists as a power independent [*Verselbständigung*] of him and without him, created by intercourse, and is transformed into social relations, into a series of powers which determine and subordinate the individual, and in their independence assume the form of general interests?[32]

This passage articulates clearly an aspect that is central to Marx and Engels' claim about the constitution of social institutions, namely that the activity which constitutes them is *social* activity or "intercourse". Indeed, a better translation for the term *Verhalten* is "relating": it is the personal *relating* of individuals that is alienated. This idea arises frequently in the *German Ideology*, for instance:

---

[30] Marx and Engels, 35–36.
[31] Marx and Engels, 82.
[32] Marx and Engels, 245, translation amended.

> *All-round* dependence, this primary natural form of the *world-historical* co-operation of individuals, will be transformed by this communist revolution into the control and conscious mastery of these powers, which, born of the action of men on one another, have till now overawed and ruled men as powers completely alien [*fremde*] to them.[33]

Thus, the first stage of the process of the alienation of social institutions—that social institutions are constituted by agents' social activity—is clearly repeated in *The German Ideology*, albeit excised of the claim that this activity constitutes the alienation or externalisation of agents' "essence".

3.2 Becoming-alien and the critique of Bauer and Stirner

Similarly, *The German Ideology* makes frequent reference to social institutions becoming independent or alien *vis-à-vis* those whose activity constitutes them. (As we will see in some of the quotations below, it appears that Marx and Engels adopt a new term—*Verselbständigung*; "becoming self-standing" or "autonomisation"—to refer to this becoming-independent.) See, for instance, the following passages:

> the division of labour offers us the first example of the fact that, as long as man remains in naturally evolved society, … man's own deed becomes an alien [*fremde*] power opposed to him, which enslaves him instead of being controlled by him.…
>
> This fixation of social activity, this consolidation of what we ourselves produce into a thinglike [*sachlichen*] power above us, growing out of our control, thwarting our expectations, bringing to naught our calculations, is one of the chief factors in historical development up till now.* The social power, i.e., the multiplied productive force, which arises through the co-operation of different individuals as it is caused by the division of labour, appears to these individuals, … not as their own united power, but as an alien [*fremde*] force existing outside them, of the origin and goal of which they are ignorant, which they thus are no longer able to control, which on the contrary passes through a peculiar series of phases and stages independent of the will and the action of man, nay even being the prime governor of these.[34]

Here, all aspects of the process of the alienation of social institutions identified above can be found: agents constituting social relations through their social activity ("man's own deed"; "what we ourselves produce"; "which arises through the co-operation of different individuals"); this becoming something alien and independent of those on whom its existence depends ("an alien power opposed to him"; "into a

---

[33] Marx and Engels, 51.

[34] Marx and Engels, 47–48, translation amended. See also Marx and Engels, 51, 86, 396.

material power above us"; "appears to these individuals,… as an alien [*fremde*] force existing outside them"); and this newly-acquired independence giving it an increased ability to influence those agents' behaviour, restricting their freedom ("enslaves him instead of being controlled by him"; "independent of the will and the action of man, nay even being the prime governor of these"). Indeed, as if the parallel were not obvious enough, Marx even inserted a passage in the margin at the point of the * in the quoted passage which begins: "This 'alienation' ['*Entfremdung*'] (to use a term which will be comprehensible to the philosophers)".

The German Ideology is primarily framed as a critique of Feuerbach, Bauer, and Stirner, the latter two of whom, we saw in chapter 1, themselves attempted to apply the Feuerbachian concept of alienation to social institutions. We saw above that Marx agreed with these writers that, at a broad level of description, a similar process to that described by Feuerbach in terms of the alienation that results in God can be seen in the creation of social institutions. However, in the *German Ideology* Marx and Engels criticise these authors at length for overlooking the crucially different ontological statuses of God and of social institutions, i.e. that unlike God, alienated social institutions actually exist. Marx and Engels thus criticise these writers for transferring Feuerbach's concept of alienation too completely to this new subject-matter, overlooking the crucial differences between them. In this sense, Marx and Engels' criticism can be seen to mirror Marx's criticism of Hegel outlined in chapter 1, section 3.1.

Marx and Engels' criticism of Bauer and Stirner is directed particularly strongly at their characterisation of the second stage of the alienation of social institutions, the becoming-independent or *Verselbständigung* of these social institutions:

> The entire body of German philosophical criticism from Strauss to Stirner is confined to criticism of *religious* conceptions. The critics started from real religion and theology proper. What religious consciousness and religious conception are was subsequently defined in various ways. The advance consisted in including the allegedly dominant metaphysical, political, juridical, moral and other conceptions under the category of religious or theological conceptions; and similarly in declaring that political, juridical, moral consciousness was religious or theological consciousness, and that the political, juridical, moral man … was religious. The dominance of religion was presupposed. Gradually every dominant relationship was declared to be a religious relationship

and transformed into a cult, a cult of law, a cult of the state, etc. It was throughout merely a question of dogmas and belief in dogmas.[35]

Because the holy is something alien [*Fremdes*], everything alien is transformed into the holy; and because everything holy is a bond, a fetter, all bonds and all fetters are transformed into the holy. By this means Saint Sancho [Stirner] has already achieved the result that everything alien becomes for him a mere *appearance*, a mere *idea*…. Saint Sancho criticises all actual conditions by declaring them 'the holy'[36]

This insufficiently careful transference of Feuerbach's account of alienation was seen by Marx and Engels to have significant political implications, leading to a form of political quietism that focused exclusively on ideological issues at the expense of the kind of political activism that dealt with actual social and political institutions:

Since the Young Hegelians consider conceptions, thoughts, ideas, in fact all the products of consciousness, to which they attribute an independent [*verselbständigten*] existence, as the real chains of men (just as the Old Hegelians declare them the true bonds of human society), it is evident that the Young Hegelians have to fight only against these illusions of consciousness. Since, according to their fantasy, the relations of men, all their doings, their fetters and their limitations are products of their consciousness, the Young Hegelians logically put to men the moral postulate of exchanging their present consciousness for human, critical or egoistic consciousness, and thus of removing their limitations. This demand to change consciousness amounts to a demand to interpret the existing world in a different way, i.e., to recognise it by means of a different interpretation. The Young-Hegelian ideologists, in spite of their allegedly "world-shattering" phrases, are the staunchest conservatives.[37]

More importantly for our investigation, however, is that Marx and Engels believed that these Young Hegelians' account prevented them from adequately theorising the process that results in this alienation:

Thus [instead] of the task of describing [actual] individuals in their [actual] alienation [*Entfremdung*] and in the empirical relations of this alienation [*Entfremdung*], [purely empirical] relations, the same happens here — the setting forth is replaced by the [mere idea] of alienation, of [the Alien], of the Holy.[38]

This idea comes out strongly, too, in *The Holy Family*, written by Marx and Engels in the previous year:

---

[35] Marx and Engels, 'The German Ideology', 24.
[36] Marx and Engels, 282.
[37] Marx and Engels, 30.
[38] Marx and Engels, 281–82.

> According to Critical Criticism [i.e. to Bauer], the whole evil lies only in the workers' "*thinking*".... But these *mass-minded*, communist workers, employed, for instance, in the Manchester or Lyons workshops, do not believe that by "*pure thinking*" they will be able to argue away their industrial masters and their own practical debasement. They are most painfully aware of the *difference* between *being* and *thinking*, between *consciousness* and *life*. They know that property, capital, money, wage-labour and the like are no ideal figments of the brain but very practical, very objective products of their self-estrangement and that therefore they must be abolished in a practical, objective way for man to become man not only in *thinking*, in *consciousness*, but in mass *being*, in life. Critical Criticism, on the contrary, teaches them that they cease in reality to be wage-workers if in thinking they abolish the thought of wage-labour; if in thinking they cease to regard themselves as wage-workers and, in accordance with that extravagant notion, no longer let themselves be paid for their person. As absolute idealists, as ethereal beings, they will then naturally be able to live on the ether of pure thought. Critical Criticism teaches them that they abolish real capital by overcoming in *thinking* the category Capital[39]

More will be said about these criticisms in chapter 4. For now, we must ask: if for Marx and Engels the becoming-independent of social institutions is not simply a matter of agents mistakenly *thinking* or *believing* them to be independent of them, what does it entail? I do not think that Marx (or Engels) adequately explicates this crucial idea in these early works. It is therefore the task of subsequent chapters to develop a framework with which we can give a more satisfactory account of it. Nevertheless, it would be useful to provide a rough account of what Marx and Engels here have in mind, which can be refined subsequently.

We saw in section 2 that in Marx's 1844 writings there was a complex relationship between mistaken beliefs and the kind of activity that constitutes alienated social institutions: mistaken beliefs were said to be both causally involved in this activity, and to result from it. However, what was key was that the becoming-alien of a social institution did not *consist in* mistaken beliefs, but in activity. This theme, we have seen, is reiterated in the *German Ideology* in Marx and Engels' criticism of Bauer and Stirner. Further, in this work the theme that alienated social institutions *produce* mistaken beliefs is also central.[40] The idea that mistaken beliefs are in some sense causally involved in the activity that constitutes alienated social institutions, however, does not appear. Nevertheless, I take this to be a consequence of the focus of Marx and Engels' argument in this work rather than an explicit change in outlook.

---

[39] Marx and Engels, 'The Holy Family', 52–53.
[40] Marx and Engels, 'The German Ideology', 36, 363.

All activity necessarily involves beliefs of one kind or another in its motivation, and Marx and Engels need not deny this for their criticism of these Young Hegelians to work. As characterised above, Marx's position seems to be that social institutions' becoming-independent consists not in agents' mistaken beliefs, but in mistaken activity, and I see no plausible way of describing activity as mistaken without suggesting that there were mistaken beliefs involved in its motivation. This suggestion is borne out in another passage quoted above, in which Marx and Engels suggest that the independence of social institutions can be overcome by transforming our relationship to them into one of "the *conscious* mastery of these powers": that is, by the powers generated through social activity ("born of the action of men on one another") coming to be a product not of activity motivated by mistaken or, here, "unconscious" beliefs, but by conscious and (presumably) correct beliefs.

On this basis, we might suggest as our rough account of the becoming-independent of social institutions to be refined subsequently that a social institution is independent or alien when constituted by activity in which agents mistakenly *treat* or *act as if* it is independent, i.e. they so act as a result of (some kind of) mistaken beliefs. This account captures both the concept's Feuerbachian inheritance—including the implicit idea that alienation involves some kind of mistake—and Marx and Engels' distinctive emphasis on activity.

4 Conclusion

We have seen that the concept of the alienation of social institutions I have drawn from Marx's work of 1843-4 continues into his work of 1845, in a way which excises from it the idea that these social institutions are externalisations of, or somehow constitute, the "essence" of human beings, and which emphasises even more explicitly the role of social activity both in the constitution of social institutions and their becoming independent. It seems, then, that we can ourselves drop the notion of the human essence from our formal account of this concept, for it clearly plays a non-essential role in it, and is an aspect of this earlier account which many contemporary readers will find objectionable.

We can, therefore, sketch our concept of the alienation of social institutions, the clarification of which will be the task of subsequent chapters, as follows:

(1) A group of agents *X* act together, thereby constituting social institution *Y*

(2) Social institution *Y* becomes independent of group of agents *X*. This results from members of group of agents *X* mistakenly acting as if the social institution was independent of their activity as a result of a certain kind of mistaken belief.

(3) That (2) is the case enables social institution *Y* to influence group of agents *X*'s behaviour in a way in which it otherwise would not. This constitutes a restriction of their freedom.

Finally, from this basic structure we can see the sense in which this process as a whole can be alternatively described in terms of a particular *mode of existence* of social institutions. This is because the phenomenon described by (2) is explicated, in its second sentence, as a particular kind of the phenomenon described by (1).

Further, it is worth commenting on a curious feature of this process. Note that its result, in (3), is said to be that the social institution affects the behaviour of the group of agents in question. And see also that, in (1), this social institution is itself constituted by the activity of this group. It thus seems that the process of the alienation of social institutions is *reflexive* in a way in which the alienation that results in religion is not. Given this reflexivity, we can see that the concept of the alienation of social institutions potentially describes situations in which alienation not only affects agents' behaviour, but in doing so changes what this institution in fact is: by changing agents' behaviour, the alienation of social institutions potentially changes that out of which the social institution is constituted, and as such the social institution itself is transformed. Of course, not all behaviour by members of groups constitutes social institutions, and so cases where no such reflexivity is involved are quite possible: we might imagine scenarios in which the fact that an institution becomes independent enables it to affect only that behaviour which has nothing to do with the constitution of the institution (say, agents writing about it in social philosophy essays), and to have no effect on their institution-constituting behaviour. Nevertheless, as we will see, I take it that Marx's central interest is in alienated social institutions that display this kind of reflexivity.

In this sense, it seems we can view the claim about alienated social institutions having a particular mode of existence in a somewhat different light. That is, it seems that the alienated social institutions with which we are primarily concerned are

those that are constituted by social activity which is affected by the independence of that very social institution. If this sounds confusing—and, indeed, potentially circular—that is because it is: it is the purpose of chapters 3 and 4 to provide a translation of the concept of alienation to a more easily-understandable level of description, at which this claim will be more comprehensible.

Part 2
John Searle and the translation of social institutions to the level of individuals

In chapter 2, I sketched the broad outlines of Marx's concept of the alienation of social institutions. This was:

(1) A group of agents X act together, thereby constituting social institution Y

(2) Social institution Y becomes independent of group of agents X. This results from members of group of agents X mistakenly acting as if the social institution was independent of their activity as a result of a certain kind of mistaken belief.

(3) That (2) is the case enables social institution Y to influence group of agents X's behaviour in a way in which it otherwise would not. This constitutes a restriction of their freedom.

This broad characterisation leaves us with many unanswered questions. First, what even is a social institution, and how are we to understand the idea that these are constituted by a group of agents acting together? Indeed, what does it mean for agents to act *together* in the relevant sense? Second, what does it mean to say that a social institution becomes independent of these agents? If this consists in agents acting together motivated or caused by mistaken beliefs of some kind, then what kind of mistaken beliefs are these? Third, how are we to understand the way in which social institutions affect or influence peoples' behaviour? And how might this constitute a restriction of their freedom? In chapter 3 I will introduce some key ideas necessary for addressing these questions. In chapter 4 I will attempt to give a clearer account of (1) and (2), and in chapter 5 I will turn to those questions concerning (3).

Many of these questions arise primarily because the language and concepts we use at the social level of description at which the alienation of social institutions has been described so far do not translate easily to the micro or individual level of description at which we are more familiar with operating, and in fact blur important distinctions at this level. Our understanding of social-level concepts such as "social

institution" have much the same character as that which St. Augustine attributes to time: "If no one ask of me I know: if I wish to explain it to one who asks I know not".[1] And if we are to adequately understand the process our concept describes, we had better be able to explain—and not merely intuitively 'know'—what we mean by these terms.

If the concept of the alienation of social institutions is to be reconstructed in such a way that describing a social institution as alienated can be understood as making concrete, comprehensible claims about it, let alone as justifying a negative evaluative judgment of it, we need—I suggest—to examine its key terms more closely by *translating* them to the more familiar level of description of individual agents. I am not here concerned with whether this approach constitutes a case of "methodological individualism", for this term has been employed to describe many importantly different approaches.

Discussions of the kind of translation between levels of description that I am talking about commonly speak not of "translation" but of "reduction", and these terms both try to capture something of roughly the same operation. Yet the term "reduction" is used with many different meanings and associated implications, these being largely responsible for differences in authors' willingness to endorse the possibility of reduction. I will here briefly outline one such difference, in order to clarify what I mean by "translation".

"Reduction" is characterised by different writers as describing two importantly different operations, these being differentiated by what features found at the individual level are to be included in the reduction. Some claim that reduction describes an operation in which properties of groups are cashed out solely in terms of properties of individuals, as when we reduce the fact that *the Ferguson family is happy* to the fact that *each member of the Ferguson family is happy*.[2] Operating with such a definition, many theorists argue—and base their rejection of "reduction" on the claim—that there are properties of social phenomena that are not so reducible: for instance, that one cannot reduce the fact that *the stock market is down ten points* solely to facts about the properties of the human beings involved in the stock market, nor can one reduce the fact *Sally and Lucy are married* to facts about the properties of Sally and Lucy in isolation. In contrast, others claim that reduction is more liberal

---

[1] Cited in Hart, *The Concept of Law*, 14.
[2] See Epstein, *The Ant Trap*, 24–25.

in what features found at the individual level are to be included, for example by reducing "statements about social collectives … to statements referring solely to individual human beings, their actions, and the relations among them",[3] a definition which, if taken seriously, clearly makes reduction a more plausible endeavour than does the former account. In what follows, I will attempt a translation of the concept of the alienation of social institutions which is a "reduction" in this second, more inclusive sense.

Searle's social ontology

In order to translate the concept of the alienation of social institutions to the level of individual agents and their relations, we need first to know what is distinctive of the *social*: what do we mean when we talk of *social* activity and *social* phenomena?

I believe a good starting point for making such a translation is John Searle's work on social ontology. Searle's work can be seen as an attempt to translate what he calls "institutional facts", these being a subset of "social facts", to the level of individual agents and their relations. All of the phenomena that Marx suggests are alienated—money, the bourgeois state, the institution of private property, and so on—involve "institutional facts" in Searle's sense, although as we will see the relation between these phenomena and such facts is not straightforward.

Before turning to Searle's account in the following chapters, I would like to consider two methodological considerations. These raise important questions concerning the suitability of Searle's framework for reconstructing the Marxist concept of the alienation of social institutions.

First, Searle's account appears to be what we might call a thoroughly *rationalist* one: it deals with rational agents applying standards of rationality to explicit, conscious mental states in reasoning about what to believe, desire, and do. This picture is not quite accurate, for Searle's concepts of the "background" and the "network" make room for non-conscious mental states to play a role in an agent's reasoning processes. Nevertheless, it remains the case that Searle's exposition operates on the basis that social institutions function in some sense because they provide rational agents with good reasons to act—that whatever the nature of the mental states which the process

---

[3] Ullmann-Margalit, *The Emergence of Norms*, 14.

of reasoning operates on and in, i.e. whether conscious or in the background, it is nevertheless a standard of *rationality* that is doing the work—, and this might give us justifiable pause for thought. If this is meant to describe the social institutions that actually exist in our social world, and these are portrayed as existing on the basis of actions of rational agents acting on good reasons, how does this leave space for evaluation and critique of these institutions? If such space is not left, Searle's account would appear to lead to a kind of conservativism which would ill suit our purposes. More specifically, does this feature invalidate Searle's framework as a tool for explicating a concept—alienation—which has frequently been associated with describing the *irrationality* of certain social institutions?[4]

Of course, I have suggested that the normativity of the concept of the alienation of social institutions draws from its connection to a concept of *freedom*, not rationality. But even if we wanted instead to portray it as invoking rationality, the problem described above need not trouble us. Given that—as I have suggested—our concept concerns social-level phenomena, it seems we would need to invoke not a concept of individual rationality, but of collective rationality. And that a social institution is collectively irrational in some sense is quite compatible with the individual acts constituting it being individually rational. The literature on game theory is replete with examples of situations in which individually rational actions lead to what we might call collectively irrational outcomes. While, for reasons discussed in chapter 4, I do think Searle's work on social ontology has conservative ideological implications, these cannot be attributed to his emphasis on rationality.

Further, rather than invalidating his framework for our purposes, Searle's rationality-talk is useful for understanding the concept of the alienation of social institutions for two reasons.

First, it seems to be correct that social institutions, including those described by Marx as alienated, affect our behaviour at least partly by producing in us feelings of commitment, of making us feel that we "should" or "ought" do certain things and not others, and these concepts seem indelibly tied to a conception of rationality: we primarily understand "ought" claims as meaning that "all things considered I have good reason" to act in a particular way, and as such can be effective only for agents capable of acting for reasons more generally. While it is crucial not to downplay the

---

[4] See, for instance, Sensat, *The Logic of Estrangement.*

importance of physical force and violence in many social institutions, both at their creation and in their continued maintenance by means of police, armies, and incarceration, it nevertheless seems true, first, that such physical force alone is inadequate to maintain such institutions (as Searle puts it, that police forces and armies are themselves systems of status functions), and, second, that an account that relies only on such physical force would seem to miss a crucial feature of how social institutions are experienced. As such, any description of alienated social institutions would seem incomplete without capturing this aspect. Without it, we would seem to be talking simply of bad or criticisable—perhaps, even, alienated— *phenomena*, but not social institutions as we normally understand this term.

Second, the concept of alienation, as we have seen, is not only a descriptive concept, but also a critical one. Like other concepts used in the tradition of critical theory which Marx's work inspired, it aims not only at describing the world, but also at changing it. Without wanting to claim that the only tools critical theories have for achieving this world-changing goal is an appeal to the rational judgement and decision-making of the individuals whose activity they describe and reflect—for instance, it would seem plausible that some critical literature employs non-rational means to this end, inspiring emotions and non-rational responses that contribute in some way to this goal—it seems that this rational appeal is the means most open to the concept of alienation with which we are concerned. In this case, it seems that a concept that can describe and criticise as alienated those social institutions which *are* created and sustained primarily through rational decision-making processes holds out the possibility of contributing to the transformation of these, in a way that a concept which describes wholly non-rational processes could not.

The second methodological concern we must address is the following. Searle's account of social ontology is often portrayed as concerned exclusively with agents' thoughts and attitudes, of holding what we might call an *idealistic* conception of the relation between social institutions and agents according to which social institutions exist insofar as we *think* or *believe* that they exist. For instance, Brian Epstein writes that for Searle "The social world, quite generally, is the social world in virtue of our beliefs about it"[5] and that for him "We, as a community, make the social world by thinking of it in a particular way".[6] On this basis, his account is criticised for

---

[5] Epstein, *The Ant Trap*, 51.
[6] Epstein, 'Précis of The Ant Trap', 128.

ignoring the centrality of activity in social and institutional life. For example, Epstein contrasts Searle to H.L.A. Hart, whose account, according to Epstein, "involve actions, not just attitudes".[7] If this portrayal were correct, Searle's account would be of limited use for our purposes, given the centrality Marx gives to social activity in his concept of the alienation of social institutions, and in light of his and Engels' criticism of what they saw as a very similar idealism underlying Bauer and Stirner's work.

Searle certainly sometimes encourages this idealistic interpretation. He says on the first page of *The Construction of Social Reality* [CSR], for example, that "In a sense there are things that exist only because we believe them to exist. I am thinking of things like money, property, governments, and marriages".[8] However, to interpret Searle's overall account as Epstein does is to drastically mischaracterise it, in two senses.

First, the central object of analysis in Searle's account is what he calls "deontic statuses" (hereafter simply "statuses"). As we will see in chapter 3, statuses involve sets of "deontic powers" such as obligations, rights, and so on. For instance, when investigating the status "money" he is interested in what rights money gives to its possessor, what such rights consist in, and how they come about. While Searle does explain the creation of deontic powers involved in statuses, and thus statuses themselves, with reference to thoughts and attitudes, this is far from the whole story. As we will see, Searle's account of these centres around the rational constraints created through *speech acts*. While such acts, like all acts, have an attitudinal component, they clearly do not consist entirely in attitudes: Searle's idealistic-sounding statements are about necessary conditions for the creation of statuses, not their sufficient conditions.

Second, even if statuses were understood as consisting entirely in attitudes, the social world consists of more than statuses alone. It also involves people actually acting in particular ways. While statuses are the central object of Searle's analysis, he also recognises the central importance of activity. As he puts it in *Making the Social World* (MSW):

---

[7] Epstein, *The Ant Trap*, 104.
[8] Searle, *The Construction of Social Reality*, 1.

> Human social reality is not just about people and objects, it is about people's activities and about the power relations that not only govern but constitute those activities.[9]

These activities are what constitute *social institutions*, discussed in chapter 4. When we are examining the social institution of money, we are interested in the extent to which agents are actually *using* certain tokens to exchange for other things, their motivations, and the consequences of their so acting. While there is an important connection between statuses and social institutions (one which, I will argue, Searle suggests an overly simplistic account of), they are distinct phenomena. Therefore, even if Searle were portrayed as an "idealist" when it comes to statuses, he clearly does not mean his account of statuses to provide an exhaustive account of the social world. However, as I will argue in chapter 4, Searle often blurs his discussions of statuses into discussions about social institutions, using important terms in dual senses that sometimes pertain to one and sometimes the other. This is likely another reason for the mistaken interpretation of his work described above.

In chapters 3-4, I will reconstruct Searle's account such that its relevance for the concept of the alienation of social institutions can be seen. Chapter 3 addresses Searle's account of statuses; chapter 4 explores the relationship between statuses and social institutions, before using this framework to translate Marx's concept of alienation to the level of individual agents. Given the previously-mentioned ambiguities in Searle's account, these chapters are necessarily quite heavily reconstructive.

---

[9] Searle, *Making the Social World*, 106.

Chapter 3
Institutional facts, statuses and deontic powers

For Searle, facts such as that Donald Trump is the President of the United States, that the paper in my pocket is money in the United Kingdom, or that I am a doctoral student at the University of Sussex are "institutional facts". These are facts about *statuses*, or what we might commonly call "social statuses": facts about particular people or objects having the status *president*, *money*, and *doctoral student*. Such facts are for Searle epistemically objective facts the truth of which is not a matter of my or your *opinion*,[1] and they are such facts by virtue of "collective acceptance or recognition".[2] Donald Trump is President because most Americans accept that he is President, and similarly so with money and being a student.

Statuses of this kind are, for Searle, essentially connected to "deontic powers", by which he means phenomena such as obligations and rights. Searle observes that we often also use the term "status" to refer to labels that we give to people that have no associated deontic powers, such as being a bore, an alcoholic, or an intellectual.[3] I will ignore such non-deontic statuses: of interest in this chapter are only those statuses that have the appropriate connection to deontic powers. The precise relation between deontic powers and statuses is a little unclear in Searle's writing: he talks variously of deontic powers being "implied", "conferred", "imposed", or "carried" by statuses, and of the latter "consist[ing] of" the former.[4] I take it that the last of these formulations best describes Searle's position, and that for him statuses consist of deontic powers.

For Searle, this connection between statuses and deontic powers is essential to these statuses' ability to factor into agents' behaviour: this is because "deontic powers always provide desire-independent reasons for action".[5] That I have the status of being a doctoral student consists in the fact that I have the right to enter the library,

---

[1] Searle, 17–18.
[2] Searle, 8.
[3] Searle, 92.
[4] See, respectively, Searle, 23, 89, 95, 102, 106.
[5] Searle, 23.

the obligation to meet certain deadlines, the right to apply to become an associate tutor, and so on. And these deontic powers provide reasons that I and others have for acting in particular ways that are independent of what I or they would like to do: I have a reason to meet my writing deadlines even if I would rather go to the beach than finish this chapter; the security guard has a reason to allow me to enter the library even if they don't much like the look of me; and so on. I will focus initially on obligations, proceeding to discuss other kinds of deontic powers such as rights once we have a clear account of these. (Indeed, for reasons that will become clear, I take obligations to be more fundamental in Searle's account than are other kinds of deontic powers.)

One problem with Searle's exposition in *CSR* and *MSW* is that certain key terms—in particular those of "acceptance" and "deontic powers", as well as the subtypes of the latter such as "obligations" and "rights"—are not adequately elaborated. This causes significant problems because Searle uses these terms in interesting and somewhat idiosyncratic ways, and reading these books with a more standard meaning of these terms in mind—I would like to argue—has led many interpreters to misunderstand his overall position. (In the case of "acceptance" in particular this problem is exacerbated by Searle's flagrantly inconsistent usage, discussed in chapter 4.) Searle offers a fuller account of ideas pertinent to these terms in *Rationality in Action* (*RIA*), published between the two books just mentioned (though itself drawing on his earlier work, especially "deriving 'ought' from 'is'"[6] and *Speech Acts*[7]). He also offers a shorter account of these ideas that draws on the RIA account towards the end of MSW, although by the time this account appears it is frankly too cursory to undo the damage already inflicted.

In what follows, then, I will begin—*contra* Searle's own approach—at the micro-level, with the idea of a desire-independent reason (section 1) and the question of how desire-independent reasons for action are created by individuals' acceptance of obligations (section 2), before building from there to the more macro-level discussion of statuses and deontic powers such as rights (section 3).

---

[6] Searle, 'How to Derive "Ought" from "Is"'.

[7] Searle, *Speech Acts*.

1 Desire-independent reasons

To understand Searle's account of deontic powers requires understanding his account of desire-independent reasons for action, which lie at the centre of his account of human rationality:

> The single most remarkable capacity of human rationality, and the single way in which it differs most from ape rationality, is the capacity to create and to act on desire-independent reasons for action.[8]

In turn, understanding such reasons requires understanding two closely connected concepts: that of *reasons* generally, and that of rational *commitment.* As Searle continues:

> The creation of such reasons is always a matter of an agent committing himself in various ways.

Searle says relatively little concerning what commitment in this sense consists in, although he says considerably more about reasons. I will, therefore, begin with reasons, before reconstructing Searle's notion of commitment and the role it plays in his account of desire-independent reasons.

1.1 Reasons

Searle attempts to give a unified account of reasons, in which reasons for Intentional mental states and actions are discussed alongside reasons for "brute facts" such as the collapse of a bridge.[9] I will focus primarily on reasons for actions, but refer also to reasons for Intentional mental states such as beliefs. By "Intentional" I mean that property of being directed at or about states of affairs in the world.[10] I capitalise this term, following Searle in his main book on the subject,[11] to clearly differentiate it from "intentional" as used to describe an act which is done "on purpose" or deliberately.

---

[8] Searle, *Rationality in Action*, 167.
[9] Searle, 107.
[10] Searle, 34.
[11] Searle, *Intentionality, an Essay in the Philosophy of Mind*. Searle does not follow this practice in later works, however for clarity I will amend all quotations with capitalisation where necessary.

Because Searle attempts to give a unified account of reasons, he eschews the usual distinction between "good" and "motivating" reasons, which do not apply to reasons for brute facts.[12] However, when focusing specifically on reasons for actions he reintroduces this distinction using his own terminology.[13] He states that his primary interest is "justificatory explanations", by which he means motivating reasons that are also good reasons. A good reason is a consideration that speaks rationally in favour of (or "justifies") an agent doing something, whether or not that agent acts upon that reason or even knows about it. A motivating reason is a consideration that leads an agent to act. Searle's claim to be focusing on both confuses matters, so I will focus on good reasons, which are logically prior. (Indeed, Searle's subsequent discussion seems primarily focused on good reasons too.[14]) Hereafter I will refer to good reasons for an agent to act simply as "reasons" unless otherwise specified.

Another unusual feature of Searle's discussion of reasons—which seems to follow from his decision to talk of good and motivating reasons together—is his claim that not only facts but also mental states (and certain "other entities") can be reasons. He therefore introduces the term "factitives" to describe these various phenomena that are possible reasons.[15] Because I am focusing on good reasons, I take it that this neologism can be dropped: (good) reasons are always facts, including facts about mental states.

Searle emphasises that reasons are "relational" in several important ways.[16] Facts can never be reasons *simpliciter*. In so far as facts are reasons, they are always reasons *for something*; that is, in the cases of interest to us, for performing some action. Further, they are always reasons for *an agent*. Because our focus is on good reasons, this last point is separate from the question of whether the agent in question is motivated to act on the fact or even knows that it exists.

Finally, reasons are relational in a sense that is relevant to the question of *what makes a particular fact a reason* (for some action and for an agent). That is, Searle takes the answer to this question also to involve some sort of relation between a fact

---

[12] See Alvarez, 'Reasons for Action: Justification, Motivation, Explanation'.
[13] Searle, *Rationality in Action*, 109–11.
[14] See, for instance, Searle, 114.
[15] Searle, 103–4.
[16] Searle, 99, 115.

and the agent for whom it is said to be a reason. What this relation consists in is the question to which we now turn.

1.2 Rational commitment and making facts reasons

Searle primarily describes the relation that makes a fact a reason for an agent not as a relation between that fact and that agent, but a different relation. He says that "a statement is a statement of a *reason for an action* only insofar as that statement is systematically related to certain other statements".[17] He spells this idea out using another neologism, that of a "total reason".[18] This refers to a set of systemically related facts that when considered together rationally speak in favour of performing an action. For instance, that I desire to stay dry, that it is raining, and that I will stay dry in the rain only by carrying an umbrella together constitute a total reason for me to carry an umbrella. To constitute a total reason, Searle says, this set must contain at least one fact about the existence of a "motivator", by which he means an "element" that has the "world-to-mind direction of fit".[19] Searle uses the vague term "element" here because, while in the simplest cases such motivators are Intentional mental states such as desires, as we will see below he claims that certain non-mental phenomena can also play this role: there can be "external motivators".[20] By world-to-mind direction of fit Searle means the following (here using the example of mental states for simplicity). Intentional mental states have a propositional content, representing the state of affairs in the world that they are about, and a psychological mode which make them the kind of mental state that they are: a belief, a desire, an intention, and so on. This psychological mode makes the mental state the kind it is by specifying conditions of satisfaction. "Conditions of satisfaction" is Searle's term for the conditions which must obtain if a given Intentional state is to be successful: "A belief is satisfied if true, not satisfied if false. A desire will be satisfied if fulfilled, not satisfied if frustrated. An intention will be satisfied if carried out, not satisfied if not carried out".[21] The direction of fit specifies whether the satisfaction of the mental state consists in it fitting the way things are in the world (mind-to-world direction of

---

[17] Searle, 115.
[18] Searle, 115–19.
[19] Searle, 118.
[20] Searle, 119–20.
[21] Searle, 37.

fit, such as in belief) or in the way things are in the world matching it (world-to-mind direction of fit, such as in desire or intention).[22]

On this account, then, what makes a fact a reason for an agent to do something is its being "systematically related" to a motivator. This description is clearly insufficient for our purposes unless we know what makes an Intentional element with the appropriate direction of fit a motivator *for a particular agent*. For, surely, not just any such Intentional element—such as the desires of a complete stranger—will do. We therefore need a description of the appropriate relation between the motivator in question and the agent for whom a fact about that motivator constitutes part of a total reason.

It is in giving an account of how elements other than Intentional mental states can be motivators for an agent that Searle introduces the idea of "commitment", which I shall refer to as *rational commitment*. Searle employs this idea in two grammatical modes, that of an agent being "committed"[23] and that of them "having a commitment".[24] The former, I take it, is the more basic, while the latter allows him to fit his account of rational commitment into his account of external motivators. While we are primarily interested in reasons for action, we must here detour into Searle's account of reasons for Intentional mental states such as beliefs, returning to reasons for action in the following section.

The fact that an agent is rationally committed to an Intentional state x is a particular kind of reason for them to form x, namely one in which *the reason derives from the rational constraints internal to some other Intentional state or states that that agent currently has or had in the past*. I will here focus on commitments of the former kind, which derive from Intentional states the agent has currently.

The idea that Intentional mental states have rational constraints built into them is fundamental to Searle's account of rationality. He argues that rationality is not a matter of following certain rules, for example, that I am rational insofar as I follow the rule of *modus ponens*.[25] Rather, the rule of *modus ponens* summarises or describes a pattern of independently rational inferences. For Searle, rationality consists in the rational constraints inherent in the Intentional states that agents

[22] Searle, 35–39.
[23] Searle, *Making the Social World*, 124; Searle, *Rationality in Action*, 119, 175, 240–45.
[24] Searle, *Rationality in Action*, 112, 119, 136; Searle, *Making the Social World*, 131.
[25] Searle, *Rationality in Action*, 17–22.

have. For instance, to have a belief is to represent something in the world (the propositional content) in such a way that the agent takes it to be the case, and this 'taking' brings with it various rational constraints. If I believe that *p*, there is a rational constraint internal to this belief that I do not believe that not-*p*. For Searle such rational constraints are not unique to human beings. Animals, too, are constrained by their Intentional states in various ways: a dog cannot rationally believe that there is food in the bowl and also believe that there is no food in the bowl, and Searle thinks that animals do have such beliefs as well as the necessary background capacities for these constraints to make a difference to them.[26] As Searle puts it:

> once you [i.e. the agent] start representing those facts, with either direction of fit, you already have norms, and those norms are binding on the agent. All [I]ntentionality has a normative structure. If an animal has a belief, the belief is subject to the norms of truth, rationality, and consistency.… If *you* have a belief, I may be indifferent to the truth or falsity of your belief, but if *I* have a belief I cannot be similarly indifferent, because it is my belief and the normative requirement of truth is built into the belief. From the point of view of the animal, there is no escape from normativity. The bare representation of an *is* gives the animal an *ought*.[27]

Searle uses the formulation that an agent "has a commitment" to an Intentional state to refer to the fact that that agent is rationally committed to some Intentional state by virtue of having adopted some other Intentional state. If I am committed to believing *p* (by virtue of the rational constraints built into some other belief), I have a commitment to believe *p*. And formulated in this way, he can portray commitments as having both propositional content and conditions of satisfaction with direction of fit:[28] they have a propositional content that describes some state of affairs in the world (that some agent believes *p*), and they are satisfied only if the way things actually are matches that propositional content (the agent actually believing *p*). In this sense, then, Searle is able to portray such commitments as motivators, i.e. as "elements" that have the world-to-mind—or, rather, the world-to-commitment[29]—direction of fit and can therefore play the role of motivator in total reasons for belief. And since a commitment is not itself an Intentional mental state, it is described as

---

[26] Searle, 57–58, 182–83.
[27] Searle, 182–83.
[28] Searle, 174.
[29] See Searle, 188, where Searle uses the formulation 'world-to-obligation' (obligations being a form of commitment).

an *external* motivator. To see that commitments in this sense are motivators, we can see how facts about them are able to make other facts reasons for the committed agent. If I believe that *if p then q*, this commits me to the belief that *q* in any situation in which I believe that *p* obtains. Thus if it actually is the case that *p*, and I perceive and thereby come to believe that *p,* I am committed to believing *q*: my commitment makes the fact that *p* is the case a *reason* for me (to believe *q*).

I am a little uneasy with Searle's move from an agent's "being committed" to the formulation that they "have a commitment", and the parallel presentation of these as "entities" that have direction-of-fit. This formulation seems to obscure from view the reference to that fact which originally committed the agent (their having of a particular Intentional state) and therefore seems unnecessarily reifying. After all, my commitment in this sense is "external" in only a very limited sense: it is an Intentional mental state, an "internal" phenomenon, which does the committing. However, if we bear in mind that the formulation that an agent "has a commitment" is simply shorthand for their being rationally committed by virtue of the rational constraints internal to another Intentional state, no harm need be done. As we will see, Searle uses this framework to analyse considerably more complex cases of external motivators, where this point will be important.

Searle's account of rational commitment offers an explanation for how agents might have reasons that are independent of any desires that they have. As we saw above, we can say that my belief that *p* gives me a reason to believe *q* without any mention of desires. This account of desire-independent reasons may not be particularly controversial when considering theoretical reasons for beliefs, for we commonly take reasons for believing some proposition to be independent of our desires. Further, it captures an intuitive sense of what we mean by commitment in our everyday language. However, Searle makes more controversial use of this framework of commitments, arguing that we can also be rationally committed to certain *acts*, and can thereby have reasons for action that are desire-independent. We can now consider this argument.

2 Creating reasons for action

Searle takes his explanation of what makes facts reasons to *act* to be the key difference between his account and what he calls the Classical Model, amongst

whose proponents he includes Aristotle, Hume, Kant, and more recently Bernard Williams amongst many others.[30] Central to the Classical Model as Searle portrays it is a particular relation between a fact and an agent that makes the fact a reason for that agent to act. On this view, this relation necessarily involves some desire (broadly construed) that the agent has: facts are reasons for an agent to act *only* in virtue of their desires. In other words, for the Classical Model all reasons for action are desire-dependent. Let us ignore the adequacy of Searle's portrayal of the proponents of the Classical Model: I will refer to it simply as a foil against which to explicate Searle's argument.

Searle acknowledges that many reasons agents have for acting are such reasons relative to that agent's desires, and thus that the Classical Model can successfully explain many reasons for action.[31] However in contrast to the Classical Model he thinks that other features of an agent can play this role too. His account draws on the notion of rational commitment outlined above.

There is a trivial sense in which I can be said to be rationally committed, and thus have a reason, to act that makes no reference to desire. For Searle action is partly constituted by an Intentional mental state, an "intention-in-action". An intention-in-action captures something like the sense of "trying" involved when we intentionally act, in contrast to mere bodily movements such as twitches which do not involve such an Intentional state.[32] Just as we can be rationally committed to forming certain beliefs by virtue of the rational constraints internal to our pre-existing beliefs, we might say that we can be rationally committed to forming intentions-in-action and thus to acting by virtue of pre-existing Intentional states. The most obvious example is that I can be committed to act by the rational constraints internal to an intention that I have to carry out that act: if I intend to take a break in five minutes' time, then, after five minutes, I have a reason to do so, and that reason need make no explicit reference to desires. (I can equally well intend to ignore my desire for a break, feeling I don't deserve one, and this will similarly give me a reason for continuing to write.) In this sense, the rational constraints internal to the having of an intention commit me to acting at the relevant time. As Searle puts it, "one may commit oneself to a policy just by adopting a firm intention to continue with that policy".[33] Similarly,

---

[30] Searle, 5, 26.
[31] See, for instance, Searle, 192.
[32] Searle, 44–45.
[33] Searle, 175.

intentions are presumably subject to the rational constraints deriving from the logical relations of the facts that form their content. If I intend to fire a gun, and believe one fires guns by pulling triggers, then I am rationally committed by virtue of the rational constraints internal to my intention to pulling the trigger.

However, commitment to action in this sense falls far short of the kind of commitment involved when we talk of *obligations*, our primary explanandum. First, as we have seen, Searle portrays obligations as giving us desire-independent reasons for action, and in the commitments to act described above desires remain nearby. If I had formed the intention to take a break to satisfy my desire to rest, then my reason to act looks distinctly desire-dependent, and the Classical Model untouched. Second, the kind of commitments described above seem too easily changeable to capture what we mean by obligations. Obligations, it seems, are unaffected by our current attitudes: not so for the commitments to act described above. While I am rationally commitment to take a break by virtue of my intention to do so, I need only change my intention to lose the commitment. My obligations are not so easily dropped.

To move from rational commitments to act as described above to the stronger kind of commitment involved in obligations, and thus to properly desire-independent reasons for action, Searle suggests that not only attitudes that I currently have, but also certain kinds of actions that I have performed, can give me desire-independent reasons to act. In particular, he argues that "speech acts" can play this role. (Indeed, he seems to claim that such acts are the only possible origin of strong rational commitments to act.[34]) The move from the rational constraints internal to attitudes to those internal to speech acts as the origin of rational commitments is not as drastic as it might first appear, because Searle's account of action holds that actions have an Intentional component—the intention-in-action—which have rational constraints internal to them just as other attitudes do. Nevertheless, that these Intentional components are part of the performance of an action makes an important difference, and for Searle explains the characteristic features of obligations.

---

[34] See, for instance, Searle, 189, 157–64.

2.1 Speech Acts and Obligations

For Searle, then, speech acts create rational commitments—and thus desire-independent reasons—to act that are significantly stronger than those examined above. Searle uses the terminology of "commitments" and "obligations" interchangeably to refer to this kind of stronger rational commitment to act, usually listing them alongside one another (and various other terms).[35] He nowhere makes clear whether these are meant as synonyms, or refer to token concepts of a common type, but the former interpretation seems more plausible. To differentiate this stronger kind of rational commitment to act from those discussed above, I will refer to them as "obligations".

Obligations, then, play a similar role in practical reason as rational commitments to believe do in theoretical reason. Just as the claim that an agent "has a commitment" to believe means that they are rationally committed to some belief, the claim that an agent "has an obligation to x" means that they are rationally committed to x-ing (and thus to the intention-in-action that partly constitutes that action). And such obligations have propositional content (the agent x-ing) with conditions of satisfaction that are satisfied only if the way things are matches that propositional content (the agent actually performs x). As with the rational commitments to believe, then, obligations are thereby characterised as motivators—"elements" with a world-to-obligation direction of fit—and can play this role in total reasons. They can, therefore, make facts into desire-independent reasons to act.

I will here outline Searle's account of how agents create obligations for themselves by means of speech acts, which will highlight the continuity between obligations and rational commitments for beliefs, before examining their unique features.

Speech acts are acts performed with the *intention* of communicating something to somebody. Searle therefore characterises them as being partly constituted by intentions-in-action of a particular kind. When I perform a simple act, such as pulling on my socks, the intention-in-action that partly constitutes this action is also simple: it has the propositional content of me pulling on my socks, and conditions of satisfaction such that the intention is satisfied if I successfully pull them on (because of that intention). However, the intentions-in-action involved in speech acts are more

---

[35] Searle, 147, 176, 187.

complex. In telling you that "Sam smokes habitually", for instance, my intention is to do more than make the appropriate sounds, although I necessarily intend this too. Rather, I intend to make these sounds *and* for these to *mean* something to you. By these sounds meaning something, we here mean that these sounds themselves are given a propositional content and conditions of satisfaction with direction of fit: I intend that the words I utter themselves refer to a state of affairs in the world—Sam habitually smoking—with the conditions of satisfaction characteristic of an assertion, which will be satisfied if the propositional content matches the world and Sam does indeed smoke habitually.[36] Thus in performing speech acts we intentionally "impose conditions of satisfaction on conditions of satisfaction".[37] We can generally do this because of the constitutive rules of language (although we can sometimes perform simple speech acts without such rules), which we—as competent language users—understand as indicating that certain words with certain ordering, etc., count as the performance of a speech act of assertion-making, while others count as speech acts of promising, or requesting, and so on.[38]

For Searle, just about all speech acts create obligations and so desire-independent reasons to act in certain ways. Some, such as promising, do so explicitly. Others, such as assertion-making—which for Searle implies an obligation to, for instance, provide evidence for the assertion if challenged—do so implicitly.[39] I will focus on the explicit example of promise-making, before sketching how this account applies to the implicit example of assertion-making.

Just as Searle's account of rational commitments to beliefs drew on the rational constraints internal to the committed agent's already-held beliefs—for instance, those internal to their belief that *p* and that *if p then q*—, his account of the origin of obligations draws on the rational constraints internal to speech acts, and in particular internal to the complex intentions that partly constitute them. When I intentionally perform the speech act of promising by way of saying "I promise to x", there are various rational constraints internal to the intention-in-action involved, determined by the meaning in our language of the words used and the constitutive rules of speech acts of promise-making. When my utterance is part of a speech act characterised by such an intention-in-action, then, I rationally commit myself to the

---

[36] Searle, *Speech Acts*, 173.
[37] Searle, *Rationality in Action*, 179.
[38] Searle, *Speech Acts*, 33–42.
[39] Searle, *Rationality in Action*, 180.

logical consequences of these rational constraints, *in the same sense* as my beliefs that *p* and *if p then q* rationally commit me to believe *q*. In particular, for Searle competent language-users know that the speech act of promising is the creation of an obligation for the speaking agent:

> the relations between promising and obligations are internal. By definition a promise is an act of undertaking an obligation. It is impossible to explain what a promise is except in terms of undertaking an obligation.[40]

Thus if I, as a competent language-user who knows this definition of a promise, intentionally perform a speech act of promising, there is a rational constraint derived from the logical relation between promising and obligation such that, by virtue of promising, I create an obligation for myself to perform the promised act. And if I create an obligation for myself, and an obligation simply is—again, for me as a competent language-user—a rational commitment to act (of a certain kind), then I am rationally committed to (being rationally committed to) perform that act. Being so rationally committed is, for Searle, what it means to have an obligation. Thus, agents create obligations for themselves by means of what Searle calls "Declarations" (or what John Austin called "performative utterances"): agents make it the case that they promise by saying "I promise", and promising consists in the explicit creation of obligations. Since obligations are a kind of desire-independent reason for acting, agents thus create such reasons for themselves by way of such speech acts:

> Within these structures [the relevant constitutive rules, here those of language] if the agent acts with the appropriate intentions, that is sufficient for the creation of desire-independent reasons. Specifically, if the agent acts with the intention that his action should create such a reason, then if the circumstances are otherwise appropriate, he has created such a reason. The crucial intention is the intention that it be a reason.[41]

In Searle's terminology, introduced above, when I perform a speech act I impose conditions of satisfaction onto conditions of satisfaction. The immediate condition of satisfaction of my intention-in-action is that I perform the relevant utterance, i.e. I say "I promise to *x*". But I also intend this utterance to be a promise to *x*. And the condition of satisfaction of the promise to *x* is that I actually *x*. In acting with this

---

[40] Searle, 197.
[41] Searle, 212.

complex set of intentions, I thereby rationally commit myself to the imposed conditions of satisfaction:

> [T]he speaker can undertake commitments when he imposes conditions of satisfaction on conditions of satisfaction. Indeed there is no way to avoid undertaking commitments. The speech act of asserting is [the creation of[42]] a commitment to truth, the speech act of promising is [the creation of] a commitment to a future action. Both arise from the fact that the speaker imposes conditions of satisfaction on conditions of satisfaction. Speech acts commit the speaker to the second set of conditions of satisfaction.[43]

Searle's account of the implicit creation of obligations through assertion-making follows a similar pattern, except that the obligations derive from a "commitment to truth"[44] that is internal to assertion-making just as the creation of an obligation is internal to promise-making. When an agent makes an assertion, i.e. a speech act partly constituted by a complex intention-in-action to impose conditions of satisfaction onto their utterance, they are said to be rationally committed to the truth of that assertion, and being so committed implies certain obligations:

> When he intentionally imposes conditions of satisfaction on conditions of satisfaction, in the manner of an assertion, he takes responsibility for those conditions being satisfied. *And that commitment is already a desire-independent reason for action.* For example, the speaker has now created a reason for accepting the logical consequences of his assertion, for not denying what he has said, for being able to provide evidence or justification for what he has said, and for speaking sincerely when he says it.[45]

Searle does not elaborate this interesting notion of "commitment to truth", which perhaps raises more questions than it answers. However, for our purposes the key point is that speech acts of all kinds can create rational commitments to act, i.e. obligations.

2.2 The special features of obligations

Searle's account, then, highlights obligations' similarity to rational commitments of other kinds, such as my rational commitment to believe q created by my beliefs that

---

[42] Searle seems to use "commitment" both for the state of being rationally committed and for the act by means of which one rationally commits oneself.
[43] Searle, *Rationality in Action*, 183.
[44] Searle, 184–85.
[45] Searle, 173.

p and *if p then q*. However, as suggested above, obligations are stronger than other kinds of rational commitment to act. To understand this difference, we must examine the unique characteristics introduced by obligations' origin in speech acts.

Unlike rational commitments resulting from mental states, the rational commitments arising from speech acts are necessarily forward-looking. My rational commitment to believe *q* depends on my *currently* believing *p* and *if p then q*: that I previously had such beliefs commits me to no beliefs now. In contrast, it seems that my obligation to *x*, now, derives from my performance of a speech act of promising to *x in the past*.

This raises a potential problem for Searle's account. It seems that Searle's argument accounts for how a speech act can rationally commit me, at the time of speaking, to acting in the future. I cannot rationally promise to *x* at time *t* and to not-*x* at *t* without breaching the rational constraints internal to these speech acts. But this does not explain how I, at time *t*, will be rationally committed to act by virtue of speech acts performed in the past. There might be nothing stopping me, at *t*, from saying: yes, when I made the promise I created an obligation to *x* in the future. But now I desire not to *x*, and see no reason why my past speech acts trump my current desires.

To avoid this objection (and others), Searle introduces the idea that agents are, perhaps amongst other things, "selves" that persist through time. Such selves are the entities that have attitudes and intentions, and which act on reasons.[46] As such, they are the subjects of the rational constraints internal to these. And since these selves persist through time, these rational constraints also operate trans-temporally.

Searle justifies his (reluctant) invocation of trans-temporal selves in two ways. The first seems rather question-begging: he argues that such selves are presupposed by the notion of "responsibility".[47] While the notion of "responsibility" is central to the contemporary Western world-view, it hardly seems a sufficiently solid basis for this argument. It is possible that a mistaken belief in the idea of a self would give us a mistaken attachment to the idea of responsibility, so the fact that responsibility presupposes a self cannot be used as an argument for the self. Searle's second argument seems more promising: human practical reasoning is essentially about

---

[46] Searle, 83–89.
[47] Searle, 89–90.

time, about structuring our attitudes and actions in time.[48] In this sense, practical reasoning would be impossible did we not take our decisions to affect our future behaviour. The very possibility of practical reason thus presupposes that the agent that operates with reasons is the same across time, or at least that an agent at two points in time is connected in an important sense through these reasons. For instance, the notion of having an intention presupposes the idea that this intention can affect my future self. This idea of an agent that operates with reasons and that persists through time is what Searle calls the self.

Whatever we think of these arguments, what matters here is that for Searle agents can create desire-independent reasons for themselves in the future which they, in the future, will be rationally committed to. As Margaret Gilbert puts it, "Decisions … have a kind of *trans-temporal* reach. They continue to give their subjects reason to conform to them up to the moment of conformity".[49] Searle discusses this idea in terms of an agent's capacity for "recognitional rationality":

> The commitments you undertake are binding on you, because they are *your* commitments. That is, because you freely and intentionally made the assertion and thus committed yourself to its truth, it is not rationally open to you to say that you are indifferent to its truth, or sincerity, or consistency, or evidence, or entailment. Recognitional rationality is enough. You simply have to recognize your own self-created commitments and their logical consequences.[50]

Obligations, then, can be created for an agent in the future by that agent's speech acts here and now; an agent can have an obligation—and thus a desire-independent reason to act—by virtue of having created, in the past, such an obligation. In such situations, it is a fact's relationship to an agent's past intentional actions that makes it a reason *for them*. It is a fact, now, that "in the past I said 'I promise.'" And this fact is a reason for me to act as promised because of this fact's internal relation to my creation, then, of an obligation.

We can now return to the suggestion that obligations are *stronger* than other kinds of rational commitment. Searle seems to take this difference to derive from two

---

[48] Searle, 3, 90–91.
[49] Gilbert, *A Theory of Political Obligation*, 130.
[50] Searle, *Rationality in Action*, 180–81.

paradigmatic features of obligations: their *irreversibility*, and their *public character*.[51]

First, the idea that obligations resulting from speech acts are irreversible derives from their temporal character, described above. Where my rational commitments derive from an attitude that I currently hold, they are operative only as long as I hold the attitude from which they derive. If I cease to believe *p* or *if p then q*, then I am no longer rationally committed to believing *q*. In contrast, my past intentional acts cannot be so easily abandoned. This is why, as Margaret Gilbert points out, such obligations need to be "rescinded" by some new intentional act.[52] It is to this that I take Searle to be alluding when he describes speech acts as involving "an undertaking that is hard to reverse".[53]

Second, Searle emphasises the "public" character of obligations. For instance, he says that:

> the speech act is more than just the expression of an intention or the expression of a belief. *It is above all a public performance*. I am telling something to someone else.[54]

So, what difference does this make to the strength of the rational commitments (obligations) created thereby? Searle does not address this question specifically nor adequately. This passage seems to suggest that it is the fact that my words are intentionally communicating something to somebody else that makes the difference to the strength of the commitments involved. But why would this make a difference? Searle does not tell us.

In fact, I take it that the key idea underlying Searle's intuition that obligations are "stronger" than other rational commitments refers not to any difference in the "strength" of their rational commitment, but to a difference in their ability to factor into our motivation. Indeed, the notion that rational commitments can be "weaker" or "stronger" perhaps ought to strike us as strange: surely I either have a rational commitment to act in a certain way or I do not. This suggestion need not diminish the importance of this difference. As we saw above, Searle's ultimate interest is in

---

[51] He describes these as 'first, the notion of an undertaking that is hard to reverse and, second, the notion of an obligation.' This second notion is unhelpful for explaining obligations given its circularity. I take this to be referring to obligations' characteristic public nature. Searle, *Making the Social World*, 81–82.
[52] Gilbert, *A Theory of Political Obligation*, 130.
[53] Searle, *Making the Social World*, 81.
[54] Searle, 83–84. See also Searle, *Rationality in Action*, 183.

providing an account not only of good (or "justificatory") reasons, but of good reasons that actually motivate agents ("justificatory explanations"). And our focus on the former in this chapter is intended to clarify part of Searle's account before ourselves turning to the relation between obligations and actual activity, and thus between good and motivating reasons, in chapter 4.

If we were to account for this difference, then, it seems we would need a richer model of human motivation than the mere assumption that humans act rationally *qua* being responsive to the rational commitments internal to Intentional attitudes and speech acts, on which Searle's analysis is based. That is not to say, of course, that Searle's analysis would be irrelevant for such an account. We might, for instance, imagine an account of social motivation according to which agents are motivated partly by a desire or need to be recognised or treated by others as rational agents. If this were one's model of social motivation, then it seems an analysis of what constitutes a "rational" agent would be of fundamental significance, telling us what it is that agents wanted to be recognised as and how this would motivate them to act. With the addition of such an account, we would be able to see how obligations created through speech acts are both stronger than purely rational commitments, in the sense that they have an additional weight in our motivation, while also being of a kind with them. Adopting such an account, we might then see Searle's discussion of rational commitment as forming an aspect of—even a basis for—a fuller account of this kind.

According to Searle's account of obligations, then, to say that an agent has an obligation is to say that that agent is rationally committed to act by virtue of a certain kind of speech act that they have performed. This definition of an obligation has interesting implications which will become relevant in the following chapter, which are worth emphasising here.

First, for Searle obligations are necessarily *self-created*: to say that an agent has an obligation is to say that that agent has created an obligation for themselves. Searle makes this explicit with the example of "a group of people in Australia completely unknown to me [which] sets up a 'rule' whereby I am 'obligated' to pay them $100 a week. Unless I am somehow involved in the original agreement, their claims are

unintelligible".[55] (The fact that Searle always seems to locate such rogues in Australia raises an important issue, addressed below.)

This point seems to underlie Searle's repeated statements that obligations (and rational commitments generally) can only be understood from the "first-person point of view":[56] it is only with knowledge of the intentions on which agents act that we can know what obligations they have, for these intentions are a necessary part of the creation of obligations. This does not undermine the claim that when I am obligated I am *objectively* committed, nor that obligations give me *good*, and not merely motivating, reasons: as we have seen, it is an irreversible fact about me that I, for instance, made a promise with the appropriate intention-in-action; while such facts are ontologically subjective in that they would not exist were it not for my own intentional acts, they remain epistemically objective in that they are facts independently of whether anyone believes they are. Rather, this point seems to concern the possibility of acquiring knowledge about such objective obligations: insofar as my obligations depend on my speech acts, which are defined with reference to the intentions-in-action that partly constitutes them and which are accessible only to me, it is not always possible for others in my community, nor outside observers, to know what I am obligated to do. As we will see in chapter 4, however, this epistemic opacity reduces somewhat when obligations are created through speech acts of certain kinds.

With this emphasis on the first-person point of view in mind, it is worth returning to Searle's battle with the Classical Model. While the crucial difference that Searle emphasises between his own account and that attributed to the proponents of the Classical Model—that he takes agents to have desire-independent reasons to act— is vital to understanding Searle's account, it is important to highlight too a crucial similarity between these accounts. This is that on both accounts facts are reasons for an agent to act only by virtue of their relation to some fact about that agent and their Intentional states. While Searle argues that an agent can have reasons that are independent of their present desires, he is *not* claiming that agents can have reasons that are reasons for them completely independent of all present and past

---

[55] Searle, *Speech Acts*, 189–90. See also Searle, *Making the Social World*, 132; Searle, *Rationality in Action*, 208.
[56] Searle, *Rationality in Action*, 178, 209.

facts about themselves. For Searle as for the Classical Model, "all effective reasons are agent created".[57] As he puts it elsewhere:

> The only way that such entities [i.e. obligations] can be binding on rational selves is precisely if the rational selves freely create them as binding on themselves.[58]

(By "freely" in this passage, Searle refers to his relatively thin sense of freedom of the will, which is for him a necessary condition of all rationality: it refers to the necessary "gap" between the reasons that count in favour of me believing or acting in a particular way and my actually so believing or acting.[59] In this context, then, it simply specifies that the creation of commitment must be by means of some attitude or action: I must have actually formed the relevant belief or performed the relevant action—both of which require a "free" decision or choice "in the gap"—, rather than merely having had reason to do so.)

Second, for Searle to say that an agent has an obligation is, *analytically*, to say that that agent is rationally committed to acting. It is part of his definition of an obligation that the obligated agent is rationally committed (in a peculiarly irreversible and public way) to act. With this definition in place, Searle does not need an additional account to explain why agents ought to act in accordance with their obligations, beyond the assumption that agents are practically rational (alongside his account of how rationality requires us to recognise our rational commitments, i.e. of "recognitional rationality"). If an agent has an obligation, this means that they are rationally committed to act, and assuming they are rational they will therefore be motivated to act as they are obligated to. The key question in understanding the motivation of rational agents is therefore shifted away from that of how agents might be motivated to act according to their obligations, and towards that of why and how rational agents might create obligations for themselves in the first place. This point appears to have been missed by many interpreters of Searle: I return to it below.

---

[57] Searle, 189. Searle seems to ignore this similarity when describing in Searle, *Making the Social World*, 131 what the Classical Model gets right.
[58] Searle, *Rationality in Action*, 206.
[59] Searle, 13–17, 61–96.

2.3 kinds of obligation-creating speech acts

In previous sections, we have examined the creation of obligations through the speech act of promising, although I noted that for Searle almost all speech acts create deontic powers or obligations of some sort. However, it is important to emphasise that for Searle not all speech acts involve language, and there are therefore many obligations we can create for ourselves without saying anything at all.

An example that Searle gives to demonstrate this is when I buy somebody a drink, bring it to the table and push it towards them.[60] In this case, it would seem that I have created (amongst other things, considered below) an obligation for myself, namely to not start drinking the drink I have given you. In doing so, I have performed a speech act, and thereby created an obligation for myself, without language. Once again, the key to seeing this lies not in any specific feature of the bodily movement involved. Rather, it lies in the combination (in the context of certain constitutive rules, discussed below and in chapter 4) of a particular kind of complex intention-in-action—in which I intentionally impose conditions of satisfaction onto my bodily movement so that this be understood by you as meaning that this drink is yours—with a bodily movement, which adds the irreversibility and publicity required for the creation of obligations.

I will argue in chapter 4 that there is at least one important difference between the creation of obligations by means of speech acts that involve language and those which do not. However, for present purposes this difference can be ignored: an agent can create obligations for themself through speech acts of both kinds.

In some passages in which he discusses promising, Searle introduces another pair of terms to describe the speech acts by means of which agents create obligations for themselves, "acceptance" and "recognition". For instance, he says that:

> No analysis of the concept of promising will be complete which does not include the feature of the promisor placing himself under or undertaking or accepting or recognizing an obligation to the promisee to perform some future course of action, normally for the benefit of the promisee[61]

> The notion of an obligation is closely tied to the notion of accepting, acknowledging, recognizing, undertaking, etc., obligations[62]

---

[60] Searle, *Making the Social World*, 89.
[61] Searle, *Speech Acts*, 178.
[62] Searle, 189.

This introduction is crucial, for these terms arise again in his treatment of statuses. However, these terms fit somewhat awkwardly with an account that emphasises the role of speech *acts*, for both are ambiguous between what we might call a "passive" and an "active" sense. For instance, I might come to "accept" my doctor's prognosis, or I might "recognise" that you are the same person I spoke to yesterday: both of these seem passive in that they refer to mere attitudes that I have as opposed to an act I perform. In contrast, I might "accept" you into our club by giving you a membership card, or the United States might "recognise" Juan Guaidó as Venezuela's President by publicly declaring him to be so: both examples employ these terms in an active sense in that they refer not to mere attitudes but to acts.[63] I take it that Searle's use of these terms in both passages employ the active sense of these terms, as suggested by their inclusion alongside "undertaking" and "placing himself under". Insofar as Searle's account is concerned with statuses and thus with the obligations and other deontic powers connected to them—the ultimate focus of this chapter—, I take it that his invocation of this terminology is again meant in this active sense. In order to clearly differentiate acceptance or recognition in this active, obligation-creating sense from the more passive, merely attitudinal sense, I will in what follows refer to the former as *committing acceptance*. The ambiguity between committing acceptance and the more passive sense of these terms will arise again in chapter 4.

The idea that we can create obligations by means of speech acts of committing acceptance fits well with some of our everyday language. For instance, we sometimes use the language of acceptance or recognition when an agent creates obligations for themselves reluctantly or without any positive second-order attitudes towards the obligated act. Searle makes this clear in his insistence that "Acceptance, as I construe it, goes all the way from enthusiastic endorsement to grudging acknowledgement".[64] This point can be perhaps best demonstrated with an example discussed by Frank Hindriks:

> The difference between acceptance and endorsement is important. It is exemplified in a particularly salient way in Al Gore's concession speech

---

[63] That recognition or acceptance have an active sense is a common theme in work inspired by Alfred Schutz. See Schutz, *The Phenomenology of the Social World*, 77. See also Thomason, *Making Sense of Reification*, especially the foreword by Thomas Bottomore, p. x. Also Luckmann and Berger, *The Social Construction of Reality. A Treatise in the Sociology of Knowledge*.

[64] Searle, *Making the Social World*, 8. See also Searle, *Speech Acts*, 194–95.

that he delivered in December 2000 after the U.S. Supreme Court had
decided in favor of George Bush as the next president in spite of the
contested and decisive outcome of the election in Florida. Gore's words
were the following: 'Let there be no doubt, while I strongly disagree with
the court's decision, I accept it.'[65]

Hindriks' subsequent discussion asserts that in light of this difference "Gore's
acceptance is a factual attitude". But this cannot be right. Gore's statement is not
simply the statement of a belief, for example that he believes the Supreme Court in
fact came to the decision that it did, or that *others* endorse its decision. (Contrast
Michael Gove's merely factual claim: "I accept that the Supreme Court have made
this judgement".[66]) Rather it is clearly either itself a declaration akin to a promise
("I *hereby* accept") or the reporting of a previous act of acceptance in precisely the
sense of *committing* acceptance: it is or reports his (reluctant) committing of himself
to act in certain ways with respect to George Bush, and thus the creation of an
obligation. While Gore wishes the Supreme Court did not rule that Bush had these
rights, he *both* (factually) acknowledges that it did *and* accepts the decision in the
sense that he has rationally committed himself to treating Bush appropriately.

With such examples, we have strayed from our narrow focus on individual
obligations, towards the more macro-level phenomena of statuses (President) and
what Searle calls "positive deontic powers" such as rights. It is to such phenomena
that we now turn.

3 Statuses and deontic powers

One necessary task for any reconstruction of Searle's account of statuses is to clarify
the various senses in which he uses the adjective "collective" with respect to
Intentional mental states generally and intentional acts of committing acceptance in
particular.

In a famous 1990 paper 'Collective Intentions and Actions'[67] Searle explicated his
concept of *collective Intentionality*. Collective Intentionality is Searle's term for the
Intentional attitude of a group, as when a group is said collectively to believe, desire
or intend something. It is therefore also used in his account of collective *activity*:

---

[65] Hindriks, 'Collective Acceptance and the Is-Ought Argument', 473–74.
[66] 'Interview with Michael Gove, 25/09/2019'.
[67] Searle, 'Collective Intentions and Actions'.

when a group acts together, it is said to act on a collective intention in this sense. Searle describes cases of collective Intentionality as cases in which each individual in a group has an attitude which they each attribute not to themself as an individual but to the group as a whole, and in which the fact that each member has such an attitude is a matter of common knowledge. In the case of collective activity, each group member is said to have a collectively-attributed intention or "we-intention".

While there seems to be an important ambiguity in Searle's use of the collective adjective even in this classic account,[68] one which continues into his later work, this need not concern us. More pertinent is that Searle's discussion is uncharacteristically vague concerning what role he takes such collective Intentionality to play in explaining statuses. In both *CSR* and *MSW* he recapitulates his account of collective Intentionality, and explicitly says that understanding this concept is necessary for understanding statuses. Yet he never adequately explains precisely what role it plays, nor how collective Intentionality relates to the notion of *collective acceptance or recognition*. Indeed, in MSW Searle complicates matters considerably by stating that, since collective acceptance or recognition can be constituted by individually-attributed attitudes, this "shows that there are some forms of collective [I]ntentionality which are reducible to I-[I]ntentionality plus mutual belief",[69] thereby seeming to abandon his classic account of collective Intentionality altogether. I will not here attempt to unify the ways in which Searle discusses "collective" attitudes. Rather, I take it that there are several importantly different ideas each referred to in his work with such language, and will explicate these separately when they each arise.

3.1 Positive deontic powers

As we have seen, Searle places an account of obligations and other "deontic powers" at the centre of his account of statuses and their ability to give agents desire-independent reasons to act. In previous sections, I reconstructed a Searlean account of obligations, according to which an agent can create obligations for themself by performing speech acts. However, for Searle statuses involve more than simply

---

[68] Margaret Gilbert deftly articulates this ambiguity as that between the collectively-attributed attitude in an individual's mind, and the Intentional attitude "of" the group: Gilbert, 'Searle and Collective Intentions'.

[69] Searle, *Making the Social World*, 58.

obligations created by individuals for themselves. Rather, they involve what he calls "*positive* deontic powers",[70] phenomena such as rights and entitlements that involve obligations on the part of more than one agent.

Searle often talks of obligations alongside such positive deontic phenomena as "rights,… permissions, authorizations, entitlements, and so on",[71] as each being examples of "deontic powers". As can be seen from this list, "power" must here be understood in the normative sense, as in the idea of "power of attorney", or as when a police officer talks of "the powers vested in me", rather than in the sense of an agent's *actual* power to do something. (This distinction will arise again in chapter 4.) It is worth reflecting on the counterintuitive nature of the inclusion of obligations alongside these other phenomena: obligations do not seem like the kind of thing that give the obligated agent any new normative "power". However, I take it that in doing so Searle has stepped away from the individual obligated agent's point of view: instead, he is talking from the point of view of a *group* or "we", and from here the inclusion makes more sense: my obligation can give *others* normative powers with respect to me.[72] For instance, my obligations as a doctoral student give the university administrators the power to demand certain things of me. It is from this point of view that we must understand Searle's account of "positive deontic powers"[73] such as rights. (I will hereafter refer only to rights; I am not sure if the other terms listed above are meant as alternative kinds of positive deontic powers or merely as synonyms, but I take it that what can be said of rights can broadly be said of them also. Searle does not analyse these deontic phenomena in a detailed way *a la* Hohfeld,[74] and his discussion would surely benefit from such an account. For our purposes, Searle's somewhat vague usage will suffice.)

While obligations were portrayed above as potentially very broad in content—I can create an obligation for myself to perform just about any act—Searle's account seems to draw primarily on obligations that are directed in some sense towards other agents: they are "typically made to others, but they can be made to oneself, as well".[75] I take it that this directedness concerns whatever difference we take the public nature of the speech acts that create obligations to make (see section 2.2) being

---

[70] Searle, 9.
[71] Searle, 8–9.
[72] Searle, 106.
[73] Searle, 8–9; Searle, *The Construction of Social Reality*, 100.
[74] Hohfeld, *Fundamental Legal Conceptions*.
[75] Searle, *Rationality in Action*, 150.

applied to you in particular: I am obligated *to you* to do as I promised, because it was to you that I made the promise. It seems likely that most cases in which my obligation is directed to you involve acts of mine that affect you and your ability to act, however this does not seem a necessary feature: if I promise you that I will brush my teeth, I am obligated to you to do so regardless of the fact that my teeth-brushing does not (or, at least, barely) affects you. This directedness carries over to Searle's discussion of rights, which—drawing on a characterisation of rights originated by Hohfeld and most famously put forward by H.L.A. Hart—takes rights to be similarly "directed".[76] This directedness of obligations and rights puts them in a reciprocal relation. As Searle puts it:

> rights are always rights *against* somebody. If … I have a right of easement to cross your property, then that is a right against you. And you have an obligation not to interfere with my crossing your property. Rights and obligations are thus logically related to each other. If X has a right against Y, Y has an obligation to X.[77]

Of course, when we talk of rights we usually refer to rights against not just one individual, but against many. My property rights consist not simply in *your* obligation not to take what is mine, but in *everybody's* obligations of this kind. Thus, as I understand it, Searle's account suggests that *rights are constituted by the obligations of one or more agent.* Put another way, rights are names for patterns of obligations across several agents: to say that I have a right *simply is* to say that certain other people have the corresponding obligations. My right to enter the University library is constituted by the librarians and security guards' obligations to allow me to do so. Deontic powers therefore name normative "power relationships between actual people",[78] namely relationships in which agents have obligations *vis-à-vis* one another.

We saw above that for Searle knowledge about an agent's obligations requires that we take their first-person point of view, for it is only with reference to the intention-in-action that partly constitutes the relevant speech act that they can be said to be rationally committed and thus to have an obligation. However, when it comes to rights this individual first-person point of view will no longer do: to know that I have a right, I must know that others have the corresponding obligations, and it seems I

---

[76] Hart, 'Are There Any Natural Rights?'
[77] Searle, *Making the Social World*, 177.
[78] Searle, 98.

cannot know this from my own point of view. However, it seems that neither can such powers be known from the external, third-person viewpoint of the anthropologist: since the obligations that constitute my right can be known only from the first-person point of view of the obligated agents, an external observer cannot know these either.[79] While Searle generally eschews epistemic questions of this sort,[80] I take it that it is something like this argument that underlies his shift, in discussing statuses and their associated rights, to taking the first-person *plural* point of view: statuses, he believes, are best understood and spoken about from the point of view of *we*. (As mentioned above, Searle's concept of collective Intentionality centres on discussion of we-intentions (we-beliefs, etc.). I am not here invoking that account, but merely describing the conceptual point of view from which Searle operates when discussing statuses.)

An important sense of what Searle means by "collective acceptance" here presents itself. If agents create obligations for themselves by means of speech acts of committing acceptance, and if rights are constituted by the obligations of several agents, then it seems that rights can be created by collective acceptance that does not require any we-intentions in the minds of individuals: if each of us commits her/him/themself to act such that we each have the relevant obligations towards someone, then that person has the relevant right against each of us. As I understand the term, this distributive sense of collective acceptance is in fact the most fundamental to Searle's account of statuses.

As with our account of obligations, then, the key question concerning rights is not how an agent can be rationally committed to acting in accordance with them—in conditions of collective acceptance so defined, it is an analytic truth that each accepting agent is so committed—but rather how they were created. In particular, when it comes to rights the relevant question seems to be: how did the various agents come to coordinate their respective obligations in this particular configuration? Why do I have the right to enter the Sussex University library, while others do not? Why would the other people involved have created the relevant obligations for themselves?

One possibility, of course, is that a particular right emerged either by happenstance or by virtue of some feature of the world that each obligated agent independently

---

[79] Searle, *The Construction of Social Reality*, 5, 98.
[80] Searle, *Making the Social World*, 6.

took to be normatively relevant: each member of a group happens to have undertaken the relevant obligations to me such that together these constitute a right possessed by me against them for me to do something. This, of course, sounds rather far-fetched, but we could concoct a relevant example. In discussing "moral obligations", Searle gives the example of an obligation he takes himself to have to help others who are in desperate need of immediate help and whom he is able to help.[81] We might posit as his motivation for creating this obligation a belief that others' suffering is a bad thing, and that, motivated by this belief, he performs a speech act that rationally commits him to helping them whenever he can. Now, if—not unreasonably—it happens that most members of his community have an identical pattern of thought, and each creates obligations for themselves to so help those in need, we might plausibly start talking of "rights": in this community, those in desperate need have a right against others to help, a right constituted by these others' intentionally created obligations. When in need in such a community, I might plausibly say to Searle that "I have a right to your help", and this statement need not be understood as an appeal to some objective, external moral truth, but simply as what I take to be a correct statement based on the presumed fact that he, as a member of the community that I believe consists of people who have each undertaken the relevant obligations, has done so.

Even granting the plausibility of this story, however, it should be clear that this mechanism for the coordination of obligations is not operative in all, or even very many, of the rights we encounter in social life. When I stand at the front of a queue, my right to be served next does not derive from any features of the situation that each person behind me in the queue happens to independently find normatively salient. And yet such rights seem just as phenomenally real as rights of the kind described above: certainly, personal experience suggests that people take them very seriously. How else, then, might groups of agents coordinate their creation of obligations such that together they constitute rights? I believe Searle's arguments concerning statuses can be best understood as addressing this question.

---

[81] Searle, 91; Cf. Searle, *Rationality in Action*, 157–65.

3.2 Statuses and status terms

Before reconstructing the role I take Searle to give to statuses, it is worth disambiguating several notions that sometimes get blurred together in his account. As mentioned above, I am reconstructing Searle's account in the opposite direction to that in which his explication travels: whereas he starts with institutional facts and statuses and works down, ending (in MSW at least) at the micro-level of agents' desire-independent reasons to act, I began my explication at this micro-level and am building up from there. It is partly for this reason that certain features of Searle's terminology seem unhelpful for us. Searle's discussion is primarily concerned with describing how *we*, considered as a group, "impose" statuses on people and objects. In contrast, my focus is on getting to this level from the point of view of individual agents: what am *I* doing when *we* impose a status? It is only from this point of view— according to Searle's own discussion of how obligations are created—that we can capture why rational agents might be motivated, by "desire-independent reasons", to act in accordance with deontic powers associated with these statuses.

Searle primarily uses two terms for discussing statuses: "status" and "status function". While his basic idea in using these terms is relatively clear, and he sometimes explicates quite clearly what he means by them,[82] in his writing both tend to be used to refer to several different —and overlapping—aspects of his account, and this confuses matters. I will here disambiguate four distinct notions that Searle discusses with these terms.

First is what I will call simply a *status*. When an object or person has a particular status, this means that we can understand them as being a particular *kind* of object or person, by virtue of their having certain properties.[83] For instance, that I have the status "student" means that I am, amongst other things, the kind of person that is a student, and that I belong to this kind by virtue of having the properties characteristic of this kind. The properties that characterise statuses consist in the possession or carrying of certain deontic powers, either deontic powers of the person who has the status or of some other agents who have a particular relation to the person or object who has the status. Thus, I am a student by virtue of my property

---

[82] Searle, *Making the Social World*, 99 perhaps articulates their meaning most clearly.
[83] See e.g. Searle, 10, 96. The description of statuses as kinds is my own reconstruction. I will focus on paradigmatic cases in which statuses are imposed on pre-existing objects or people. Searle also discusses exceptions to such cases: Searle, 20, 108–9.

of being obligated to act in certain ways, of possessing certain rights, and so on. Similarly this paper has the status "money" by virtue of carrying rights for its possessor. In this sense, the existence of the relevant deontic powers is constitutive of the status: that I have these rights and these obligations constitutes my having the status "student"; that this paper carries these rights is constitutive of it's having the status "money". It is statuses in this sense that Searle's notion of "institutional facts" is primarily concerned with: institutional facts are facts about certain people or objects having statuses (i.e. carrying deontic powers such that they are of the relevant kinds).

Second is what we might call the *function of a status*.[84] As we have seen, statuses consist in agents having deontic powers. For Searle, it is the "function" of a status to "carry" these deontic powers: the point of agents having statuses is that when they have such statuses they have deontic powers.[85] That this "function" is built into the very definition of a status perhaps explains Searle's use of the phrase "status functions", however it is useful to keep the two notions separate.

To talk of "functions" of statuses in this sense, it seems we must presuppose some goal or purpose of the group whose activity imposes a status. Such purposes are another thing that could be meant by the "function" of a status. However, this implies a notion of collective purposes that I do not think Searle wishes to commit himself to: while, as we will see, he takes it that some statuses are imposed through cooperation, in which talk of a collective purpose might seem plausible, he denies that all impositions of statuses are imposed in this cooperative manner. As such, I will avoid talking of the function of statuses in this second sense: it is sufficient that we understand that statuses *in fact* (and by definition) carry deontic powers, without needing to imply that this carrying serves a particular purpose. (We will encounter this distinct notion of a "function" again in chapter 4.)

The third important notion in Searle's account is what I will call a *status term*.[86] When we talk of statuses, we are necessarily talking about phenomena occurring at the macro level: it is *we* who impose statuses on people and objects. Status terms, on the other hand, are symbolic devices used in the thoughts and attitudes of individual agents. I name these "status *terms*" because these symbolic devices are

[84] See Searle, *Making the Social World*, 59, 94.
[85] Searle, 8–9.
[86] e.g. Searle, 95, 102, 124; Searle, *The Construction of Social Reality*, 46, 72, 73–74.

paradigmatically words, although Searle points out that they need not be.[87] Status terms *represent* possible statuses and their characteristic properties, and thus also represent the set of deontic powers associated with them. These deontic powers thus constitute part of the definition of the status term for the agent who is using it. As Searle puts it: "Typically the associated function is definitionally implicit in the expression that names the status",[88] i.e. the deontic powers the carrying of which is the function of the status are included in the definition of the status term. For instance, in thinking about whether I want to become a student, I can use the status term "student" in my reasoning: if I become a student I will have less free time (because "student" represents amongst other things a set of obligations to submit work); but I will also be able to develop myself academically (because "student" represents amongst other things a set of rights to teaching and support).

Finally, Searle sometimes uses these terms to refer to what I will call the *function of a status term*.[89] As we have seen, status terms *represent* bundles of deontic powers to agents: in some of Searle's discussion, so representing is portrayed as the *function* of status terms. This use of the term "function" seems less problematic here than it does in the second notion of the function of a status, for it implies no collective purpose: the function of a status term is to represent the set of deontic powers associated with a status for the agent in whose thoughts the status term is used; it is a function relative to that agent's purpose of representing and reasoning about the world.

3.3 The coordination of obligations constitutive of statuses

Our central question, then, concerns how we get from the micro level of description, at which we can talk of individual agents being rationally committed to act and thus as having obligations, to the macro level of description at which certain people and objects have statuses which carry deontic powers including both obligations and rights. As we might expect from the definitions above, Searle's account emphasises the role of status terms in connecting these levels of description.

---

[87] Searle, *The Construction of Social Reality*, 73–74.
[88] Searle, 114.
[89] Searle, 39–40, 60, 75.

I take it that Searle's answer to our question of how groups of agents might coordinate their respective creations of obligations such that together these obligations constitute positive deontic powers such as rights is that, often, this happens by way of agents committing themselves to acting in the appropriate way through speech acts of committing acceptance the content of which employs a status term. I will here attempt to spell out how such an account might work, and the kind of speech acts that might be involved in creating and sustaining statuses.

It seems that for Searle an agent can create an obligation for themself by means of any of a number of different kinds of speech act. The most explicit kind, in which an agent makes an explicit performative utterance similar to saying "I promise to treat agent *x* as having status *y*", would seem to mirror the promising case outlined above too closely, and indeed to be too rare, to be worth explication here. I will therefore focus on cases in which the speech act in question is a form of assertion. To make it as clear as possible, I will lay out how I take Searle to get from the micro level of individual speech acts to the macro level of statuses by way of the use of status terms in a series of numbered steps.

1. Recall that for the agents that use them, status terms represent the deontic powers that are characteristic of a possible status.

2. As we have seen above, deontic powers consist in agents being rationally committed to act in certain ways: my obligation consists in the fact that I am rationally committed to perform the relevant act, while my right consists in the fact that others are similarly rationally committed.

3. Given 1. and 2., status terms represent agents being rationally committed to performing certain acts.

4. Status terms can be used by agents to represent propositions concerning (perceived or potential) statuses. Given 3., the proposition that *x* has status *y* entails the proposition that agents that stand in certain relations to *x* have obligations, and thus rational commitments, to act in certain ways.

5. Speech acts, in which agents intentionally impose meaningful propositions with conditions of satisfaction onto their utterance, bodily movement, etc., rationally commit (in a strong, public sense) the speaking agent to those imposed conditions of satisfaction. For example, if I make an assertion (that is, if I say to one or more others, with the appropriate intention to

communicate) that "it is raining", I impose a meaning onto the noises that I make, namely the proposition "it is raining" with the word-to-world direction of fit, such that this utterance is satisfied if the proposition is true and it is in fact raining. In making this assertion, I rationally commit myself to these conditions of satisfaction, and thus to the truth of that proposition. This rational commitment is said to include a rational commitment to certain kinds of acts, such as providing evidence when questioned.

6. If I make an assertion predicating a status term of some object or person, such as "person x has status y", I rationally commit myself to the truth of this proposition. Given 4., this rationally commits me to the truth of the proposition that agents that stand in certain relations to x have obligations and thus rational commitments to act in certain ways.

7. If I stand in a relation to x such that if x had a particular status this would entail obligations *for me* (and I know that I stand in such a relation), then if I make an assertion which involves the application of the corresponding status term to x, then I am rationally committed to the truth of the proposition that certain agents *including me* have obligations and thus rational commitments to act in certain ways.

8. If I am rationally committed to the truth of the proposition that I am rationally committed to act, then I am rationally committed to act. Since an obligation consists in an agent being rationally committed to act, I therefore have an obligation to act.

9. If every agent who stands in the appropriate relation to x performs an equivalent speech act as in 7., they are each rationally committed to performing the act specified for them in the definition of the status term. Thus each agent has the relevant obligation, and these obligations together constitute the rights associated with the status. x thus has the y status by virtue of possessing the rights characteristic of this kind.

We can spell this out with a brief example. Take the status "teacher". Say that a teacher is a kind of person who has an obligation to teach certain other people ("students"), and the right to receive coursework at the end of term, a right that is constituted by the obligations of students to submit such coursework. Assume that three agents x, y and z use the status terms "teacher" and "student" to represent

these possible statuses and their constitutive deontic powers. If agent x states that "I am the teacher and y and z my students", and agents y and z each states that "x is the teacher and I am their student", they each rationally commit themselves to the truth of these assertions. Thus x rationally commits themself to (the truth of the proposition that they are rationally committed to) teaching y and z, and each of y and z rationally commit themselves to (the truth of…) submitting coursework at the end of term. Given these rational commitments, x actually has the obligation to teach y and z, and y and z each have the obligations that constitute x's right to receive coursework. As such, x has the properties constitutive of the status "teacher", and we can say that x is a teacher.

I said previously that Searle uses the term "acceptance" in an active sense—which I called "committing acceptance"—to describe speech acts by means of which agents rationally commit themselves to act. From what has been said above, we can see that 7. describes a case of committing acceptance. That is, just like promises, statements involving status terms can rationally commit agents to act, i.e. can create obligations for the speaking agent. Thus, acts of committing acceptance can be speech acts of either of two different kinds: explicit, as in promises, or implicit, as in statements involving status terms.

Further, 9. constitutes a case of *collective acceptance or recognition* of the kind that Searle describes as sufficient for the creation of statuses:

> As long as there is collective recognition or acceptance of the institutional facts, they will work. They work because they consist of deontic powers, and the deontic powers will function if they are accepted.[90]

> The institution and the institutional facts within the institution require continued recognition or acceptance because they exist only as long as they are so recognized or accepted.[91]

Thus, Searle uses this language to describe the creation of the status "President":

> The essential component of that fact [that Barack Obama is "President"] is that people regard him and accept him as president, and consequently they accept a whole system of deontic powers that go with that original acceptance.[92]

---

[90] Searle, *Making the Social World*, 106.
[91] Searle, 103.
[92] Searle, 170.

Above, we saw that for Searle speech acts need not involve language: recall the example of my creation of an obligation for myself by pushing a drink towards you. We can now see that such speech acts can create not only obligations, but also rights: if I push a drink towards you, I create a right for you *vis-à-vis* me, namely the right to consume the drink without interference. Interestingly, whereas above I emphasised the role of status terms and their meaning for the agents involved, in simple cases of the creation of statuses through speech acts that do not involve language, it seems that no already-existing status term need exist. For instance, Searle says that we can create the status "leader" without anybody using the *term* "leader": if we all accept that you have certain rights, then this constitutes you as having the rights characteristic of the status "leader", even if this term or a synonym doesn't arise for us until long after the fact.[93] Of course, this still requires agents to have a conception of obligations and rights, for they must be able to intentionally impose the relevant meaning onto their act of acceptance. I discuss this below. The introduction of the idea that speech acts of committing acceptance can be non-linguistic raises significant complications, discussed in chapter 4.

Nevertheless, it seems that cases in which this kind of process occurs without language are limited in complexity by the kinds of rational commitment an agent can sensibly take themselves to be communicating the creation of by means of non-linguistic speech acts. With language, agents can create statuses of significantly higher complexity.

This account brings out several interesting features of statuses, which Searle makes much of in his discussion.

First, statuses are "ontologically subjective": they depend for their existence upon the intentions-in-actions partly constitutive of the speech acts that create them. And this ontological subjectivity co-exists with the possibility that epistemically objective statements can be made about them: it is not a matter of anybody's opinion that I am a student; rather, it is a matter of epistemically objective fact that each relevant person has rationally committed themselves to performing the acts necessary for me to have this status (even if there is an epistemic difficulty concerning how one might know all the relevant facts).

---

[93] Searle, 19–20.

Second, statuses of a certain complexity depend upon agents having a shared language: as we have seen, the creation of statuses through agents' speech acts of committing acceptance involving status terms requires that each agent is using *the same* status term, i.e. a status term that represents the same set of deontic powers. For instance, in the teacher example above, each person used the status term "teacher" to represent the same set of deontic powers, and it is for this reason that their respective speech acts created obligations that together constituted the relevant rights. If each agent uses the same *word*, "teacher", but each takes this word to represent different sets of deontic powers, then even if they each accept that the same person is "a teacher" this will not create the rights that any of them associate with that term. Further, all statuses depend upon agents having a shared understanding of obligations, rights, and rational commitments generally, and this is true even where agents rationally commit themselves non-linguistically. Nevertheless, at a more fundamental level Searle believes that agents require language in order to represent deontic powers such as obligations at all. That is, we need to have encountered the idea of an obligation as consisting in rational commitments to act in order that we can represent these in such a way as to rationally commit ourselves.

> To get to the point that you can recognize an obligation as an obligation, you have to have the concept of an obligation, because you have to be able to represent something as an obligation, that is, something that gives you a reason for action independent of your inclinations and desires. You need not have the actual word "obligation" or some synonym, but you must have a conceptual apparatus rich enough to represent deontology.[94]

Because obligations are not objects that can be perceived in the world, we can encounter this idea only through language.[95]

Finally, Searle emphasises in his most recent work that statuses necessarily involve "speech acts that have the same logical form as Declarations".[96] Declarations, as we saw above, are speech acts that make something the case by representing that it is the case. Thus, for example, the speech act of promising is characteristically carried out by means of the Declarative speech act "I promise": I make it the case that I promise (creating obligations for myself) by representing in speech the fact that I

---

[94] Searle, 95–96.
[95] Searle, *The Construction of Social Reality*, 69–70, 77.
[96] Searle, *Making the Social World*, 12–13, 85–86, 96.

promise. With the argument I have outlined above, we can see the parallel in the case of the creation of statuses: *we* make it the case that a person or object has a status by way of each representing them as having this status in speech acts using status terms. Searle's caveat that these have "the same logical form" as Declarations is included because, as we have seen, these speech acts need not be explicit "Declarative speech acts" such as when an agent says "I promise to treat agent x as having status y": rather, they can be speech acts that rationally commit the agents to the relevant deontic powers only implicitly, as in the examples used above. Given that, as I argued above, Searle often uses the same terms to refer to both statuses and status terms, this feature of statuses as created through representations such that they have the "logical form" of Declarations often leads to statements that sound somewhat circular:

> [An object or person] has to be represented as having the status function it has; otherwise it cannot have that status function[97]

However, in other passages Searle is more careful to differentiate between statuses and status terms, and here the apparent circularity disappears:

> there is a special relation between the imposition of these status-functions and language. The labels that are a part of the Y expression, such as the label "money," are now partly constitutive of the fact created. Odd as it may sound, in the creation of money, the linguistically expressed concepts, such as "money," are now parts of the very facts we have created.[98]

While, then, there is no circularity in Searle's account, there is still the fact that in practice it would be unlikely for us each to have matching conceptions of a particular status term, defined for each of us in terms of a complex bundle of deontic powers, without the status that this status term represents already existing. As such, it seems that the account of the creation of complex statuses outlined above is best suited to an account of how statuses are maintained and *re*created, rather than to origin stories.

There remains one important feature of Searle's discussion of statuses that we have not yet touched upon, namely the requirement that the acceptance of the agents involved is carried out in conditions of *common knowledge*. I do not think that Searle

---

[97] Searle, 111.
[98] Searle, *The Construction of Social Reality*, 51.

gives sufficient justification for the centrality of this feature. I will suggest an important role for it in the following section.

Searle's account covers much of what I have reconstructed above relatively quickly, but adds considerable complexity onto this framework. In particular, his account most famously includes a discussion of what he calls "constitutive rules" of the form "x counts as *y* in *c*".[99] We need not go into detail in explicating this notion, but it is worth sketching how these fit into his overall account. They play two roles in his account.

First, the meaning of status terms as described above is determined by the constitutive rules of language. The shared use of these rules in language means that we can perform speech acts in a publicly accessible way.[100]

Second, in the context of statuses the central point of Searle's discussion of constitutive rules is that I can perform acts of committing acceptance not only towards token objects having a status, but also towards *classes* of objects (*x*'s) having a status (*y*) in a particular context (*c*). In doing so, I rationally commit myself to the obligations constitutive of token objects having a status *indirectly*. For instance, I might commit myself to the obligations that are partly constitutive of you being my leader not by saying "you are my leader", but by saying that "the oldest living child of the deceased leader (*x*) is my leader (*y*)": if you satisfy the conditions specified by the *x* term, then my rational commitment to this rule logically entails—and thus rationally commits me to—treating you as my leader, and thus commits me to the deontic powers this entails. A further layer of complexity is then added with the term "institution", which Searle sometimes describes as referring to a *set* or "system" of constitutive rules.[101] For instance, we might talk of the institution of the monarchy consisting in the constitutive rules that stipulate what conditions an agent must meet to have the status "monarch", what conditions an agent must meet to have the status "citizen", and so on. On this definition of an institution, then, the acceptance of institutions entails deontic powers just as acceptance of constitutive rules and statuses do: an accepted status consists in deontic powers; accepted constitutive rules entails the acceptance of token statuses; acceptance of an institution entails the acceptance of the relevant constitutive rules. This sense of "institution" is thus

---

[99] Searle, *Making the Social World*, 13–14.
[100] Searle, *Rationality in Action*, 186; Searle, *Speech Acts*, 33–42.
[101] Searle, *Making the Social World*, 10.

crucially distinct from the sense in which I will talk of "social institutions" in the following chapter: there, we will return to the relation between, on the one hand, statuses, constitutive rules, and institutions in this sense and, on the other, social institutions.

The key point to emphasise from this account is that an object or person has a status—a status *exists*—insofar as the relevant deontic powers actually obtain, i.e. insofar as people have each rationally committed themselves in the appropriate way. To say that a status exists is thus to make a claim about the obligations—and thus the rational commitments—of various agents. Agents' rational commitment to act in accordance with the deontic powers definitive of statuses is therefore entailed *conceptually* by the claim that a particular status—in this sense—exists. Those interpreters of Searle who find in his account a lack of explanation for how statuses can rationally commit agents and give them desire-independent reasons for action are thus looking for an explanation that would be tautological according to Searle's definition of his terms.[102] Once again, the key question seems to be not how statuses rationally commit us, but rather why we would ever rationally commit ourselves in such a way as to constitute them. Of course, matters are complicated significantly in the case of statuses, first because there are other importantly distinct senses in which we often talk of statuses "existing" or of objects "having" a status, and second because we often talk of statuses existing amongst a group in which (only) a *significant proportion* of group-members accept the status, leaving us with the question of what relation those minority of group members who do not so accept the status have to it. Both problems will be discussed in the following chapter: for now, it is important to keep in mind that what has been said above pertains only to statuses and deontic powers insofar as these relate to those agents that actually accept them in the sense described above.

---

[102] For such claims, see Hindriks, 'The New Role of the Constitutive Rule', 185; Hindriks, 'Collective Acceptance and the Is-Ought Argument', 473–74; Gilbert, 'Searle and Collective Intentions', 44; Stahl, 'Institutional Power, Collective Acceptance, and Recognition', 358–59.

3.4 Explaining collective acceptance

So far, I have given an account of *how* agents might coordinate their creation of obligations such as to constitute positive deontic powers such as rights by means of status terms. Our next question is: *why* might they do so?

Searle does not attempt to give a comprehensive account of the kinds of reasons for which agents might accept statuses. Indeed, he explicitly says in both books on the topic that, given the variety of statuses we create and interact with in social life, such a comprehensive account is not possible.[103]

However, one thing seems clear: Searle does not take agents to rationally commit themselves to statuses willy nilly. We can, I think, consider one important idea—one which Searle does not explicitly consider but which I take to be implicit in some of his discussion—that might contribute to an understanding of why rational agents might accept statuses in the sense outlined above. This is that it is characteristic of statuses that agents accept them at least partly *because others do too*. This, I take it, is implied by my claim above that statuses are primarily a means by which agents coordinate their creation of obligations.

In order to spell this idea out a little more systematically, I would like to introduce the idea of the *social grounding* of speech acts of committing acceptance. By the "grounds" of an action I mean those considerations which justify that action in the mind of the agent whose act it is.[104] In this sense, it describes the relation between an act and the agent's motivating reason for performing it. Take, for example, my previous discussion of Searle's moral obligation to help those in desperate need whom he is able to help. In this example, I posited that his motivation for creating this obligation might be that he believes that others' suffering is a bad thing: as such, this belief *grounded* his acceptance of the obligation.

The belief that grounded Searle's moral obligation in the above example, of course, makes reference in its propositional content to other people, namely to those in need. Nevertheless, the acceptance motivated by such reasons is not socially grounded in the sense I would like to develop. Rather, socially grounded acceptance is that which

---

[103] Searle, *The Construction of Social Reality*, 92; Searle, *Making the Social World*, 107.
[104] See Brennan et al., *Explaining Norms*, 58. This conception of social grounding draws on Brennan et al's account of the practice-dependence of 'social norms'. I take the difference between our accounts to result from those between statuses and social norms; see Brennan et al., 66–72.

is grounded in considerations concerning others' acceptance of the status. Take, for instance, the obligations a shopkeeper takes themself to have *vis-à-vis* money. We might spell this out as something like: the shopkeeper has committed themself to the act of taking as payment for their goods (only) those pieces of paper that satisfy the relevant conditions for having the status "money". That many such shopkeepers, as well as bank tellers etc., have so committed themselves constitutes the rights that I have by virtue of owning money. Now, why would the shopkeeper ever commit themself to such a strange set of acts? We might spell out their reasoning in caricatured form as follows: they believe that others possess items which they themself desire to have; and they believe that these others will accept (only) these particular bits of paper in exchange for these items. The shopkeeper therefore desires these bits of paper as a means to getting that which they desire, and is motivated by this to rationally commit themself to (only) taking these bits of paper as payment for their own goods. If this was roughly the shopkeeper's reasoning, their acceptance would be a socially grounded attitude, for while the description of the acts to which the shopkeeper has rationally committed themself makes no reference to others' acceptance (they have simply committed themself to accepting certain bits of paper and not others), this acceptance was grounded in a belief that does.

Now, recall that above I briefly described Searle's classic account of the collective Intentionality of a group, according to which this consists in members of a group having Intentional attitudes that they attribute to the group rather than to themselves in isolation. On this account, agents have attitudes—I'll use the example of intentions, which is Searle's primary focus—of the form "we-intend". One example he uses is that of two musicians playing a duet together: in such a situation, it is said not to be the case that each has an intention-in-action of the form "I intend that I will play my part on the condition that my partner plays theirs". Rather, each musician has an intention-in-action of the form "we intend to play the duet". Two important details must then be added to the account. One potential problem is that there seems to be something strange about the idea that a collectively-attributed intention could ever motivate *me*, as an individual, to act. To counter this concern, Searle suggests that such collectively-attributed intentions are systematically related to intentions to do one's component part. This relation is said to be such that I we-intend to do something by way of or by means of my playing my individual part. A second potential problem is that of how an agent can take themself to have an intention that covers an entire group, even a group of two as in the duet case. Searle's

response is that an agent can have such an intention only presupposing that the relevant others have the corresponding collectively-attributed intentions. Thus it is not that I intend that we play the duet: rather, I assume that you and I together form a "we" that intends to play the duet, and this entails an assumption that you similarly take yourself to be part of this same "we". Searle therefore spells out the attitude one musician might have when playing a duet as follows: "I have a collective intention-in-action that *we* play the duet by way of *me* playing the piano, *in a context where I take it for granted that you are playing the violin*".[105] Where each member of this group has such a complex collectively-attributed intention with the same collective content (i.e. the same group activity such as "playing the duet": the component parts each individual must play can differ), and where each member's intention of this kind is common knowledge in the group, Searle describes the members as *cooperating*, and we can sensibly talk of the group having a collective intention.

As I said above, it is not immediately clear what relation Searle takes this account to have to his account of statuses and collective acceptance. As we saw above with the example of money, it is quite possible to spell out the acts of acceptance that together constitute the collective acceptance necessary for a status to exist without including any collectively-attributed intentions-in-action. (And, as we saw, Searle in MSW says explicitly that collective acceptance need not include any such intentions-in-action.[106]) I would like to suggest that the best way to understand Searle's invocation of his classical account of "collective Intentionality" *vis-à-vis* his account of statuses is that he believes that speech acts of committing acceptance of statuses are often collectively-attributed, and that in such cases this collective attribution is a result of the way in which such speech acts are socially grounded. In this way, he takes this collective-attribution to be closely connected to that which is distinctive of the deontic powers constitutive of statuses, namely, that the speech acts of committing acceptance that create them *need to be socially grounded*.

The idea that an agent's complex intention-in-action that partly constitutes an act of committing acceptance can be collectively-attributed seems plausible. Take, for example, a group of casual footballers who decide that they need a referee. They draw straws, and one player—Anna—is selected for the role: Anna thus has the

---

[105] Searle, *Making the Social World*, 52; for discussion of these details, see Searle, 50–55.
[106] Searle, *Making the Social World*, 58.

status "referee" imposed upon her, and each player, by virtue of their acceptance of this status by means of a speech act employing the relevant status term, rationally commits themselves to various actions such as to abide by Anna's decisions. It seems plausible that in such a situation the intentions-in-action partly constitutive of the speech act of acceptance of each player is best expressed by the agent not along the lines of "I accept that Anna is referee as long as the other players do", but instead as simply "we accept that Anna is referee", i.e. it is a we-intention-in-action.

The connection between Searle's classic account of collective Intentionality and his account of statuses, then, is an empirical and not a conceptual one. As we saw with the example of the shopkeeper, collective committing acceptance of the kind necessary to create a status need not consist of collectively-attributed speech acts. However, it seems (as a matter of empirical rather than conceptual fact) that such speech acts by different agents are unlikely to be sufficiently coordinated so as to create a status unless each individual's act is grounded in the knowledge that others have performed, will perform or are performing a speech act of the same kind, i.e. unless each speech act is socially grounded. When such knowledge is present, agents (again, empirically and not conceptually) are likely to think of themselves as acting together, such that the intentions-in-action of their speech acts will often be we-intentions. Indeed, Searle seems to think that such we-intentions epitomise the kinds of speech acts that create statuses.

It thus seems that Searle's classic account of collective Intentionality applies to many cases of collective committing acceptance that creates statuses, which explains why Searle recapitulates this account in his books on social ontology. But it is not a conceptually necessary part of his account of collective committing acceptance (as he recognises explicitly in MSW). Rather, because in practice the individual acts of committing acceptance that create obligations need to be socially grounded if these are going to cohere with others' creations of obligations to constitute statuses, the intentions-in-action involved in the creation of statuses will often be we-intentions, and we can therefore often speak of collective Intentionality in Searle's classic sense in such contexts.

Statuses' characteristic quality of existing by virtue of speech acts of committing acceptance that are socially grounded explains Searle's inclusion of the requirement for mutual belief in all kinds of collective acceptance, i.e. the claim that when there is collective acceptance of a status we must (as a matter of empirical fact) each

believe that the others accept the status, and believe that the others believe that we each accept the status, and so on.[107] It perhaps also partly explains why the kind of rational commitments with which Searle is concerned are so closely connected to speech acts: whereas in my own practical reasoning I can get by without obligations—I can intend to do something tomorrow and this will give me a reason tomorrow to do that thing—we tend to create obligations for ourselves in social contexts and for reasons that involve others. As we saw above, in the example of promising I commit myself to an obligation by means of a speech act which also communicates this creation of commitment to others, and the fact that this form of acceptance of the relevant obligations is socially grounded perhaps explains why this is the case: I usually promise others on the understanding that doing so will have some sort of social effect, whether this is in the form of encouraging reciprocal promises or some other effect. And it seems that this is usually the case also for the acceptance of statuses.[108]

While, then, other forms of self-committing, such as promising, are also often socially grounded, it is the social grounding of the collective committing acceptance of obligations *via* status terms that differentiates obligations so created in an important respect from the kind of "moral obligations" discussed previously such as Searle's obligation to help those in need.

Taking a step back, we can see that Searle's account of statuses resembles a rather elaborate social contract theory of statuses: statuses such as President exist only because all of those who are subject to Presidential deontic powers have accepted this status through speech acts.[109] Indeed, it seems to resemble an "actual contract theory" such as Hobbes's or Locke's, although with several important differences.[110] First, his account replaces the notion of promising in Hobbes or Locke with the notion of committing acceptance, which is significantly broader. And second, Searle implicitly introduces what we might think of as a Humean emphasis on others'

---

[107] Searle, 49, 57.

[108] On the importance of the communication of our created commitments, see Searle, *The Construction of Social Reality*, 77; Searle, *Rationality in Action*, 179, 186, 203.

[109] Searle, *Making the Social World*, 165; Searle, 'Language and Social Ontology', 443–44 and *passim*. See also Leezenberg, 'The Contract of Language'.

[110] See in particular Hobbes, *Leviathan*, 114; Locke, *The Second Treatise of Government*, chap. 8.

beliefs:[111] for Searle, agents accept a status not on the *condition* that others do so, but motivated by a *belief* that they do, i.e. their acceptance is socially grounded.

As reconstructed so far, however, it differs from most social contract theories insofar as its subject matter is not social institutions such as the state, but rather statuses understood as bundles of obligations, with obligations understood as rational commitments that agents have actually created for themselves.[112] Given Searle's definitions of these terms, something like a social contract theory of obligations and statuses seems to follow more or less analytically: if statuses are defined in terms of rights and obligations, and rights as consisting in obligations, and obligations as created by means of speech acts of committing acceptance, then of course statuses are created by means of speech acts of committing acceptance. We will see in the following chapter how this account relates to the more traditional subject-matter of social contract theories—that is, social institutions—and will see that in that context Searle's resemblance to traditional social contract theories is less justifiable.

---

[111] See chapter 4 section 1.1

[112] As Hobbes put it, 'And covenants, without the sword, are but words, and of no strength to secure a man at all.' Hobbes, *Leviathan*, 111.

Chapter 4
Understanding social institutions and their alienation

In chapter 3, I focused on the topic that I take to be central to Searle's work on social ontology, namely *statuses*. These were said to be the subject-matter of institutional facts such as that Donald Trump is the President of the United States and that I am a doctoral student at the University of Sussex, and on Searle's account consist in bundles of deontic powers such as obligations and rights. Such deontic powers were seen to be constituted by the rational commitments arising from agents' speech acts of committing acceptance.

In the introduction to part 2 we encountered a characterisation of Searle's work as *idealistic* put forward by Brian Epstein, for whom Searle's position is that "We, as a community, make the social world by thinking of it in a particular way".[1] It would seem, then, that, given a certain way of understanding what it means to "think" of the social world, Epstein's portrayal of Searle as concerned with only agents' thoughts and not their actions is partly plausible *insofar as* his object of analysis is statuses. While for Searle the obligations that constitute statuses require for their creation speech *acts*, such that even here Epstein's portrayal seems wide of the mark, it is true that for Searle statuses consist in rational commitments, not acts. We can thus construct examples in which agents have the relevant deontic powers and thus statuses as a result of speech acts performed in the past, in which the status exists without any relevant activity happening here and now. Further, as we have seen Searle's focus in analysing the speech acts that rationally commit agents is primarily on the role played by the intentions-in-action that partly constitute these acts, and to that extent his examination emphasises the role of the "ideal" aspect of these. However, not only does this portrayal ignore the important role played by *acts* in the creation of statuses, it further ignores the fact that Searle does not take the social world to consist only of such statuses.

---

[1] Epstein, 'Précis of The Ant Trap', 128.

For Searle, the social world has at its very heart the action and behaviour of agents: "to understand society, you have to understand collective human behavior".[2] While for Searle statuses imposed on people and objects are crucial for understanding this collective behaviour,

> the whole point of the creation of institutional reality is not to invest objects or people with some special status valuable in itself but to create and regulate power relationships between people. Human social reality is not just about people and objects, it is about people's activities and about the power relations that not only govern but constitute those activities.[3]

As such, the Epsteinean criticism seems to mistake Searle's frequent emphasis on statuses for an implicit claim that statuses are all that the social world consists in.

Searle's analysis of statuses, then, is part of a larger attempt to understand that which motivates agents' behaviour. If Searle is not guilty of the kind of idealism of which Epstein accuses him, he is however guilty of a number of other transgressions. First, what has been said above should make clear that, on his own terms, Searle's account is crucially incomplete. While acknowledging its importance, he pays insufficient attention to the role of activity in social life and the role of agents' beliefs concerning others' activity and motivations. His account of the relationship between statuses and the other aspects of the social world is thus somewhat simplistic. Second, that so many writers join Epstein in reading Searle as idealistic suggests that Searle is not sufficiently clear about what his explanandum is. In particular, his account employs several terms—including such central terms for his account as "power" and "function"—in crucially ambiguous ways, sometimes referring to aspects of statuses, sometimes to phenomena that consist in agents' actual activity. These ambiguities can easily give the impression that he is claiming more for his account than he actually intends to deliver, that his account of the ontology of *statuses* is in fact an account of the ontology of what I will call *social institutions*.

This chapter aims, first, to arrive at a clear definition of social institutions, and then to use this definition in translating the concept of the alienation of social institutions to the level of individual agents and their relations. In section 1 I will introduce the notion of a social institution and suggest a provisional definition, drawing on Searle's account. In section 2, I will examine a crucial ambiguity in Searle's discussion of

---

[2] Searle, *Making the Social World*, 26.
[3] Searle, 106.

social institutions, one which we must resolve if we are to be clear about what social institutions are. This ambiguity concerns the ways that status-representing attitudes motivate the activity that constitutes social institutions, and corresponds to an ambiguity in his definition of a social institution. In section 3, I will employ our clarified definition of social institutions to translate the second stage of the process of alienation described in chapter 2 to the level of individual agents and their relations, in which social institutions are said to become independent of the activity of those agents whose activity constitutes them.

1 Activity and social institutions

In understanding social life, we often want to talk not only of statuses but also about how agents actually act. In asking what it means to say that I am a student, we might ask not only what rights and obligations this entails, but whether people—both myself and those against whom I have rights, such as the librarians or those staffing the IT helpdesk—*actually act* according to such rights and obligations. Do these people actually let me into the library and fix my computer?

When analysing statuses and positive deontic powers such as rights in chapter 3, we saw that these were phenomena that existed at a macro level of description: they consisted in the obligations of several agents. Analogously, there are crucial macro-level features of agents' activity without an account of which our view of social reality would be incomplete, although as before our aim remains to translate these to the level of description of individual agents and their relations.

One way in which we commonly refer to such macro-level features is in terms of *social institutions*. When we invoke the existence of social institutions in daily life, or indeed in non-Searlean social theory more generally, I take it that part of what we are describing is that people—not just an individual, but the members of a community or some subset of these—actually act in particular ways. Of course, such talk implies more than this: social institutions differ in important ways from other macro-level phenomena constituted by social activity such as conventions, widespread conformity to norms, and so on. However, that social institutions, like these other phenomena, are constituted by activity seems clear: in this sense, they are in stark contrast with statuses.

Searle acknowledges this way of talking of social institutions in some places. For instance, he says that "social institutions such as governments and corporations are dependent on and derived from the mental phenomena *and behavior* of individual human beings" (my emphasis).[4] Thus, to characterise social institutions seems to require a focus *both* on agents' actual activity *and* their motivations. Further, I take it that Searle's analysis suggests an interesting way in which we might differentiate "social institutions" from the other macro-level phenomena mentioned above, namely that they intrinsically involve—in some sense to be determined—statuses. While I differ in detail with him on the role of statuses in social institutions, I will nevertheless follow Searle in this general insight.

As mentioned in chapter 3, Searle also uses the term "institution" in an importantly different way, namely to refer to sets of constitutive rules.[5] Given his account of constitutive rules (see chapter 3 section 3.3), this use of the term suggests that institutions are ultimately constituted by obligations and thus, like statuses, are not constituted by activity. It thereby encourages the Epsteinean interpretation.

In what follows, then, I will take it that a social institution consists in a pattern of activities amongst a group of people that intrinsically involves (in a way yet to be specified) statuses.

1.1 Defining social institutions

In this section, I will suggest an initial definition of social institutions drawn from Searle's discussions, one which I take it he would broadly agree with.

A social institution is a pattern of activity amongst the members of a group or community: thus, to say that a social institution exists is to say that (a proportion of) group members are actually acting in a particular way. However, not any pattern of activity constitutes a social institution: the activity must be activity of a particular type and which is motivated in a particular way. Our central question, then, involves how to characterise the activity and motivation that is peculiar to social institutions. (I will not here treat the question of what proportion of group members must be so acting. Searle talks notably vaguely about what this proportion is: for instance, he

---

[4] Searle, 4.
[5] Searle, 10, 23.

writes that the existence of social institutions requires that "the individuals directly involved and a sufficient number of members of the relevant community" are acting appropriately.[6] I will return to this question in section 3 below.)

I take Searle's central claim to be that social institutions in my sense intrinsically involve statuses in the following broad way: a social institution consists in the activity of a group in which each agent a) acts as the deontic powers constitutive of a status or set of interrelated statuses would commit them to act were the represented status to exist, and b) does so motivated in some way by an attitude that represents the relevant status or the deontic powers constitutive of it. It is worth highlighting that on this definition agents are said to act as the deontic powers constitutive of a status would commit them to act if the represented status actually existed: as such, to say that a set of agents so act, and thus that a social institution exists, does not require that the status represented by the status term nor its constitutive deontic powers do in fact exist (in the sense analysed in chapter 3). In what follows I will therefore talk of agents acting in accordance with the *perceived* deontic powers of a *perceived* status represented by a status term. By a perceived status for a given agent, I mean a status that the agent believes that others in the relevant community believe that they and the agent are rationally committed to acting in accordance with. As such, some person or thing can have a perceived status for a given agent regardless of whether or not they actually have that status in the sense described in chapter 3: only if the agent and others in the community have each rationally committed themselves by way of a relevant speech act of committing acceptance does the object that has the perceived status for the agent also actually have that status. Similarly with individual deontic powers: I have a perceived obligation to x if I take it that others believe I am rationally committed to x-ing; I also have an actual obligation if I am so committed. The significance of this distinction will become apparent in the argument that follows.

This definition of a social institution can usefully be contrasted with what we might call a "Humean convention".[7] This is a pattern of behaviour amongst a group in which each agent acts such as to partly constitute the pattern of behaviour in question because they take it to be in their interest to do so assuming that others

---

[6] Searle, *The Construction of Social Reality*, 117.

[7] In using 'convention' in this sense I am drawing on Lewis's usage of the term rather than Hume's own. See Lewis, *Convention*.

also act in the appropriate ways (because it is in their respective interest to do so) and in which this complex of interests is a matter of common knowledge. Hume uses this notion to explain, amongst other things, the social practices constitutive of private property:

> I observe, that it will be for my interest to leave another in the possession of his goods, *provided* he will act in the same manner with regard to me. He is sensible of a like interest in the regulation of his conduct. When this common sense of interest is mutually express'd, and is known to both, it produces a suitable resolution and behaviour. And this may properly enough be call'd a convention or agreement betwixt us, tho' without the interposition of a promise; since the actions of each of us have a reference to those of the other, and are perform'd upon the supposition, that something is to be perform'd on the other part. Two men, who pull the oars of a boat, do it by an agreement or convention, tho' they have never given promises to each other. Nor is the rule concerning the stability of possession the less deriv'd from human conventions, that it arises gradually, and acquires force by a slow progression, and by our repeated experience of the inconveniencies of transgressing it. On the contrary, this experience assures us still more, that the sense of interest has become common to all our fellows, and gives us a confidence of the future regularity of their conduct: And 'tis only on the expectation of this, that our moderation and abstinence are founded.[8]

On this account, there is no suggestion of anything resembling deontic powers. Indeed, Hume implicitly excludes such notions by emphasising that conventions do not involve promises. Rather, these phenomena are explained solely with reference to agents' desires and beliefs about other agents' desires, beliefs and behaviours. Conventions in this sense, then, are not social institutions.

Of course, social institutions might build up around conventions. Driving on the left may originally have been simply a convention. But we might now be partly motivated to drive on the left because we accept a set of deontic powers such that one ought not to break the convention. Indeed, most countries have laws reinforcing these conventions, and the law certainly is a social institution. Nevertheless, it is important to keep the concepts of convention and social institution distinct.

Social institutions can further be contrasted with widespread activity in accordance with *norms*. Geoffrey Brennan *et al* have given an account of norms inspired by H.L.A. Hart, according to which norms consist in a significant proportion of the members of a group having certain *normative attitudes* and knowing that a

---

[8] Hume, *A Treatise of Human Nature*, 315.

significant proportion of the group members have such an attitude.[9] There are of course similarities between norms so defined and social institutions. First, both centrally involve normative propositions, i.e. propositions about what agents ought or should do: in social institutions, these are encapsulated in status terms; in norms these are in the content of normative attitudes. Second, insofar as norms often motivate agents to act in accordance with these normative attitudes,[10] both often involve agents acting in accordance with these normative propositions.

However, there are important differences between social institutions and widespread activity in accordance with norms. In particular, while social institutions involve activity in accordance with normative propositions, these must be normative propositions *of a particular kind*, namely those encapsulated in a status term. In contrast, the normative propositions that are the content of the shared normative attitudes constitutive of norms are not so limited. We might, for instance, talk of a norm in a community according to which one ought to help those in need. In such a situation we would not talk of a social institution existing. While we might formulate *post hoc* statuses defined with reference to these normative propositions—for instance, we might develop the status term "help-needer"—these do not play any basic role in agents' motivation: they are motivated not by any attitude representing the status, but by the normative attitude directly. Insofar as their motivation includes reference to a status, this reference adds nothing substantive to the picture, but summarises a set of already existing normative attitudes.[11]

2 Two definitions of social institutions

Another possible difference between social institutions and widespread activity in accordance with norms pertains to the role that attitudes representing the relevant normative propositions play in motivating activity and the kind of attitudes these are. On the definition of norms discussed above, agents must *have* "normative attitudes". We now turn to the different possible ways in which status-representing attitudes might motivate agents to act in accordance with them so as to constitute social institutions.

---

[9] Brennan et al., *Explaining Norms*, 28–35.
[10] See Brennan et al., 15.
[11] This discussion draws from Brennan et al., 69–70.

There are two distinct claims that Searle might be read as making regarding the way that status-representing attitudes motivate the activity that constitutes social institutions, which correspond respectively to a broader and a narrower definition of social institutions. On the broader definition, the relevant activity is motivated by any of a range of attitudes that represent a perceived status or its constitutive deontic powers. This broader definition can perhaps be read as underlying the following passage:

> It is common, for example, to read that certain ant colonies have slaves or that beehives have queens.… [B]ut it is important to keep reminding ourselves that for a community literally to have slaves or literally to have a queen, the participants would have to have the apparatus necessary to represent something as a queen or as a slave. Just behaving in certain ways, where behavior is construed solely in terms of bodily movements, is not sufficient for a community to have a queen or to have slaves. In addition, there would have to be a certain set of attitudes, beliefs, etc., on the part of the members of the community, and this would seem to require a system of representation such as language.[12]

On the narrower definition, the relevant activity is always motivated by the rational recognition of obligations to act in accordance with a perceived status. This might be read as underlying his discussion of the emergence of a social institution of a boundary amongst a tribe:

> If our imagined tribe just is not disposed to cross the boundaries as a matter of inclination, they do not in our sense have an institutional fact. They simply have a disposition to behave in certain ways, and their behavior is just like the case of animals marking the limits of their territory. There is nothing deontic about such markings. The animals simply behave in such and such ways, and 'behave' here means they simply move their bodies in specific ways.
> But if we suppose that the members of the tribe recognize that the line of stones creates rights and obligations, that they are *forbidden* to cross the line, that they are *not supposed to* cross it, then we have symbolization.[13]

What are we to make of these two definitions of social institutions?

I take it that the second, narrower definition is intuitively implausible. Certainly, it seems there are many phenomena that we would think of as social institutions but that we act in accordance with not as a result of a feeling that we *ought* do so, but for more prosaic, prudential reasons. And in particular our experience of many of the

---

[12] Searle, *The Construction of Social Reality*, 37.
[13] Searle, 71.

social institutions with which we are primarily concerned—those which Marx suggests are alienated—would seem to be of this kind. Indeed, I can think of few existent social institutions that are constituted entirely by activity motivated by the rational recognition of deontic powers, with games being perhaps the most promising example, although of course many social spaces try to approximate this model as closely as possible.

Further, it seems that Searle recognises many of these other kinds of motivations, and builds them into his account. For instance, as we will see below, he discusses the kinds of motivations for which people acted in accordance with the perceived deontic powers involved in the social institutions of the Third Reich, which includes the claim that many people simply "went along with it as a matter of nationalism, indifference, prudence, or even just apathy".[14]

Why, then, does Searle elsewhere seem to suggest, as the narrow definition has it, that social institutions are constituted exclusively by activity motivated by attitudes of rational recognition of obligations to act in accordance with perceived statuses? In section 2.1 and 2.2 I will suggest some explanations for this. In 2.3 I will clarify the broader definition which I favour.

2.1 "Power", "function", and "acceptance or recognition"

The first explanation for Searle's occasional endorsement of the narrow definition of social institutions is that, as a result of several ambiguities in his account, the distinction seems less significant to him than it in fact is. It will be useful to unpick some of these ambiguities.

One problem we face in clearing up the relationship between statuses and social institutions is that facts concerning statuses and facts concerning social institutions often overlap considerably, to such an extent that we often use the same expressions to refer to both together. For instance, the statement "Anna is our referee" might express two quite different things: that we each have committingly accepted that Anna has the right to make decisions in our game, and so are each rationally committed to acting accordingly (whether or not we *actually* do what she says); or that Anna tends to make all the decisions in our game, and we tend to go along with

---

[14] Searle, *Making the Social World*, 57.

her decisions.[15] This distinction mirrors the common one employed with respect to the law between *de jure* and *de facto*: on the first meaning, we might say, Anna is our *de jure* referee; on the latter, she is our *de facto* referee. We might, of course, often make this statement with both ideas in mind: that Anna is our referee in the sense that we have each committingly accepted that she has the right to make decisions *and therefore* do what she says. Nevertheless, it is crucial to keep the two distinct senses of such claims apart, for they each entail importantly different things.

Searle's discussion employs a number of statements that have this dual meaning, most significantly statements employing the terms "power", "function", and "acceptance or recognition". These ambiguities cause considerable problems not only for interpreting Searle's work, but more significantly for his account of the relationship between statuses and social institutions.

First, as we have seen the concept of *deontic powers* is central to Searle's account of statuses. But Searle often also employs the term "power" to refer to what we might call *actual power*: the power *qua* ability or capacity to actually do things or to get others to do things.[16] This distinction can be spelled out with reference to Anna the referee: Anna might have the deontic powers constitutive of the status "referee" (regardless of whether or not people actually do as she says); or she might have the actual power to get people to do as she says (regardless of whether she has the deontic powers). Again, there are of course many situations in which somebody has an actual power *by virtue of* having a deontic power, as when we have committingly accepted that Anna has the right to make decisions and therefore do what she says. Nevertheless, the two senses of power must not be conflated.

Second, and related to the dual uses of "power", Searle's term "function" has a dual reference. As we saw in chapter 3, Searle often describes statuses as having some function, and his language of referring to statuses and their functions both with the term "status function" confused matters. The "function" of a status was said to be to "carry" the deontic powers constitutive of it, i.e. to unify a set of deontic powers around a single person or thing. Here, however, we encounter yet another sense in which Searle uses the term "function", namely to refer to the purposes[17] for which

---

[15] Of course, for this to be a social institution, this activity must be motivated by attitudes that in some way represent Anna as having the relevant perceived status.

[16] See especially Searle, *Making the Social World*, chap. 7.

[17] Searle, 59.

we develop statuses: in this sense the function of the status "boundary" is said to be to regulate inhabitants' and outsiders' movements in and out of a territory,[18] while that of money is said to be to serve as a medium of exchange.[19] These "functions" are described in terms of the purpose for which agents created a status, but crucially these purposes are *satisfied* not by the creation of statuses alone—as we have seen, it is possible to say that Anna has the status "referee" without anyone actually doing what she says and thus without the purpose for which we created this status being satisfied—but by agents actually acting accordingly. For this reason, we might instead refer to the purpose for which a status is created as the function of the relevant social institution rather than the function of the status.

Third, and most crucially, is the notion of individual or collective "acceptance or recognition". In contrast to these terms, Searle's uses of "power" and "function" seem relatively unambiguous, for the terms "acceptance" and "recognition" have considerably more referents in Searle's writing.[20]

Within Searle's account of statuses, these terms have several meanings. First, as we saw in chapter 3, is that of *committing acceptance*, i.e. the speech act by means of which agents create obligations for themselves. Second, once we turn our attention to such an obligation's role in motivating me to act, another sense arises: in order to be motivated by an obligation so created, I need to recognise or accept (I will use "recognise" here, but Searle makes no systematic distinction between the two terms) that the situation is such that my obligation is relevant. For instance, if I am rationally committed to act according to the deontic powers constitutive of a status that some person or object has, then I have to recognise that the person or object in front of me is that which has that status: that it is Anna and not her twin sister that is telling me I was offside. Sometimes the way that I recognise that this object has a status (y) is by having first recognised a constitutive rule that any x in circumstances c has this status.[21] Another feature of the situation that we might need to recognise in this sense derives from the temporal element that is central to Searle's account of deontic powers: once I have committed myself to some future act, I then—in the future—need to recognise that the situation is one in which this act is called for.[22] If

[18] Searle, *The Construction of Social Reality*, 39.
[19] Searle, 98.
[20] cf. Laitinen, 'Recognition, Acknowledgement, and Acceptance'.
[21] Searle, *Making the Social World*, 181–82.
[22] Searle, 131.

we have a rota for refereeing duties, and as part of this I have committed myself to treating Anna as the referee when we play football on the third Saturday of every month, then when Anna starts to adjudicate my behaviour I need to recognise both that we are playing football and that it is the third Saturday of the month, and thus that the situation is such that I ought to abide by Anna's decision. Let us call acceptance or recognition of this second broad kind *normative acceptance*. I take it that Searle has in mind attitudes of normative acceptance when he speaks of "recognitional rationality"[23] in these contexts (see chapter 3 section 2.2).

These uses of "acceptance" and "recognition" are internal to Searle's account of how statuses are created and how we come to be motivated to act in accordance with their constitutive deontic powers: while his use of the same terms for each of these seems unnecessarily confusing, it need cause us no concern. However, there is yet another sense in which Searle employs these terms, one that refers not to statuses but to social institutions.

Agents can have attitudes concerning not only statuses, but also concerning the existence of social institutions. I can, for instance, believe not only that Anna is our referee in that she has the right to make certain decisions, but can believe that all the other players believe that Anna is our referee and therefore actually follow Anna's decisions. Indeed, I might believe that Anna and these other players would likely rebuke or sanction me for not abiding by her decisions. In this sense, I can recognise or accept that a social institution exists amongst this group according to which players treat Anna as referee. Let us call attitudes of this kind *descriptive acceptance*. Committing acceptance is a speech-act. By contrast, both normative acceptance and descriptive acceptance are attitudes.[24]

Crucially, the facts that attitudes of descriptive acceptance describe *can themselves be reasons for me to act*. First, given what was said in chapter 3 about the social grounding of acts of committing acceptance of statuses, such facts can be reasons for me to committingly accept a status if I interpret others' activity as constituting their own committing acceptance. In this sense, such facts can ground my creation of an obligation to act as Anna says: my descriptive acceptance partly motivates my creation of desire-independent reasons. Second, facts concerning others' activity *vis-*

---

[23] Searle, *Rationality in Action*, 117–19. Searle, *Making the Social World*, 128–31.

[24] Given our current emphasis on motivating reasons, descriptive acceptance is here understood non-factively: I can be motivated by mistaken descriptive acceptance.

à-vis a perceived status can also be desire-*dependent* reasons for me to act according to the deontic powers constitutive of that status. For instance, in one passage in MSW Searle says that "institutional structures" require collective recognition or acceptance, and that:

> 'collective recognition or acceptance' … marks a continuum that goes all the way from enthusiastic endorsement to just going along with the structure. At the time of the Nazi regime, for example, members of the Nazi Party enthusiastically endorsed the institutional structure of the Third Reich. But there were lots of people in Germany at the time, who, while not endorsing the institutional structure, went along with it as a matter of nationalism, indifference, prudence, or even just apathy.[25]

Here, agents are described as acting motivated by descriptive recognition of the existence of an institution because of *prudential*—i.e. desire-dependent—reasons as well as acting motivated by normative recognition. Similarly, in my more prosaic example, we might say that if I recognise that the other players all behave in a certain way with respect to Anna, and I desire not to stand out (be rebuked/sanctioned), this recognition gives me good reason to do likewise.

Let us formalise the above distinctions. Call *motivating* attitudes those attitudes that factor into an agent's decision to act according to a set of perceived deontic powers constitutive of a perceived status such that that agent's activity partly constitutes a social institution (if in a context in which others so act). All attitudes discussed in what follows are motivating attitudes unless otherwise specified. Amongst motivating attitudes are different kinds of attitudes of acceptance. First, we have attitudes of what I have called *normative acceptance*: agents' acceptance or recognition—by means of "recognitional rationality"—of their previously created obligations, i.e. agents' perception that the descriptive features of the situation are such that their prior act of committing acceptance rationally commits them to act according to the perceived deontic powers in question. Second, we have attitudes of *descriptive acceptance*. (We can of course also have attitudes of descriptive acceptance that are not motivating. For instance, in order to protest against a social institution we must first descriptively accept it.) Attitudes of descriptive acceptance can be motivating attitudes through different possible mechanisms. First, they might partly ground acts of committing acceptance, if the activity constitutive of the descriptively recognised social institution is interpreted as constituting the other

---

[25] Searle, *Making the Social World*, 57.

agents' own committing acceptance. Second, they might themselves—if connected to desires such as not to stand out or be sanctioned—directly factor into the decision to act according to the perceived deontic powers. Let us call attitudes of descriptive acceptance of this second kind *prudentially motivating descriptive acceptance*.

In the case of each of the terms "power", "function", and "acceptance", then, Searle alternates between a sense which essentially refers to statuses and at least one alternative sense which does not. This causes serious problems for interpreting his work, for often the context does not make clear which sense he has in mind and therefore what he is claiming. I will not here belabour the point that this lack of systematic differentiation causes many interpretive problems. Rather, I will focus on the effect it has on Searle's implicit definition of social institutions.

It seems that the ambiguities in Searle's use of "acceptance" might explain his occasional invocation of the narrow definition of social institutions described above. Consider the following claims, which are each true according to the Searlean account I have been reconstructing.

> 1. Collective committing acceptance is necessary and sufficient for the creation of a status and its constitutive deontic powers.
>
> 2. Widespread attitudes of *normative or descriptive* acceptance are necessary for the existence of a social institution.
>
> 3. Therefore if a social institution exists, there is widespread *normative or descriptive* acceptance of its perceived deontic powers.

Now consider that if we lose track of the distinctions between the different senses of "acceptance", it seems as if the following is also true:

> 4. if a social institution exists, there is collective *committing* acceptance of its perceived deontic powers.
>
> 5. Therefore those whose activity constitutes a social institution are all rationally committed to act in accordance with those deontic powers.

If we add that for Searle rational agents will, by virtue of recognitional rationality, be motivated to act according to their rational commitments, and that—as specified in the introduction to section 2—we are for the sake of analysis assuming that we are dealing with rational agents, then it seems that a claim very close to the narrow

definition of social institutions follows: if a social institution exists, agents are acting motivated by rational recognition of deontic powers. Indeed, this position is derived from the claim that a social institution exists according to the broader definition, i.e. from the claim that there is widespread attitudes of either normative or descriptive acceptance. As such, the distinction between the two definitions of social institutions collapses under the weight of the ambiguities described above.

2.2 "going along with" and promising

There is a second possible explanation for Searle's invocation of the narrow definition of social institutions. This is that it is rendered plausible by an argument he implicitly makes to the effect that agents' mere acting in accordance with the perceived deontic powers constitutive of a perceived status is sufficient to create rational commitments, and thus to create actual deontic powers and actual statuses. This implicit argument draws on a parallel between activity of this kind and promising. If this argument were successful, the narrow definition of social institutions would seem plausible. In this section I will spell out this implicit argument and show that it is invalid.

Searle talks in several places as if agents' mere acting in accordance with the perceived deontic powers constitutive of a perceived status is sufficient to create actual statuses and deontic powers. Consider, for instance, the following passage:

> Hitler and Stalin, for example, were both constantly obsessed by the need for security. They could never take the acceptance of their system of status functions for granted, as a given part of reality. It had to be constantly maintained by a massive system of rewards and punishments and by terror.[26]

Here, it seems that Searle takes the relevant "system of status functions" to be maintained by activity motivated by descriptive acceptance of the "system of rewards and punishments and by terror". Acts of obedience motivated by terror are portrayed as counting as speech acts that create statuses. Or see this passage:

> Many people simply go along, unreflectively, with social situations in which they find themselves. But this can amount to a form of inauthenticity or even bad faith, *because they are creating desire independent reasons which are rationally binding on them* but which

---

26 Searle, 165.

they might not have created if they had thought about the question. [my emphasis][27]

Here, "going along with" a social institution is portrayed as sufficient to rationally commit agents to the relevant acts and thereby to create deontic powers. As he puts it elsewhere, "participation in human institutions reinforces the deontology".[28] So whereas in chapter 3 we saw that only acts of a very particular kind—that is, speech acts partly constituted by a particular kind of complex intention-in-action—were able to create obligations for agents, here it seems that just about any kind of act in accordance with the perceived deontic powers of a perceived status is able to do so.

It is worth pointing out how conservative this suggestion is. It seems to imply that any agents who do not, say, take to the streets in armed rebellion are thereby under a *rational commitment* to act in accordance with the deontic powers constitutive of a given status, akin to the rational commitment created by promising. It implies that an exhausted anarchist who puts down their placard for a few weeks would have thereby rationally committed themself to continuing to act according to the perceived status against which they had previously been protesting. It is perhaps for this reason that Searle's examples of institutions to which he himself is *not* rationally committed are always located in Australia, for it seems that if they were closer to home then given his view he might commit himself to them quite by accident!

Call all activity in accordance with the perceived deontic powers of a perceived status *institutional acts*. The passages quoted above suggest that all institutional acts lead to rational commitments and thus deontic powers. Whatever reasons motivate agents' original institutional acts, these acts would create obligations for them to continue to so act, and thus to partly constitute the status associated with the social institution. This position looks to be in clear tension with Searle's explicit account of how agents create obligations for themselves. In chapter 3, we saw that only acts of a very particular kind—those he names *speech acts*, which are characterised by being partly constituted by certain kinds of complex intentions-in-action—were sufficient to rationally commit agents to act.

Of course, the claim that agents *can* create obligations for themselves by means of institutional acts seems plausible: institutional acts can also be speech acts as described in chapter 3. But it seems that Searle's argument goes beyond the claim

---

[27] Searle, 132.
[28] Searle, 141.

that agents *can* create obligations in such ways. Rather, it seems to be that *all* institutional acts *necessarily* create such obligations.

I take it to be intuitive that this claim fails: that, for instance, our tired anarchist is not rationally committing themself in any way equivalent to promising. This intuition is perhaps captured in our legal concept of *duress*, and in the philosophical literature in frequent attacks on the Hobbesian and Lockean notion of *tacit consent*, a notion which Searle's comments in the passages with which we are here concerned seem to mirror.[29] Further, it seems that Searle's account of the creation of obligations through speech acts explains precisely *why* this claim fails: while agents *can* create obligations by means of institutional acts by imposing the appropriate meaning onto their act, there are many cases of institutional acts in which agents do not act with the necessary kind of complex intention-in-action, as in the case of the tired anarchist.

If these statements seem in such obvious tension with Searle's basic account of the creation of obligations, how can he sensibly make them? How can he possibly suggest that institutional activity *necessarily* entails obligations and the statuses they constitute, that agents merely "going along with" the perceived deontic powers associated with a social institution rationally commits them to continuing to do so, given the lengths to which he has gone to specify the kinds of complex intentions-in-action necessary for the creation of obligations which are so clearly lacking in such cases?

My suspicion is that Searle has taken the comparison with the promising case too literally, and that this lies behind the statements we are considering here. To see this, consider the following potential objection to Searle's account of promising. As we saw in chapter 3, Searle's account of promising is such that the performance of the utterance "I promise" creates an obligation for an agent because of the rational constraints internal to the speech act that this utterance is a part of, and that these rational constraints derive from the complex intention-in-action involved in which the agent intentionally imposes a meaning, with conditions of satisfaction and direction of fit, onto their utterance and is therefore rationally committed to these imposed conditions of satisfaction. We therefore saw that there is an epistemic

---

[29] See, for instance, Simmons, *Moral Principles and Political Obligations*, 79–100. Simmons' discussion of the distinction between acts which are "signs of consent" and those which "imply consent" (p89-93) broadly parallels my criticism of Searle below.

difficulty when it comes to knowing about others' obligations. The possible objection is that our intuitions tell us that an agent creates an obligation for themself when they say "I promise" *whatever their intention*: a liar's intention to break a promise does not absolve them of their obligation, but simply means that they are unlikely to act accordingly.

Searle responds to this objection as follows. When I promise, I paradigmatically have two connected intentions:

(1) I intentionally create an obligation for myself (I impose conditions of satisfaction of creating an obligation on the performance of the relevant utterance)

(2) I intend that I will act in the future according to the obligation as a result of having created it.[30]

In sincere promises, intention (2) is paradigmatically part of my motivation for performing the relevant speech act: I usually make a promise in order to give myself a reason to act as promised in the future (and to communicate this creation of a reason). However, when I lie or insincerely promise I perform the intentional speech act described in (1) but do not have intention (2). The speech act of insincerely promising makes sense only on the understanding that promises create such obligations: to intentionally say "I promise" in the appropriate conditions *just is* to intentionally create an obligation, and there is nothing else that saying "I promise" can mean in such conditions. As such, intention (1) alone is sufficient to create obligations, and insincere promising is made possible by this fact. Of course, there are situations in which the appropriate conditions are not present: if I say "I promise" while acting on stage, neither of these intentions are present, and I am under no obligation to do as the "promise" implied. But in this situation I do not promise at all, rather than promise insincerely.

Searle's response seems plausible as a response to the objection above about promising. As a result, it seems that the intuition described above is perfectly compatible with Searle's account of promises: when I perform the speech act of promising, I *necessarily* create obligations for myself whatever my intentions concerning how I will actually act in the future.

---

[30] See Searle, *Rationality in Action*, 184–85, 197; Searle, *Speech Acts*, 60–62.

The argument I wish to attribute to Searle is that this response carries over to institutional acts more broadly. If it did so, it would seem that Searle's various suggestions that all institutional acts necessarily create obligations would be plausible: just as saying "I promise" creates obligations whatever my intention, so too would acting in accordance with the perceived deontic powers associated with a social institution (equivalent to (1) above) necessarily create obligations whatever my intention (as in (2)). It is this move that I take Searle to have implicitly made when making the statements considered above.

However, this argument relies on aspects of the promising case that are absent in the case of many institutional acts. As such, while this response is plausible in the case of promising, we are not justified in assuming that this kind of response carries over to institutional acts in general.

As presented above, Searle's response concerning promising centred on the complex intention-in-action partly constitutive of the speech act of promising (described in (1)), in which I intentionally impose the meaning that I am creating an obligation onto my performance of the relevant utterance. However, the content of the intentional act described in (1) is complex, and in order to highlight the important difference between promising and institutional acts generally we must unpick this complexity.

To do so, we must recall a concept introduced only briefly in chapter 3, namely that of *constitutive rules*. As I put it there, constitutive rules are rules governing the imposition of token statuses, such that agents can rationally commit themselves to act according to the deontic powers constitutive of a status by means of committing acceptance of the relevant constitutive rule. For example, if I accept, *qua* committing acceptance, the constitutive rule that metal discs coined by the Royal Mint count as money, then I am rationally committed by this acceptance to act according to the deontic powers constitutive of money *vis-à-vis* any token metal disc coined by the Royal Mint. Since constitutive rules can be the subject of committing acceptance, we can talk of agents being rationally committed to constitutive rules, just as we can talk of them being rationally committed to token statuses and to token acts.

A primary use to which Searle puts the notion of constitutive rules is to explain agents' use of language. Language involves sets of constitutive rules, for instance the rule that utterances of the sounds "snow is white" count as statements to the effect that snow is white in English, or that utterances of the sounds "I promise" count as

promises. We will see below that there is an important difference between the constitutive rules of language and the kinds of constitutive rules mentioned above, relating to how agents come to be rationally committed to them.

With this notion in mind, we are able to unpick Searle's account of promising as follows:

> (p1) I intentionally perform the utterance "I promise".
>
> (p2) I intentionally commit myself or I know I am already rationally committed to the constitutive rule that performing this utterance in this context rationally commits me to the promised act, i.e. creates an obligation.
>
> (p3) Given (p2), in intentionally performing the utterance (p1) I intentionally create an obligation for myself.

These steps (p1)-(p3) together constitute the complex intention-in-action referred to in (1) in our original account above. We might add that in cases of sincere promising, the motivation for performing the speech act characterised in (p1)-(p3) is the creation and communication of a desire-independent reason for myself in the future to act. As such, in paradigmatic cases of sincere promising we also have:

> (p4) I intend that I will act in the future according to this obligation.

This mirrors intention (2) in our original account.

Let us, then, see how we might reconstruct this argument for institutional acts:

> (i1) I intentionally act as a perceived deontic power constitutive of a perceived status specifies.
>
> (i2) I intentionally commit myself or I know I am already rationally committed to the constitutive rule that performing this act rationally commits me to act in accordance with these perceived deontic powers in the future, i.e. creates an obligation.
>
> (i3) Given (i2), in intentionally performing the act (i1) I intentionally create an obligation for myself.

Again we can add that in some cases the intention motivating (i1)-(i3) is the intention to create a desire-independent reason for myself in the future to act, *a la* the intention referred to in (2) above:

(i4) I intend that I will act in the future according to this obligation.

The inclusion in (p2) and (i2) of the fact that the agent either intentionally commits themselves to, or is already rationally committed to, a constitutive rule highlights that there is more to be said about the matter, namely the question of how agents come to be so rationally committed. And it is here that the apparent parallel between the two cases comes to an end.[31] In the case of promising, the prior rational commitment to the relevant constitutive rule in (p2) *is presupposed* because in Searle's analysis we are told to assume that the promising agent is a "competent speaker of the language".[32] Because the constitutive rule described in (p2) is part of the constitutive rules of language, and because we assume that the speaking agent is a competent language user, we assume that the promising agent is rationally committed to this constitutive rule. These rational commitments are part of what it means to be a competent language user, and we do not feel the need to ask why agents might rationally commit themselves to them. Another way to put this is that, as we have seen, the creation of obligations is part of the very meaning of "promising" (according to the constitutive rules of language to which we are rationally committed as competent language users), such that there is nothing else that an agent saying "I promise" in normal conditions can possibly mean. As Searle puts it:

> language is precisely designed to be a self-identifying category of institutional facts. The child is brought up in a culture where she learns to treat the sounds that come out of her own and others' mouths as standing for, or meaning, something or representing something.[33]

Given this presupposition, if we have (p1), (p3) follows directly. This is why it seems reasonable to present intention (1) in our original account without further explication, and to say that agents who promise have necessarily created obligations for themselves. However, for institutional acts this is clearly not the case: we have no reason to presuppose that any agent that acts in accordance with the perceived deontic powers of a social institution either sees themselves as rationally committing

---

[31] Interestingly, Searle emphasises this difference in his distinction between linguistic and nonlinguistic institutional facts. See Searle, *Making the Social World*, 93, 101, 109–15. See also Hindriks, 'The New Role of the Constitutive Rule', 192–97.

[32] Searle, *Making the Social World*, 111.

[33] Searle, *The Construction of Social Reality*, 73.

themself to, or as already rationally committed to, the constitutive rule that so acting creates obligations for them as in (i2). Indeed, we know from experience that we often have intentions like (i1) without any sense that this involves the kind of committing acceptance that might create the rational commitments in (i2), as our exhausted anarchist attests. Indeed, our anarchist might even vote in an election hoping to bring about a governmental breakdown, performing an institutional act with the explicit intention of ending the relevant system of perceived statuses rather than rationally committing themself to them. Searle's implicit parallel between the Declarative speech act of promising and the creation of deontic commitments by means of institutional acts therefore does not necessarily hold.

The proper comparison between promising and institutional acts, therefore, is not that an agent who performs an institutional act has the intention described in (i3) and is thus obligated regardless of whether or not they have the intention described in (i4), as the initial parallel seemed to suggest. Rather, such an agent intentionally performs the act described in (i1) without necessarily being rationally committed as in (i2) and thus not intending as in (i3) at all.

Of course, as we have seen an agent *can* create obligations through institutional acts *if* they rationally commit themself such that the situation described in (i2) obtains along with the relevant intention-in-action (i1). And such situations surely do exist: if I sit opposite you at a chess board, look you in the eye, and move the Queen's pawn forward two spaces, I am likely to be doing so as part of a speech act of committing acceptance towards the rules of chess and the deontic powers these entail, such that I rationally commit myself to the relevant constitutive rules. Continuing to follow these rules for a dozen turns before suddenly using my rook to knock over all of your pieces would thus rightly be seen as a breach of the rational constraints internal to my previous speech act. However when it comes to social life more generally, we cannot assume that all institutional acts constitute or are accompanied by the kind of committing acceptance described above.

It is worth saying that this parallel with promising even fails for *linguistic* speech acts about "non-linguistic" institutional facts, because unlike in promising the words associated with such facts (words representing the function and powers associated with a status term) have multiple possible referents. For instance, they might refer either to the perceived deontic powers constitutive of a perceived status, or to the actual powers created by a corresponding social institution. If I say to a bully: "OK,

OK, you're the boss, I'll do what you want" this might be a speech act of committing acceptance that the bully has a particular status, and thereby the intentional creation of obligations for myself. But it might mean simply that I descriptively accept that the bully, by virtue of the relevant social institution constituted by others' acting in the appropriate ways with respect to them, has certain *actual* powers beyond their physical strength and that I had better do what they say: the statement might be predictive rather than declarative.

2.3 The role of statuses in social institutions

Where does this this leave the narrower definition of social institutions? I suggested above that not only is this definition intuitively mistaken, but that Searle seems to recognise this. It would seem, then, that we would need good reason to abandon the broader definition in favour of it.

In section 2.1, I suggested that Searle's occasional invocation of this narrower definition might be explained by an ambiguity in his use of the term "acceptance". In section 2.2, I showed that Searle sometimes seems to claim that all institutional acts create rational commitments, and suggested an argument that Searle might be read as giving for this claim. If this claim were true, it might give us good reason to adopt the narrow conception, for it would not only explain how social institutions narrowly-defined might sustain themselves over time, but would seem to suggest that all of those patterns of activity that would be called social institutions on the broader definition would necessarily end up turning into social institutions in the narrowly-defined sense. However, I argued that this implicit argument fails.

For this reason, it seems we have no good reason to adopt the narrower definition of social institutions, and are best placed adopting the more intuitive, broader definition: social institutions are patterns of activity amongst a group in which each agent acts in accordance with the deontic powers constitutive of a perceived status, and in which each agent's motivation for so acting involves any of a broad range of attitudes that represent a perceived status or its constitutive deontic powers in some way. Further, given the distinctions outlined in section 2.1, we can now be more specific about this broad range of attitudes: they can be attitudes of either normative or descriptive acceptance.

This broader definition might appear to make social institutions closer to Humean conventions than they would be on the account suggested by the narrower definition. After all, according to it an agent's activity can be partly constitutive of a social institution even if they are motivated only by desires and beliefs concerning others' behaviour and attitudes. However, the difference that remains between the broader definition and Humean conventions is crucial. When I am motivated to do as Anna says because I want to play football with the group and, while not thinking she has any *rights* to make decisions, descriptively accept that every other player does what she says because *they* normatively accept that she has such rights, the notion of Anna's rights and the obligations these entail still play an important part in my motivation. Indeed, if I did not believe that these others took themselves to have such obligations I might act differently: rather than following Anna's decisions, I might try to convince the other players to eject Anna for being imperious. Further, my belief that others have so accepted Anna's rights *might be mistaken*; and further still, *every* player might in fact be doing what Anna says based on similarly mistaken attitudes of descriptive acceptance. In such a situation, the notion of statuses and their constitutive deontic powers continue to do important explanatory work: it is only our beliefs about others' attitudes towards these perceived rights that can explain why we are acting as we are. And an account of Humean conventions would not be able to capture such a situation adequately.

Searle often frames human institutional life in contrast with hypothetical examples featuring well-trained animals. For instance, he says:

> Suppose I train my dog to chase dollar bills and bring them back to me in return for food. He still is not buying the food and the bills are not money to him. Why not? Because he cannot represent to himself the relevant deontic phenomena. He might be able to think "If I give him this he will give me that food." But he cannot think, for example, now I have the *right to buy* things and when someone else has this, he will also have the right to buy things.[34]

My general point here is that agents can be motivated by *representations of* deontic powers (representations not available to Searle's dog) without taking themselves to be subject to deontic powers *qua* being *rationally committed* to them. I can be motivated by either normative acceptance of statuses and deontic powers or by descriptive acceptance of social institutions, and neither kinds of motivation are

---

[34] Searle, 70.

available to Searle's dog. To understand how social institutions exist and affect our social lives, we must acknowledge both kinds of mechanism, rather than—as Searle seems to do in his implicit narrower definition—insisting that we either act on desire-independent reasons or act like well-trained dogs.

We now have the conceptual tools with which we can carry out our translation of the concept of the alienation of social institutions to the level of individual agents. Before doing so, it is worth recapitulating the account given over the previous two chapters.

On the account of social institutions I am proposing, social institutions consist in patterns of activity amongst (a proportion of) a group, in which each acting agent acts in accordance with the perceived deontic powers constitutive of a perceived status, and in which each agent so acts motivated by an attitude that represents the perceived status or the deontic powers constitutive of it. I have argued that the relevant kind of motivation of these agents must be understood broadly, as involving attitudes of either of two broad types. On the one hand, the agents whose activity constitutes a social institution can be motivated by *normative acceptance*, i.e. by the rational recognition of their obligations to so act, obligations created through prior speech acts of committing acceptance. The process by means of which agents create obligations for themselves using status terms was examined in chapter 3. On the other hand, this activity can be motivated by *prudentially motivating descriptive acceptance* that the social institution exists in their community. Such agents use the relevant status terms to represent to themselves the fact that others in their community are acting in accordance with the perceived status, as well as to represent to themselves the possible motivations of these other agents.

Given the account described above, according to which social institutions can be constituted by the activity of agents with importantly different kinds of motivation, the relation between social institutions and statuses is complex, and indeed more complex than Searle himself seems to take it to be. Insofar as a social institution is constituted by activity motivated by normative acceptance of obligations previously created through committing acceptance, the perceived statuses involved in that social institution actually obtain: people have actually rationally committed themselves to act in the appropriate ways, thereby creating the obligations constitutive of the status. Insofar as a social institution is constituted by activity involving prudentially motivating descriptive acceptance, these statuses do not actually obtain: agents use the relevant status terms in their practical reasoning,

but have not rationally committed themselves to act according to them by means of committing acceptance.

This account, I take it, satisfies our purpose of developing an account of social institutions which can account for the normative dimension of social institutions—that we often act according to them because we feel we 'ought or 'should'—, while making clear both the potential origins of such motivations in our speech acts of committing acceptance and the possibility (indeed the commonplace) that we often act according to them for other, less normatively-loaded reasons.

At the end of chapter 3, I pointed out that Searle's account of statuses amounted to a kind of actual contract theory of statuses, according to which statuses exist by virtue of speech acts of committing acceptance by those whose obligations constitute them. In this chapter, we have seen that this social contractarianism need not follow through to our account of social institutions that employs this notion of statuses: social institutions *can* involve such speech acts, but they need not. However, we have also seen in section 2.2 that this is not always apparent to Searle: he sometimes seems to write as if social institutions *necessarily* involve deontic powers and thus speech acts of committing acceptance. Insofar as he takes this to be the case, it seems he mistakenly represents his actual contract theory of statuses as carrying through to his account of social institutions, thereby resembling those accounts of social contracts that employ a problematic notion of tacit consent.

3 Social institutions' independence

In chapter 2 I reconstructed the basic structure of Marx's concept of the alienation of social institutions as follows:

(1) A group of agents *X* act together, thereby constituting social institution *Y*

(2) Social institution *Y* becomes independent of group of agents *X*. This results from members of group of agents *X* mistakenly acting as if the social institution was independent of their activity as a result of a certain kind of mistaken belief.

(3) That (2) is the case enables social institution *Y* to influence group of agents *X*'s behaviour in a way in which it otherwise would not. This constitutes a restriction of their freedom.

In sections 1 and 2 above, I have provided a definition of social institutions, employing a broadly Searlean framework to suggest that a social institution is a pattern of activities amongst a group in which (a significant proportion of) agents act in accordance with the perceived deontic powers constitutive of a perceived status, motivated by attitudes that represent this perceived status in some way. And this definition gives us a clearer account of (1) above: a social institution exists amongst group X in the sense that (a significant proportion of) members of X act in accordance with its perceived deontic powers (motivated by attitudes that represent these). In this section, we turn to (2): what does it mean to say that a social institution becomes *independent* of the activity of this group?

As formulated above, (2) involves two claims, one at the macro level of description (the first sentence) and one at a micro level of description (the second). In keeping with my methodology throughout, I will here focus on the micro-level description of the second sentence. However, we will need to return below (in section 3.4) to the macro level, to show how our account at the micro level might account for the original intuition in Marx's writing from which our account derives, that according to which in the contemporary world social institutions, while constituted by our activity, have become in some sense independent of us.

It is worth pointing out that even the second sentence that is our focus contains a macro level term, "their". To see why it remains at this stage, consider the fact that each individual member of X is not only an individual, but a member of X, as well as potentially a member of any number of sub-groups within X. When we say that members of X mistakenly treat the social institution as independent of "their" activity, then, it seems that this is ambiguous between their each treating it as independent of their own activity *qua* an individual, or of their own activity as a component of the collective activity of X or of a sub-group of X of which they are part.

What, then, does it mean to say that members of a group mistakenly *acts as if* the social institution was independent of their activity? And what kind of mistaken beliefs must be involved in so acting? It is to these questions that we now turn. It is worth saying that in analysing agents' mistaken beliefs and how these might motivate them to act, I will rely on a relatively thin conception of human agency. In particular, I will treat an agent's desires as relatively stable and unquestioned, focusing on the idea that to be mistaken is to have a *mistaken belief*. Nevertheless,

as I will argue in chapter 5, even limited to this thin conception of human agency the concept of alienation can have important normative implications.

It is important to emphasise that Marx's claim is not necessarily that every member of *X*, nor even every member of *X* who acts in accordance with the relevant social institution, is motivated to act directly by a mistaken belief of the relevant kind. I will address in 3.3 the complicated question of the relation between those members of *X* who are so motivated and those who are not. In the following discussion, however, I will focus primarily on those members of *X* who do so act motivated by such mistaken beliefs, for it is them with reference to whom our concept is defined.

Our claim, then, is that certain agents mistakenly act as if a social institution was independent of "their" activity. To clarify this claim, we need to understand two key components of it. First is the idea that agents each act *as if* a social institution were independent (of "their" activity). How might a social institution's dependence or independence affect an agent's actions such that they can be described as acting "as if" it were independent? Second is the idea that their so acting is *mistaken*. The notion of an act being mistaken implies, I take it, that there is a mistaken belief involved, namely, in this case, a mistaken belief concerning the independence of the institution. We will need to spell out the content of such beliefs. Further, by virtue of the description of this belief as mistaken, it must *actually be the case* that the institution in question depends upon "their" activity: this ought to narrow down the kinds of beliefs we are concerned with. I will address the relevant kind of mistaken beliefs first (section 3.1), before going on to examine the kind of mistaken activity motivated by such beliefs (in section 3.2).

3.1 The dependence of social institutions on members of *X*

To get a clearer picture of the mistaken beliefs involved when members of *X* mistakenly treat *Y* as independent of "their" activity, there are a number of terms we must clarify. First, we must specify what we mean by *independence* and, correspondingly, dependence (3.1.1). Second, we must be more specific about exactly what it is about *Y* that agents mistakenly think is independent of "their" activity (3.1.2). Third, since members of *X* are said to be *mistaken* in acting as if social institution *Y* were independent of "their" activity, it must be the case that *Y is in fact*

dependent on their activity, and we therefore need to ask whose activity a social institution might be so dependent on (3.1.3).

3.1.1 Defining independence

I will define dependence broadly as the relation between entities or events, *A* and *B*, in which *A* depends on *B* just if *A* would not have existed or occurred if *B* had not existed or occurred. Conversely, we would then say that *A* is independent of *B* just if *A* would have existed or occurred even if *B* had not existed or occurred. When we are considering agents' mistaken beliefs about an institution's dependence or independence of "their" activity, I take I that it is this broad definition that is applicable. We will have reason to differentiate between two important kinds of dependence below (section 3.4).

3.1.2 What about *Y* do members of *X* mistakenly think depends on "their" activity?

As specified in (2) the content of members of *X*'s mistaken beliefs is rather vague. They mistakenly believe that *Y* is independent of "their" activity: but what is it about *Y* that they mistakenly believe to be so independent? And, if these beliefs are mistaken, what is it about *Y* that is in fact dependent on "their" activity?

The most obvious answer is that it is the *existence* of *Y* that is dependent on "their" activity (and which they mistakenly believe is not): after all, as we have seen, *Y* is constituted by them (or a significant proportion of them) acting in accordance with the relevant perceived deontic powers. From our discussion above about what it means to say that a social institution exists, this suggests that we are concerned with the fact that (a significant proportion of) agents are actually acting in accordance with the perceived deontic powers constitutive of a perceived status motivated in the relevant way.

This, however, raises an immediate problem: what is the relevant proportion of agents that must be so acting? Clearly, a social institution can be said to exist within a group despite the fact that not everyone in the group acts in accordance with its perceived deontic powers. Consider Anna's status as referee once more: if Anna blows

the whistle for full-time, and 21 of the 22 players stop playing, we would feel comfortable in our assertion that Anna is the referee in this group. But what if only two players stop playing? We would no longer feel comfortable in our assertion. But it remains true that this social institution exists within the group: there is a proportion of players—two of them—who are acting in accordance with the relevant perceived deontic powers and motivated by the relevant kind of attitude. Further, this raises the question of how we define our group X in the first place. If not every member of X acts so as constitute social institution Y, then how do we say which non-acting agents are nevertheless members of X, and which are not?

Our example of the relation between Anna's position as referee and the players present on that day suggests a way of defining the relevant group, namely as that group of agents who are affected by the existence (or non-existence) of the institution. In contrast to every person in the park that day—including dog-walkers, joggers, and children playing on the swings—, the group of players are those who are affected by the existence of the institution; they are those whose own behaviour will be affected by what institution exists in this group. A dog walker might be affected by the fact that a game of football is in progress: they might be forced to walk the long way home to avoid walking through the middle of it. But they are unlikely to care, such that their behaviour is affected, whether Anna is the referee or not. This definition of X, as that group of agents who care about the existence of the social institution, is ultimately justified by our focus in this chapter on the concept of alienation: as we have seen, the outcome of this process, in (3), is agents' activity being affected by the social institution, and as such we are interested primarily in those agents who care about the existence or non-existence of Y, such that their activity could be so affected.

Further, it seems that this way of defining X might help to resolve our worry about what proportion of X must be acting according to the perceived deontic powers of a social institution to say that it exists in X in the relevant sense. That is, we might say that a social institution exists in X in the relevant sense when enough members of X are acting according to its perceived deontic powers that this matters to the members of X such as to affect their behaviour. Thus, we might say that while the group of players care whether or not the social institution "exists" in the sense that, say, most people are acting according to its perceived deontic powers, none of them might care that only two people are so acting. To capture this distinction, I will say that a social institution exists *amongst* a group when enough of its members are

acting accordingly that this fact matters to the members of the group as a whole. We might contrast this with the claim that a social institution exists merely *within* a group, as when only two players do what Anna says.

Of course, the exact proportion of members of X that must be acting in the appropriate way for this to matter to any member does not seem to be easily quantifiable. Indeed, it seems that for certain kinds of social institution, a higher proportion is necessary than for others. (It is this ambiguity, I take it, that motivates Searle's frequent ambiguous language on this issue.[35]) I would like to suggest that this is because different social institutions play different roles for agents: they are used for different purposes, relative to different desires, and with different levels of acceptable risk. For instance, my belief that most people will treat my banknotes as money is sufficient to motivate me to do so: that some overly-cautious English shops will not accept my Scottish £10 notes is not sufficient reason for me not to, for I believe that most shops will. Other institutions have higher risks, and I therefore have a higher threshold for saying that the institution exists such that I care about its existence and can use it. For instance, if I believe that a significant minority of people drive through red lights, I am unlikely to treat traffic lights as indicating that it is safe to cross the road: I would still look both ways before crossing. If I believed that everybody stopped at red lights, I could use them to save myself the trouble of looking. The proportion at which I would say that a social institution exists *in the sense that matters to me* is therefore different for money and traffic lights respectively.

For this reason, let us say that what we are concerned with is a social institution's existence amongst X in the sense that a sufficient proportion of agents are acting according to its perceived deontic powers that this fact matters to members of X. While we cannot say for all social institutions what this sufficient proportion is, we might assume that for any particular social institution, and given the concerns of the agents in X and their reasons for caring about whether or not the institution exists in this sense, there is some particular proportion that matters to them. Let us call this, somewhat inelegantly, *the necessary proportion*.

In what follows, then, we will be concerned with members of X's beliefs concerning the dependence of the existence of Y on "their" activity in this sense. Thus, the

---

[35] See, for instance, Searle, 48, 117.

existence of institution *Y* depends on "their" activity just if the necessary proportion of people would not be acting in accordance with its perceived deontic powers—i.e., *Y* would not exist amongst *X*—were "they" not so acting.

3.1.3 On whose activity does Y depend?

To say that the members of *X* mistakenly act as if *Y* were independent of their activity implies that *Y in fact does* depend on their activity. And we here must deal with the remaining "their" in our account. Do members of *X* mistakenly act as if *Y* is independent of *their own* activity, as individuals? Or as independent of the collective activity of *X* as a whole, or of the activity of some sub-group of which they are a member? On whose activity might *Y* depend such that the members of *X* can be mistaken about this relation?

It seems we have four options (although I will introduce a fifth below). First, an agent might believe a social institution is independent of the activity of the largest possible relevant group, namely *X*. Second, an agent might treat an institution as independent of the activity of those members of *X* who are in fact acting according to its perceived deontic powers. Third, an agent might treat an institution as independent of some sub-group of which they are a member. And fourth, the agent might treat the institution as independent of *their own* activity, as an individual. Let us consider these four options. (I will consider the fourth before the third.)

Let us start with *X*: in what circumstances is it correct to say that a social institution depends for its existence upon the activity of *X* as a whole? I take it that our intuitions are that in general social institutions are always dependent upon the activity of a group in this sense. However, it is worth saying that this rests on how we interpret "the *activity of* the group". If, for instance, by this we mean the activity of every single member of this community, then it would seem that very few social institutions depend for their existence on the activity of the group so defined: as we have seen, most social institutions can exist even in the face of some proportion of the community not acting accordingly.

Taking less stringent definitions, however, things look different. If by the activity of the group we mean the activity of a majority of its members, then we can say that for any social institution for which the necessary proportion is fifty percent—and, I

take it, this is most social institutions—*Y* depends on the activity of *X*. Without most of the players doing what she says, it would seem hard to justify the claim that the social institution according to which Anna is referee exists amongst this group of players, even if a handful of players do do what she says. On this account, then, any agent who believes that such a social institution exists independent of the activity of *X* is mistaken in doing so.

Second, we can consider the interpretation of "their" in (2) according to which agents are said to mistakenly believe that an institution is independent of the activity of those agents whose activity actually constitutes it, i.e. those who are actually acting in the appropriate way. This interpretation is not particularly helpful. First, if this group is defined in terms of a specific set of individual agents—those who are currently acting appropriately—, then very few institutions depend on such a determinate group: most social institutions, and certainly all that are of central interest to us, can survive a change in their membership, such that it is rarely the case that an institution depends upon the activity of those specific individuals whose activity currently constitutes it. On the other hand, if this is understood more flexibly, as claiming that the social institution depends upon there being a group of agents whose activity constitutes it, then it is true *a priori* but uninteresting: this claim is implied by our very definition of what a social institution is.

Third, what of the interpretation of "their" according to which individual agents are said to mistakenly treat a social institution as independent of *their own* relation to it? Can social institutions ever depend, in the sense outlined above, on a particular individual?

There are at least two possible approaches to answering this question, and our answer to it depends on which we take. I will first outline the most intuitive approach, according to which it is not possible to describe a social institution as dependent upon an individual.

At first blush it seems obvious that any social institution cannot depend upon the activity of an individual agent. While if I act according to the perceived deontic powers of a status my action—as part of the necessary proportion—partially constitutes the existence of the relevant social institution, it is not the case that if this social institution exists I am necessarily acting according to it, for if I don't so act but everyone else does, the institution still exists. (This is not the case for social institutions which require unanimity amongst *X*, however I take it that no social

institutions that are of concern to the concept of alienation are of this kind.) According to this approach, then, it seems that an agent cannot mistakenly believe a social institution to be independent of their own activity: the existence of a social institution cannot depend on an individual's activity, so a belief that it is independent of that activity is necessarily correct.

To understand the second approach, we must see that the first relies on an implicit methodological decision, namely the decision to view the situation from the point of view of the institution, in contrast to that of the agent(s) involved.

To see the significance of this decision, consider that, from the point of view of the institution, precisely which agents' activity constitutes the institution is irrelevant: all that the institution requires is that a certain proportion of people accept and therefore act according to its perceived deontic powers, and these acting agents are from this point of view completely anonymous. However, *from the point of view of a particular agent,* there is no such anonymity: the agent can potentially know about the attitudes and motivations of many of their neighbours. From this subjective point of view, we *can* portray an institution as dependent upon the activity of a particular individual in certain, very narrow, circumstances.

Take for example a social institution for which the necessary proportion is fifty percent, and which exists in a community (as defined above) consisting of three agents, 1, 2 and 3. From the point of view of the institution, it does not matter who acts according to its constitutive rules: as long as at least two do so, the institution exists. As such, while the institution depends for its existence on the activity of this group, it does not depend upon that of any agent in particular. If, however, we take the point of view of one of these agents, things look different. Say that agent 1 has a motivating attitude of acceptance concerning the institution and will act accordingly, while agent 2 does not. In this situation, from the point of view of agent 3 the institution *does* depend on her activity: again taking the attitudes of 1 and 2 as given, if the institution exists then 3 is necessarily acting according to the relevant deontic powers. In such a situation, for agent 3 to treat the institution as independent *of her* is to be mistaken. Conversely, if both 1 and 2 have a motivating attitude of acceptance towards the institution and will act accordingly, then the institution will exist regardless of how 3 acts: in this context, again taking the attitudes of 1 and 2 as given, the institution *is* independent of her, and treating it as such therefore no mistake.

Understood in this way, then, it seems that—in admittedly very unique and unlikely situations—an individual agent can mistakenly treat a social institution as independent of their own activity.

We can now turn to the interpretation of "their" according to which agents mistakenly treat a social institution as independent of the activity of a sub-group of which they are a part. In order for beliefs or acceptances that a social institution depends upon the activity of such a sub-group to be mistaken, again it must be the case that the social institution in question is in fact dependent on the activity of such a group: is this possible?

That social institutions can so depend on sub-groups seems intuitively plausible. For instance, it seems right to say that the social institutions of capitalism depend on the activity of the proletariat: without the activity of the proletariat, the social institutions of capitalism would not exist. However, it seems that there is an important intuition underlying the plausibility of this example. This is that when we talk of the proletariat in this context, we assume that the various proletarians have common interests (at least insofar as these social institutions go), and that as such it makes sense to talk of them *as a group*. The social institutions of capitalism might similarly depend upon all people born on weekends: if everyone born on a weekend did not act according to these institutions' perceived deontic powers, these institutions would likely collapse. But this claim has none of the intuitive appeal of the proletariat example. For reasons that will become clear in section 3.2, I take it that it is only the stronger form of dependence on sub-groups involved when that sub-group is defined as a group of potential cooperators with which we are concerned.

To clarify, let us consider another example involving Anna's position as referee. Say that for us to be able to say that this institution exists, over half of the players must act according to the relevant deontic powers. Say also that the players consist of two groups. One group, "the Regulars", consists of a group of eleven regular players who play together in the park each week and cooperate in deciding who they will treat as referee. The other consists in a rag-tag collection of eleven players who happen to have been in the park that day. When we ask on whose activity the social institution depends, we can see that alongside the dependence on the activity of the community as described above, we can also talk of the social institution depending upon the Regulars. Because the Regulars make up half of the group and decides how to act cooperatively, as a sub-group, and because for the institution to exist in the relevant

sense requires more than half of the players to be acting accordingly, we can say that the institution depends on the activity of the Regulars.

Further, it is not only such large sub-groups upon whose activity a social institution can depend: as with the discussion of dependence on individual agents above, we might take the point of view of the relevant sub-group itself in order to see that in situations in which certain individual agents or other sub-groups' activities are already determined, we can characterise even relatively small sub-groups as having a dependence relation with social institutions. (Think, for instance, of small parties in coalition governments.) As with the discussion of dependence on X, again a dependence relation need not require every single member of the sub-group to act appropriately. For instance, if the Regulars in fact consisted of considerably more than half of the players, it would still be the case that the institution depends upon the activity of the Regulars, but here there is scope for a handful of dissenters amongst them. There is, rather, a complex relation between the proportion of the relevant sub-group who must act accordingly and the necessary proportion of X who must do so, one which it is unnecessary to examine in detail here.

This raises a fifth sense in which an agent might mistakenly believe a social institution to be independent of their activity. This has to do with an agent's framing of the possible point of view available to them, and in particular the beliefs that frame whether cooperation with other agents to form a cooperating sub-group is an option that is available to them.[36] To see this, consider the following example. Take a community consisting of five agents 1, 2, 3, 4 and 5, and a social institution with a necessary proportion of fifty percent. As previously, let us adopt the point of view of agent 3. Let us say that the motivations of agents 1 and 2 are given: both accept and act according to the institution. Is there a relation of dependence here between agent 3 and the social institution? In an important sense, this depends on the motivations of agents 4 and 5. If agents 4 and 5 are, in fact, potential cooperators with agent 3, then we can portray the social institution as in fact dependent upon the activity of potential group 3-4-5. Taking the activity of 1 and 2 as given, the existence of the institution depends on the activity of the group 3-4-5: if this group acts according to the deontic powers, the institution exists (and if the institution exists, taking the attitudes of 1 and 2 as given and the agents 3, 4 and 5 as a group, then group 3-4-5

---

[36] My discussion in this section of the 'framing' of the decision situation draws on Michael Bacharach's work. Bacharach, Gold, and Sugden, *Beyond Individual Choice*.

are necessarily acting according to it). In this sense the social institution partly depends upon agent 3 *qua* member of this sub-group. If, however, agents 4 and 5 are not potential co-operators in this sense (if, for instance, they have different goals with respect to this institution), then there is no relation of dependence: however agent 3 acts, the social institution may exist (if either 4 or 5 act according to it) or may not (if neither do).

In this scenario, then, agent 3's treating of the social institution as independent of her activity, based on a belief that agents 4 and 5 are not potential cooperators, is either correct or mistaken depending on whether or not agents 4 and 5 are in fact potential cooperators. This has significant implications, for it means that not only mistaken beliefs about the institution and its dependence directly, but also mistaken beliefs about the other members of the community's motivations, can determine whether 3's treating of the institution as independent of themself (considered as either an individual or a potential group member) is mistaken or not.

Above we saw that given our definition of social institutions, as well as certain assumptions about necessary proportions and our (less stringent) definition of group activity, an institution's existence necessarily depends on the activity of X, and so an agent's belief that an institution exists independently of the activity of X is necessarily mistaken. However, when it comes to the dependence of social institutions on individuals and sub-groups things are more complex. Agents' beliefs that a social institution is dependent or independent of themselves or of a sub-group of which they are a part can be either mistaken or not, for social institutions might or might not so depend on them.

3.2 What is it for agents to act *as if* the institution were independent?

The previous section focused on the possible contents of the beliefs agents might have regarding the dependence or independence of a social institution on "them". Here, we must turn our attention towards the activity that results from such beliefs: what does it mean to say that these agents mistakenly *act as if* the social institution Y were independent of "their" activity, as a result of mistaken beliefs of any of the kinds described above?

Our focus here must again be directed by our goal: agents' acting as if a social institution were independent is said, in (3), to enable the social institution Y to affect the behaviour of members of X in a way in which it would otherwise not. This, I take it, has two consequences for the kind of activity with which we must be concerned in (2), and hence also for the other characteristics of those members of X with whom we are primarily concerned.

First, it seems clear that the agents in X must be motivated by their mistaken beliefs to act in a way in which they would not act were they not to mistakenly believe the institution to be independent of their activity. If this were not the case, their activity could not have any effect on members of X that it would not have otherwise had. That is, the mistaken belief must be a *decisive* reason for these agents, and not merely one of the reasons that they act upon.

Second, it seems we are interested primarily not only in activity of members of X for which their mistaken belief was a decisive reason, but in such activity *of a particular kind*. In particular, we are interested in activity that enables social institution Y to affect the behaviour of the members of group X as a whole. Since our definition of X was as that group who care about the existence (in the relevant sense) of this social institution such that its existence might affect their behaviour, it seems we are primarily interested in activity amongst X *that makes it the case that the social institution exists* in the relevant sense, i.e. that makes it the case that the necessary proportion of people are acting according to the relevant perceived deontic powers. As such, we are primarily interested in members of X who, as a result of the kinds of mistaken beliefs described above, *act in accordance with these perceived deontic powers*, thereby partly constituting the social institution.

We need, then, to understand how a member of X might be decisively motivated to act in accordance with the perceived deontic powers of institution Y by a mistaken belief concerning the independence of Y from "their" activity, in any of the senses outlined above. Why might the independence of an institution be a decisive reason for an agent?

Let us first focus on the unusual situation in which I mistakenly believe a social institution is independent of my activity as an individual, in the sense described above with reference to the group of agents 1, 2 and 3. Given what we have said above about the kind of activity with which we are primarily concerned, we are considering a situation in which were I to correctly believe the institution depended

upon my activity I would decide not to act in accordance with its perceived deontic powers, but in which were it to exist independently of my activity I would have a decisive reason to act according to these. If I am to be so affected by my mistaken belief, it seems that I must have a motivational make-up such that I have a reason for acting according to the relevant perceived deontic powers *only* if the necessary proportion of other agents in that community are so acting or will so act. Given the discussion in the previous chapters, it seems that this motivational make-up can be of two kinds.

First, I may not want to act according to the perceived deontic powers associated with the social institution in question, nor want the social institution in question to exist in my community, but may have desire-dependent reasons for so acting if it does so exist. If I have this constellation of preferences, and if the institution exists, and if I believe that whether or not it persists depends upon whether I act according to its perceived deontic powers, then if I am acting rationally I will not so act. If, on the other hand, I believe that its existence is independent of my activity, then I will be motivated to do so. As examples of such a situation, we can think of the kinds of desire-dependent or prudential reasons cited earlier: my belief that a social institution will persist independent of how I act, along with a desire not to stand out or to be sanctioned for nonconformity, will together motivate me to act according to its perceived deontic powers, in a way in which I would not act were I to take the institution's persistence to depend upon my activity. (Scenarios of this kind will be examined more closely in chapter 5, when we consider the extent to which agents' being caused to act by this kind of scenario constitutes a restriction to their freedom.)

Second, we can imagine a similar scenario involving the social grounding of my creation of obligations by means of committing acceptance of a status. Say that I can envision two separate possible social institutions, A and B, involving two separate sets of deontic powers. Say that I prefer A, but would prefer B to no such institution existing. If, as in the example above, in a community of 3 agents I take the situation to be such that which deontic powers and which corresponding social institution exist depend on how I act—say that agent 1 is acting according to A and 2 acting according to B—then I will be motivated to act according to the deontic powers of A, my preferred social institution, thereby (assuming the kind of intention-in-action appropriate for committing acceptance) creating the relevant obligations for myself. If, on the other hand, I mistakenly believe that both agents 1 and 2 are in fact acting

according to B, then I take the social institution to exist independent of my activity, and as such will be motivated to act in accordance with and thus commit myself to the deontic powers associated with B, my non-preferred option.

I take it that what can be said of agents with the first kind of motivational make-up can be said, *mutatis mutandis*, of those with the second, more complex kind. As such, in what follows I will focus only on motivations of the first kind.

What, then, of agents who mistakenly believe a social institution is independent of the activity of a cooperative sub-group of which they take themselves to be a part? We saw in chapter 3 (section 3) that Searle's account of collective Intentionality suggests that agents often act not only on intentions that they attribute to themselves, but rather on intentions that they attribute to a group of which they take themselves to be a part. Where several people act on shared collective intentions of this kind in conditions of common knowledge, the agents can be said to be *cooperating*. Insofar as agents are part of a sub-group of potential cooperators, it seems that a social institution can be dependent upon the collective activity of this group in much the way that it was said to be dependent upon the individual's activity above. Let us posit a sub-group of potential cooperators who have the kind of motivational make-up described above, i.e. who do not want a social institution to exist but who have a decisive reason to act according to it in situations in which it does exist. An agent who does not want a social institution to exist, who takes themselves to be part of a sub-group of cooperators who similarly do not want it to exist, and who takes the existence of the social institution to depend upon the activity of this cooperative sub-group, will be motivated not to act—as part of the sub-group's not acting—in accordance with the relevant perceived deontic powers. In contrast, an agent who does not want a social institution to exist but who takes the existence of the social institution to be independent of the activity of such a group will be motivated to act according to its perceived deontic powers.

Similarly, as we saw above, an agent who is mistaken about the existence of a potential group of such cooperators can be considered, by virtue of their mistaken framing of their options as necessarily from the point of view of themselves as an individual rather than as a member of such a group, to be mistaken in believing that the social institution was independent of their own activity. In such a situation, we might say, the social institution's persistence depends on this sub-group of potential cooperators not learning of one another's willingness to cooperate, and thus on their

continued activity. As with above, if such an agent does not want a social institution to exist but has reason to act in accordance with its perceived deontic powers in the case that it does exist, they will be motivated to act in accordance with it by virtue of this mistaken belief.

Finally, we have agents who mistakenly believe an institution exists independently of its relation to the activity of group *X* as a whole. This can be viewed as a larger-scale version of the mechanism just described: individuals who do not want the institution in question to exist might be motivated by the relevant mistaken beliefs to act in accordance with it rather than campaigning in their community for it to change.

Mistaken beliefs of this last kind play a prominent role in the literature. I take it that it is mistaken beliefs of this kind that are central to the Young Hegelian attempts to apply the concept of alienation to social institutions, and it also lurks behind some of Searle's own comments on the subject. I will discuss these briefly below. However, it is worth noting here that mistaken beliefs of this kind in fact seem rather rare as decisive reasons for agents acting in accordance with social institutions. We can, of course, come up with examples of such mistaken beliefs. Say, for instance, that I believe that the Queen's powers over me exist not by virtue of the social activity of the citizens of the United Kingdom, but by virtue of God's will, and that this belief decisively affects my behaviour: I would ignore these powers were I to believe that their origins were more mundane. In such a situation, I have a mistaken belief that the institution of the monarchy exists independently of the social activity of those on whom it in fact depends in the sense outlined above. More plausibly, perhaps, are certain agents' attitudes towards the social institutions of race and gender, insofar as agents might act towards people of a certain race or gender in a particular way because of a belief that some status associated with these institutions are "natural". Nevertheless, even here it seems that often when beliefs of this kind are invoked to justify certain acts, these are best understood as rationalisations rather than truly decisive beliefs. Regardless, in the kinds of social institutions we have been primarily considering, it seems such attitudes are few and far between. As Ian Hacking puts it regarding money:

> I do remember from childhood a schoolteacher who belonged to a sect named British Israelites, and who maintained that the pound sterling

was God-given. I have never encountered anyone else unaware of the fact that money is 'socially created'[37]

3.3 Reflexivity and modes of existence

In the second sentence of step (2), then, we have a situation in which (some) members of X act according to the perceived deontic powers of a social institution because of mistaken beliefs concerning its independence of their activities. In (3), we see the outcome of this: members of X as a whole in turn have *their* behaviour affected. (To the extent that those members of X who have the relevant kind of mistaken beliefs might ground these on the behaviour of other mistaken members of X, it seems that this can include these agents too.) And this is said to constitute a restriction of their freedom.

We will examine these consequences in more detail in chapter 5. For now, we can sketch the basic idea. We said above that X consists of agents who care whether social institution Y exists amongst this group (in a particular sense: they care whether or not the necessary proportion of agents act according to the relevant perceived deontic powers). To the extent, then, that the activity of those members of X with the relevant mistaken beliefs are acting in accordance with the perceived deontic powers of Y *makes it the case* that Y exists in this sense amongst this group, the activity of members of X more broadly will be affected. The motivations here can be explicated just as can the motivations of those with mistaken beliefs, except here there need be no mistaken beliefs involved: members of X who have the kind of motivational make-up described above, in which they either have desire-dependent reasons to act according to Y if and only if this institution exists amongst X, or take the existence of Y as a ground for an act of committing acceptance regarding the relevant deontic powers, will act according to these deontic powers if and only if the social institution actually does exist in this sense. Their behaviour will therefore be affected by Y as a result of the activity of those members of X who act on mistaken beliefs.

In chapter 2, I described the potential *reflexivity* of Marx's account of the alienation of social institutions insofar as both the beginning of the process (a social institution

---

[37] Hacking, 'Review Symposium on John R. Searle', 85.

depending for its existence on agents' social activity) and its outcome (agents' behaviour being affected) contain a common element. Now, given our more complete account of this process, we can see that in fact this reflexivity is of a significantly stronger kind. Given the outcome that the process of alienation is said to have—the mistaken activity of some members of X enabling Y to affect the behaviour of the members of X as a whole, or in Marx's more frequent terminology, "dominating" and "enslaving" them—we can see that it is not a matter of these members' mistaken activity enabling a pre-existing Y to have this affect, but rather that these agents' activity *makes it the case that the social institution (continues to) exist(s)* (in the relevant sense), and it is its existence that affects the members of Xs behaviour as a whole. In this sense, the alienation of social institutions can be said to be more than a process that occurs to social institutions. Rather, it describes a *mode of existence* of social institutions: it describes a certain way in which a social institution might exist, namely by being necessarily partially constituted by activity motivated by the kind of mistaken beliefs described above. Further, insofar as the members of X who have no mistaken beliefs are motivated to act in accordance with these deontic powers, their activity, too, partly constitutes the social institution.

This sheds some light on Marx and Engels' critique of Bauer and Stirner in *The German Ideology*. As we saw in chapter 2, these authors adopted the Feuerbachian notion of alienation and used it as part of an explanation and critique of the origin of certain social institutions, much as I have argued Marx did. However, what these authors got wrong, for Marx and Engels, was in assuming too direct a link between agents' mistaken beliefs and members of group Xs acting as they would not otherwise act. That is, for these authors, each agent who acts according to a particular alienated social institution does so because *they themselves* have a mistaken belief, much as the religious behaviour of Christians was for Feuerbach a result of their own mistaken beliefs. Interestingly, in this sense Searle gives a very Young Hegelian explanation of the phenomenon we are examining:

> There are all sorts of institution where people cheerfully accept what would appear to be unjust arrangements. One thinks of various class structures, the low position of women in many societies, and vastly disproportionate distributions of money, property, and power. But one feature that runs through a large number of cases is that in accepting the institutional facts, people do not typically understand what is going on. They do not think of private property, and the institutions for allocating private property, or human rights, or governments as human creations. They tend to think of them as part of the natural order of

things, to be taken for granted in the same way they take for granted the weather or the force of gravity. Sometimes, indeed, they believe institutions to be consequences of a Divine Will.[38]

However, as we have seen, the connection between mistaken beliefs and the members of X's activity is more complex than this. And, indeed, this is true even for members of *X* who do have such mistaken beliefs: if I am a member of *X*, and *I* correct my mistaken beliefs, it is likely that the social institution continues to exist. By correcting my belief, then, I simply move from being one of those members of *X* described in (2) to one of those described in (3), and this likely means acting in much the same way as I did before. For these members of *X*, it is not their mistaken beliefs that motivate them to act according to the perceived deontic powers associated with *Y*, but the fact that *Y actually exists* (in the relevant sense). As Marx and Engels put it (in a passage reviewed in chapter 2), the workers that are oppressed by the social institutions of capitalism:

> are most painfully aware of the *difference* between *being* and *thinking*, between *consciousness* and *life*. They know that property, capital, money, wage-labour and the like are no ideal figments of the brain but very practical, very objective products of their self-estrangement and that therefore they must be abolished in a practical, objective way for man to become man not only in *thinking*, in *consciousness*, but in mass *being*, in life. Critical Criticism [i.e. Bauer], on the contrary, teaches them that they cease in reality to be wage-workers if in thinking they abolish the thought of wage-labour; if in thinking they cease to regard themselves as wage-workers and, in accordance with that extravagant notion, no longer let themselves be paid for their person. As absolute idealists, as ethereal beings, they will then naturally be able to live on the ether of pure thought. Critical Criticism teaches them that they abolish real capital by overcoming in *thinking* the category Capital[39]

3.4 Ys independence of X

We can now return to the first sentence of (2), which captures Marx's intuitive claim that the social institutions of the contemporary world have become independent of us. We must here specify this claim a little more precisely, and ask whether our explication of this idea in the preceding sections can successfully capture its intuitive appeal.

---

[38] Searle, *Making the Social World*, 107.

[39] Marx and Engels, 'The Holy Family', 53.

First, let us address the kind of dependence involved in this claim. In section 3.1.1 I outlined a broad definition of dependence, according to which *A* depends on *B* just if *A* would not have existed or occurred if *B* had not existed or occurred. However, within this broad relation of dependence we might differentiate between two important kinds, and this distinction is crucial to clarifying our first sentence of (2).

First we have what we might call, following Brian Epstein, *metaphysical dependence*.[40] Metaphysical dependence is concerned with a relation of *constitution*: *A* metaphysically depends on *B* just if *A* is necessarily partially constituted by *B*, that is, just if *A* would not exist or occur if it were not partially constituted by *B*. In this sense, water metaphysically depends on oxygen: water would not exist if oxygen did not partially constitute it. Similarly, a duet metaphysically depends on each player's playing: the duet will not be being performed if either player does not play.

Second, we have *causal dependence*. *A* causally depends on *B* just if *A* would not exist or occur if it was not caused by *B*. The dishes being clean causally depends on me cleaning them: they're not going to clean themselves. But my cleaning them in no way constitutes their cleanliness.

As we saw in (1), social institution *Y* is constituted by the activity of group of agents *X*. It seems, then, that in (2) we cannot possibly be talking of metaphysical independence, for *Y* cannot both be fully constituted by the activity of *X* and be metaphysically independent of it. It is this fact that, as noted in chapter 2, gives this claim its confusing and counterintuitive appearance, that we are claiming that it is both independent and metaphysically dependent upon the activity of *X*. It seems, then, that we must be concerned with causal dependence: *Y*, while metaphysically dependent on the activity of *X*, is causally independent of *X*. As we saw above, we are primarily concerned with *Y*'s dependence specifically on agents' *activity*. In turning away from metaphysical dependence and towards causal dependence, however, it seems we need to focus our attention not primarily on *X*'s activity (for this is what constitutes, not causes, *Y*), but on that which causes this constitutive activity, namely, on the *decision* of *X*. It seems, then, that we are best clarifying this first sentence of (2) as: social institution *Y* becomes causally independent of the decision of group of agents *X*.

---

[40] Epstein, *The Ant Trap*, 107.

How, then, might a social institution that metaphysically depends on the activity of group X be described as becoming causally independent of the decision of this group? I think some sense can be given to this claim as follows.

Recall that above a group can be said to act "together" in importantly different ways. On the one hand, they can act cooperatively: in Searle's terminology, they can each act on a collectively-attributed intention in conditions of common knowledge, such that together they have a collective intention as a group. I take it that in such a situation, we can sensibly talk not only of the activity of the group, but of a *group decision*. (That is not to say that all instances of collective intending result from group decisions, but simply that we can sensibly talk of group decisions only in the context of the creation of collective Intentional attitudes.) On the other hand, they can act together distributively: each agent acts on an individually-attributed intention, and taken together this can be described as them acting "together". In such a situation, it seems we cannot sensibly talk of the intentional activity of the group in the same sense as above, and therefore neither can we talk of a group decision: the only decisions involved are individual decisions.

On this basis, I take it that we can see how the phenomena described above at the micro level, in which some members of X act according to the perceived deontic powers of Y as a result of a mistaken belief about its independence of their activity (in any of the senses described above), such that this activity makes it the case that Y exists amongst X (which in turn motivates some other members of X to so act), can be intuitively redescribed at the macro level of description in terms of Y's causal independence of the decision of X.

Y metaphysically depends on the activity of X: Y would not exist amongst X if it were not constituted by the activity of X. It further seems that there is a potential relation of causal dependence here, however it is a somewhat complex relation. From the macro level point of view we are here considering, Y does not causally depend upon the decision of any member of X. However, neither does it causally depend upon the group decision of X to act appropriately, for as we have seen, if X does not act together cooperatively, but only distributively, Y might still exist. Nevertheless, there is a more complex kind of causal dependence here. If X decides as a group not to act so as to constitute Y, then Y will not exist. Therefore Y causally depends on X not deciding as a group not to act in the appropriate way.

With this complex form of causal dependence in mind, let us review what effect the kinds of mistaken beliefs viewed above might have. If an individual agent with the kind of motivational make-up we have examined mistakenly believes that Y is independent in any of the senses described above, they will decide, *as an individual*, to act in accordance with the perceived deontic powers of Y. For that agent, Y exists independent of their activity, and so there is no question of them cooperating with others in deciding not to act appropriately. As such, this agent's mistaken belief frames their situation as one in which X deciding together as a group not to act so as to constitute Y is not an option, and the agent therefore decides how to act as an individual. Further, if sufficient members of X are encouraged by mistaken beliefs to decide how to act as individuals in this way, then the possibility of X more broadly deciding as a group not to act so as to constitute Y actually disappears: group X is actually unable to decide how to act as a group, because many of its members have framed their situation such that they must decide how to act as individuals. In this sense, the complex causal dependence of Y on X's decision described above actually disappears, and Y, while being constituted by the activity of X, is in no sense causally dependent on X's decision.

This way of talking, of course, reintroduces the vague macro level terms such as "group activity" which I have been dispensing with where possible, and reintroduces some of the problems raised by such terms. I therefore take it that in discussing the alienation of social institutions, we are best placed sticking to the micro level account given above, in which step (2) describes some members of X mistakenly acting as if the social institution was independent of their activity (in any of the senses outlined above) as a result of a certain kind of mistaken belief, and step (3) capturing the fact that other members of X are motivated to similarly act as a result of the existence of Y that results from this. However, that this process can be alternatively described at this macro level as one in which Y has become causally independent of the decision of X despite being constituted by X's activity demonstrates the extent to which this micro-level description captures much of Marx's original account.

It seems that this account can also capture much of what we saw in chapter 2 Marx took the consequences of this independence to be. For instance, we saw that when social institutions are alienated, individual workers were said to have to "submit to

this force from egoistic need, from necessity",[41] which is explained well by the account I have given above. Similarly, we saw that:

> To say therefore that *man* is alienated [*entfremdet*] from himself is identical with the statement that the *society* of this estranged [*entfremdeten*] man is the caricature of a *true community*, of his true species-existence....[42]

Again, the idea that when social institutions are alienated, this consists in agents being unable to act cooperatively to decide jointly how to act seems to capture the claim that these agents' social life is not a "true community", i.e. one characterised by cooperative activity as species-beings. Finally, I take it that this account is able to explain the connection Marx took there to be between the alienation of social institutions and the freedom of the agents whose activity constitutes them, to which we turn in chapter 5.

4 Summary

Given the account above, then, we can specify the concept of the alienation of social institutions more precisely.

(1) A group of agents *X* act together, thereby constituting social institution *Y*. This means that the proportion of members of *X* who are acting in accordance with the relevant perceived deontic powers is large enough to matter to the members of *X*.

This social institution is *alienated* when:

(2) *Y* becomes causally independent of the decision of *X*. This means that *Y* exists in the sense outlined in (1) only by virtue of being partially constituted by the activity of members of *X* who so act only because of a mistaken belief concerning the independence of the social institution *vis-à-vis* either (i) themself as an individual, (ii) a cooperative sub-group of which they take themself to be a part, (iii) *X* as a whole, or (iv) are mistaken about the available opportunity for cooperative activity with others.

---

[41] Marx, 'Economic and Philosophical Manuscripts', 269.
[42] Marx, 'Excerpts from James Mill's *Elements of Political Economy*', 265–66.

When this is the case:

(3) That (2) is the case enables Y to influence group of agents X's behaviour in a way in which it otherwise would not. This constitutes a restriction of the freedom of members of X.

In chapter 5, we will turn to clarifying (3).

Chapter 5
Freedom and alienation

In chapter 4 I reconstructed Marx's concept of the alienation of social institutions in terms of the mode of existence of social institutions that it describes. However, in chapter 2 we saw that Marx's concept of alienation is not only a descriptive concept, but also a normative one: to call a social institution alienated is to criticise it. I suggested that in Marx's early work there are two distinct normative threads, one—most commonly discussed in the literature—a kind of perfectionism based on his account of human beings as species-beings, and another based on some conception of freedom. It is the latter thread, I claimed, that is most closely connected to the concept of the alienation of social institutions. As we saw in chapter 2, for instance, in the social institution of money:

> the *mediating function* or movement, human, social activity, by means of which the products of man mutually complement each other, is *alienated* [*entfremdet*] and becomes the property of a *material thing* external to man, viz. money.… The relation between things, human dealings with them, become the operations of a being beyond and above man.… His slavery reaches a climax.[1]

Freedom, however, can mean—and has been explicated in the philosophical literature as meaning—a great many things, and Marx nowhere gives an adequately detailed account of what he means by it. In this chapter, I would like to present a conception of freedom from which the concept of alienation I have reconstructed in previous chapters can draw its normative force. It is this conception, I argue, which allows Marx to criticise alienated social institutions as restricting freedom despite surface appearances.

---

[1] Marx, 260–61, translation amended.

1 Methodology and meta-theoretical considerations

The philosophical literature on freedom abounds with competing theories, differentiating between conceptions of freedom along numerous axes. It is, however, largely unanimous in its methodology, which primarily involves appealing to our intuitions and inherited beliefs about freedom to distil from these a more philosophically coherent conception that can be applied to more difficult borderline cases about which there is little intuitive unanimity. I see little viable alternative to this methodology, and think we must understand Marx's invocations of freedom as similarly relying on an appeal to an intuitive concept of freedom that he believed his readers shared with him. Of course, Marx desired to achieve more than to clarify the intuitions of his readers or solve tricky borderline cases: rather, he wanted to harness these intuitions for revolutionary means, and connecting freedom with the alienation of social institutions was one way of doing so.

My contention is that the intuitive concept of freedom that Marx appeals to is broadly speaking the same as that to which contemporary liberal invocations of freedom appeal. We can therefore draw on contemporary liberal philosophical literature in clarifying this intuitive concept, before seeing how its connection to the concept of alienation extends Marx's critique beyond that of these theorists.

This, however, is complicated by the fact that our intuitions concerning freedom are not unanimous. In particular, the extent to which freedom connotes a *normatively relevant* concept varies. We can distinguish three levels of such normativity.

First, we sometimes use "free" in a *normatively-neutral* sense, as when we say that a boat came free of its moorings. Similarly, we might say that I am not free to jump over the moon: while perhaps awkward, this use does not seem incorrect, and neither does it stir in us any particularly normative reactions. This normatively-neutral sense of freedom maps onto our notions of "ability": the boat is able to float away; I am not able to jump over the moon.

Many philosophers describe their concepts of freedom as "descriptive" and thus normatively-neutral in this sense.[2] However, while a normatively-neutral sense of "freedom" exists in our language for topics such as boats and superhuman acrobatics, these authors intend their discussions of freedom to apply also to politics, and this

---

[2] Perhaps most explicitly, Oppenheim, '"Constraints on Freedom" as a Descriptive Concept', 305.

normatively-neutral conception seems ill-fitting for such contexts. As William Connolly has convincingly argued, the whole point of invoking freedom is to say something normatively significant:

> In the ordinary language of political life and in more formal systems of political inquiry the normative dimensions in the idea of freedom are not attached to it as 'connotations' that can be eliminated; without the normative point of view from which the concept is formed we would have no basis for deciding what 'descriptive terms' to include or exclude in the definition. Debates about the criteria properly governing the concept of freedom are in part debates about the extent to which the proposed criteria fulfil the normative point of the concept and in part about exactly what that point is. To refuse to bring these considerations into one's deliberations about 'freedom' is either to deny oneself access to the very considerations that can inform judgment about the concept or to delude oneself by tacitly invoking the very considerations formally eschewed.[3]

Second, we sometimes employ "freedom" in what we might call an *evaluative* sense. For instance, I might describe a farmer on fertile land as freer to grow what they want than another with less fertile soil. This statement employs the descriptive notion of freedom—it says one farmer is able to grow a wider variety than another— but adds to it the evaluative suggestion that the former farmer is the better-off; that it would be good for the farmer with poor soil to be freed from this obstacle.

Phillippe Van Parijs's work can be understood as invoking this sense of freedom.[4] For Van Parijs, all obstacles to an agent doing what they might want to do— including jumping over the moon—restrict their freedom. And all such restrictions to freedom are an evil: I would be better off without the restriction of my freedom to jump over the moon. However, on this account, to identify an agent as unfree is not grounds for condemnation of the social system that allows or causes this unfreedom. Rather, for Van Parijs social systems are to be evaluated according to (i) their ability to satisfy certain formal freedoms; (ii) the extent to which it is "lexicographically maximin" *vis-à-vis* freedom, i.e. whether it allows for the person with least opportunities to have at least as many opportunities as the person with the least opportunities under any other feasible arrangement (and where these are equal, the same for the person with the second least opportunities, and so on); and (iii) whether it provides each person the greatest possible opportunity to do whatever they might

---

[3] Connolly, *The Terms of Political Discourse*, 141.
[4] Parijs, *Real Freedom for All*, chap. 1.

want to do.[5] Here, then, an agent's unfreedom is normatively relevant; however, compared to the third level discussed below this relevance is *indirect*. All imaginable social systems entail regrettable unfreedoms in this evaluative sense, and the description of someone as unfree is therefore not sufficient to criticise or condemn the social system that enables this unfreedom. If the farmers' differential freedoms resulted not from differences in the soil but of different distributions of resources, then while we can say that the latter farmer is worse off by virtue of being less free, we are not making any condemnation of this distribution: to do so, we first must know whether this distribution is part of a system that is "lexicographically maximin" *vis-à-vis* freedom.

Third is a directly *normative* sense of freedom, in which to describe somebody as unfree is to criticise or demand justification of that which obstructs them. It is in this sense that we describe a slave as rendered unfree by their master, someone whose partner is abusive and controlling as having their freedom curtailed, or when we discuss attacks on press freedom. When we invoke freedom in this sense, we are doing more than asserting that certain descriptive features obtain or that certain people would be better off without these features (although we are also saying both these things). Rather, we are adding a normative *criticism* or *condemnation* of that which obstructs. Thus, to suggest that I am unfree to jump over the moon *in the same sense as* enslaved people are unfree would be to do violence to the concept with which the latter are described, for it is to erase the strong normative component involved in the latter description and replaces it with a merely evaluative statement.

This sense of freedom is not limited to such serious cases, however. Neither is it our only normative value, and we therefore often invoke freedom in this sense even where our final judgement is not one of condemnation. We might, for instance, describe a protestor blocking a road as restricting my freedom to drive down it, even if I support their cause and am glad they are protesting. Here we are still saying something stronger than that I would be better off otherwise. Rather, we are saying that their action requires justification: their act would be criticisable or worthy of condemnation were it not justified by their worthy cause. As David Miller puts it:

> When we say of an obstacle that it renders a person unfree to act, we make a charge that stands in need of rebuttal. Reasons have to be given for the continued presence of the obstacle. Of course such reasons may not be far to seek. Many restrictions of freedom are justified…. It is a mistake to think

---

[5] Parijs, 25.

that to describe a state of affairs as involving unfreedom is to settle a political argument; it is, however, to make a *move* in a political argument.[6]

This normative sense of freedom underlies many philosophical accounts of freedom, especially those of libertarians such as Hayek and Nozick, as well as of several "responsibility theorists" (see section 4) such as Miller, Benn and Weinstein, and Kristjánsson.[7] Further, it is this sense of freedom from which, I argue, Marx's concept of the alienation of social institutions draws its normative significance. In Marx's work, it is clear that the description of a social institution as restricting agents' freedom is *in itself* to be understood as a criticism, without qualification: he does not feel the need to add that this unfreedom, unlike others, is lamentable or criticisable. However, as we will see, the Marxist invocation of freedom differs from these various accounts in an important way, namely that it describes certain— alienated—social institutions as restricting freedom, as opposed to limiting the source of such restriction to the acts of individual agents. (There are, of course, other normative senses of freedom, motivated by different normative purposes: for instance, Connolly is primarily concerned with cases in which this purpose is to determine whether an agent can be held responsible for their act.[8] While this represents another intuitive sense of freedom, it is not that with which we are concerned.)

It is beyond the scope of this chapter to philosophically ground or justify this normative sense of freedom (although one possible source of such grounding will be suggested later). As I suggested above, I take Marx's account of the alienation of social institutions to appeal for its normative significance to a concept of freedom that he took to be widespread amongst his readers, his transformative aims requiring not an appeal to a philosophically impeccable concept but to one that might move people to act. It seems likely that Marx himself was similarly motivated by a concern with freedom; were he not, alienation's connection to this concept might instead be viewed as an instance of immanent critique, demonstrating to the members of his society that the social institutions they together create are incompatible with the values they hold.

---

[6] Miller, 'Constraints on Freedom', 69.

[7] Hayek, *The Constitution of Liberty*; Nozick, *Anarchy, State, and Utopia*; Nozick, 'Coercion'; Miller, 'Constraints on Freedom'; Benn and Weinstein, 'Being Free to Act, and Being a Free Man'; Kristjánsson, *Social Freedom*.

[8] See, for instance, Connolly, *The Terms of Political Discourse*, 167.

My goal, then, is to explicate a normative conception of freedom that reflects our intuitions, can play the role I ascribe to it *vis-à-vis* the concept of the alienation of social institutions, and is compatible with Marx's brief comments on the subject.

Alongside this distinction between descriptive, evaluative and normative senses of freedom is a commonly-cited distinction between positive and negative freedom, which is orthogonal to the previously-discussed distinction. It is often said that Marx's concept of freedom is primarily a *positive* one.[9] This draws primarily from Marx's engagement with Hegel, for Hegel's account of freedom as self-determination can be seen as concerned with positive freedom.[10] While we must not overlook the very significant differences between Hegel's and Marx's accounts, nor to ignore Marx's explicit criticisms of Hegel, there is a great deal of insight to be found from examining these connections, and I do not wish to diminish their importance. However, my focus is on that particular conception of freedom that Marx invokes when describing the unfreedom of individuals with respect to alienated social institutions, and, I take it, this is not a positive but a negative conception. Let us briefly examine this distinction.

Discussions of positive and negative freedom most frequently draw on Isaiah Berlin's essay on the subject.[11] Berlin's essay begins with a relatively clear distinction between these two concepts, but soon adds various other philosophical concepts that have, he thinks, historically been associated with positive freedom, which muddy the distinction. Let us, then, focus on an account of this distinction given by Charles Taylor, which most clearly captures Berlin's initial distinction.

Taylor characterises the distinction between positive and negative freedom as that between an "exercise-concept" and an "opportunity-concept". Positive freedom concerns whether, when an agent acts, they do so freely:

> one is [positively] free only to the extent that one has effectively determined oneself and the shape of one's life.[12]

[9] Lukes, 'Emancipation'; Walicki, *Marxism and the Leap to the Kingdom of Freedom*; Carter, 'Positive and Negative Liberty'.
[10] See Chitty, 'Hegel and Marx'.
[11] Berlin, 'Two Concepts of Liberty'.
[12] Taylor, 'What's Wrong with Negative Liberty', 213.

It is thus concerned with "the actual exercise of directing control over one's life".[13] In contrast, negative freedom concerns not what one actually does, but what one can do, i.e. with one's range of options or opportunities:

> Being free is a matter of what we can do, of what it is open to us to do, whether or not we do anything to exercise these options…. [Negative] Freedom consists just in there being no obstacle. It is a sufficient condition of one's being free that nothing stand in the way.[14]

The first sentence of this passage seems to suggest that negative freedom is concerned with agents' inability of any kind, including both external constraint and internal weakness, whereas the second and third sentences seem to suggest a concern with only external obstacles. The literature on negative freedom is primarily concerned only with obstacles that are external to the agent (although of course what this externality means is a matter for debate). Given our interest in the extent to which social institutions—which are external to agents—restrict their freedom, I will focus on "obstacles" of this kind, without suggesting that these are the only possible restrictions to freedom.

Drawing on Taylor's distinction, then, my claim that Marx's concept of the alienation of social institutions draws on a negative conception can be understood as meaning that when Marx describes a social institution as alienated, he implies that by virtue of this alienation it restricts the options available to agents (in a particular way).

There are, of course, important and interesting connections between positive and negative freedom so defined, as both Taylor and Berlin recognise.[15] Some of these connections, as well as some of the additional features that Berlin associates with positive concepts of freedom, will be explored in section 3 below.

Hereafter, unless otherwise specified "freedom" refers to a normative and negative conception of freedom. In order to clarify this conception, I will structure my discussion around Gerald MacCallum's meta-theoretical framework. MacCallum argues that all significant distinctions between competing conceptions of freedom can be understood as disagreements concerning three variables:

> Whenever the freedom of some agent or agents is in question, it is always freedom from some constraint or restriction on, interference

---

[13] Taylor, 214.
[14] Taylor, 213.
[15] Taylor, 213, 229; Berlin, 'Two Concepts of Liberty', 178–79, 204. See also Gray, 'On Negative and Positive Liberty', 509.

with, or barrier to doing, not doing, becoming, or not becoming
something. Such freedom is thus always *of* something (an agent or
agents), *from* something, *to* do, not do, become, or not become
something; it is a triadic relation. Taking the format "*x* is (is not) free
from *y* to do (not do, become, not become) *z*," *x* ranges over agents, *y*
ranges over such "preventing conditions" as constraints, restrictions,
interferences, and barriers, and *z* ranges over actions or conditions of
character or circumstance.[16]

While MacCallum's claim that this framework describes *both* negative and positive
concepts of freedom is, I think, mistaken—it cannot capture Taylor's idea that
positive freedom is an exercise- rather than an opportunity-concept—this framework
is useful for differentiating between different conceptions of negative freedom, by
making clear the variables we must specify. A complete conception of negative
freedom must give an account of who the agent of freedom is (x), what kind of actions
or qualities an agent is restricted from doing or becoming when their freedom is
restricted (z), and what kind of obstacles can restrict their freedom (y).

2 (x) The agent of freedom

Alongside the characterisation of Marx as primarily concerned with positive freedom
is a frequent suggestion that his is a *collective* conception, in which the agent is a
group such as "society"[17] or "humanity".[18] This suggestion again primarily draws its
plausibility from Marx's Hegelian inheritance. I will therefore put this theme to one
side, for our focus is on the negative conception which I shall argue underlies Marx's
concept of the alienation of social institutions.

Even when examining Marx's invocations of what I shall interpret as negative
freedom, however, one might suggest that here, too, Marx's agent of freedom is not
an individual human being. In particular, this suggestion might arise from Marx's
infamous attack on the "universal rights of man" in *On the Jewish Question*. Here,
Marx denounces these as only protecting freedom defined as:

> The right to do and perform everything which does not harm others.
> The limits within which each individual can move *without* harming
> others are determined by law, just as the boundary between two fields

---

[16] MacCallum, 'Negative and Positive Freedom', 314.
[17] Schmitt, 'Marx's Concept of Alienation', 163.
[18] See for instance Walicki, 'Marx and Freedom'; Gray, 'Against Cohen On Proletarian
Unfreedom', 85.

> is determined by a stake. The liberty we are here dealing with is that of man as an isolated monad who is withdrawn into himself.[19]

and

> the right to enjoy and dispose of one's resources as one wills, without regard for other men and independently of society: the right of self-interest. The individual freedom mentioned above … leads each man to see in other men not the *realization* but the *limitation* of his own freedom.[20]

However, this criticism must not be understood as attacking either the idea of freedom as a predicate of individuals or as negative. Rather, these passages target a *particular version* of such concepts, in which the agent of freedom is a particular *kind* of individual. Instead of taking the agent of freedom to be individuals *qua* species-beings, the "rights of man" take this agent to be individuals *qua* egoistic "isolated monads", individuals as they are encouraged to be in civil society, whose needs and desires are restricted to "self-interest". This mistaken conception of human beings, Marx argues, leads to a set of rights that are self-defeating as a framework for ensuring freedom, for it ignores human beings' social aspects, portraying these as belonging to an abstract political sphere that is unconnected to questions of individual freedom. For instance, he writes that according to these rights:

> While the '*unlimited* freedom of the press' … is guaranteed as a consequence of the right to individual freedom, the freedom of the press is completely destroyed, for 'the freedom of the press should not be permitted when it compromises public freedom'. This therefore means that the right to freedom ceases to be a right as soon as it comes into conflict with *political* life, whereas in theory political life is simply the guarantee of the rights of man, the rights of individual man, and should be abandoned as soon as it contradicts its *goal*, these rights of man.[21]

Of course, this criticism itself does not necessarily reveal Marx's own conception of freedom, for it can work purely as a form of immanent critique, arguing that these rights fail on their own terms.[22] But neither does it constitute a dismissal of all conceptions of freedom in which the agent is an individual. Indeed, in the context of

---

[19] Marx, 'On the Jewish Question', 229.
[20] Marx, 229–30.
[21] Marx, 231.
[22] Marx, 231–32.

this criticism, Marx himself seems to employ a conception of freedom of just this kind:

> This *man*, the member of civil society, is now the foundation, the presupposition of the *political* state. In the rights of man the state acknowledges him as such.
>
> But the freedom of egoistic man and the acknowledgement of this freedom is rather the acknowledgement of the *unbridled* movement of the spiritual and material elements which form the content of his life.[23]

Here, first, Marx acknowledges that these rights correctly portray individuals as the foundation of social institutions. Second, he argues that these individuals in fact have their freedom, properly construed, restricted by the "*unbridled* movement" of these social institutions. In contrast, Marx instead favours a conception of freedom that takes individuals as species-beings, for it is:

> Only when real, individual man resumes the abstract citizen into himself and as an individual man has become a *species-being* in his empirical life, his individual work and his individual relationships, only when man has recognized and organized his *forces propres* as *social forces* so that social force is no longer separated from him in the form of *political* force, only then will human emancipation be completed.[24]

Precisely what it means for individuals to require emancipation from their own "social force" in the form of "political force"—or for individuals to be restricted by the "*unbridled* movement" of social institutions—is, of course, the goal of this thesis.

As we saw in chapter 2, elsewhere Marx frequently invokes freedom in contexts in which the agents of freedom are clearly individuals. Indeed, in the *Economic and Philosophical Manuscripts* Marx describes the unfreedom of individuals as *in contrast with* the interests of their broader social group:

> The more they [workers] want to earn the more they must sacrifice their time and freedom and work like slaves in the service of avarice. In doing so they shorten their lives. But this is all to the good of the working class as a whole, since it creates a renewed demand. This class must always sacrifice a part of itself if it is to avoid total destruction.[25]

Here, there is little room to suggest that "freedom" is a predicate of groups as opposed to individuals, for individual workers' freedom is said to be in stark contrast with "the good of the working class as a whole".

---

[23] Marx, 233.
[24] Marx, 234.
[25] Marx, 'Economic and Philosophical Manuscripts', 284.

That is not to say that Marx never describes social groups as free or unfree. For instance, in *On the Jewish Question* he discusses the freedom of the state:

> the *state* can liberate itself from a restriction without man himself being *truly* free of it, that a state can be a *free state* without man himself being *a free man*.[26]

This reflects a common practice of describing groups as free or unfree in a way that is not reducible to the freedom of group members.[27] However, it seems that this practice primarily employs either the descriptive or evaluative, and not the normative, senses of freedom. Indeed, I take it that normative conceptions of freedom *necessarily* are, or are derived from, conceptions of freedom in which the agent is an individual. This is because, as I will argue below, there is a necessary connection between the normativity of a claim concerning freedom and the having of desires. While we do often invoke freedom in a normative sense in talking about groups—as when we demand the freedom of a colonised country, for instance—I take it that in such cases we are using such language as either shorthand for, or to describe a phenomenon derived from, the individual freedoms of the people that make up the group in question. This mirrors the senses in which we might talk of the desires of a group, as either shorthand for the desires of the group members or to refer to the "desires of the group" understood in terms of the Searlean account of collective Intentionality outlined in chapter 3, where such group desires are constituted by collectively-attributed desires in the minds of individuals in conditions of common knowledge.

This is not to say that the conception of freedom we develop must be "individualistic" from root to tip. Indeed, I will claim in section 5.3 that the conception of freedom we are concerned with is precisely "social" in an important sense. Nevertheless, the key point here is that however social this conception turns out to be, it is still *my*—or *your*—freedom, as an individual, that is at issue.

MacCallum briefly discusses another way in which this *x* variable might differ from those conceptions discussed above, namely by being *smaller* than individuals. On some accounts the agent of freedom is the 'true self', consisting of an individual's deepest, most strongly held, or most rational beliefs and desires.[28] This theme,

---

[26] Marx, 'On the Jewish Question', 216.
[27] Hindriks, 'The Freedom of Collective Agents'.
[28] MacCallum, 'Negative and Positive Freedom', 324.

crucial in the history of theories of freedom, will be discussed in section 3, where we consider one role that such sub-human agents have historically played.

3 (z) Freedom to what?

We now turn to MacCallum's *z*: when we are free or unfree, what are we free or unfree *to do*?[29] Implicit in my discussion above was that a normative conception of freedom cannot plausibly hold that obstacles to *all* possible actions restrict freedom. We saw this in the suggestion that gravity's preventing me from jumping over the moon cannot be described as rendering me "unfree" in a normative sense. Can we capture some of this distinction by focusing on what kinds of action are obstructed in each case?

Perhaps the most intuitive answer draws on agents' *desires*: obstacles to freedom prevent us from doing *what we want to do*. Many accounts of freedom, especially amongst the earlier literature on the subject, endorse this position. For instance, Mill says that "liberty consists in doing what one desires",[30] while Voltaire says "When I can do what I want, there is my freedom".[31]

Let us sketch an account based on this response in relation to three acts that I might perform, *X*, *Y*, and *Z*. If I desire both to *X* and *Y* but not to *Z*, then an account that maintains the connection between freedom and desire holds that an obstacle to *X*-ing or *Y*-ing restricts my freedom, while an obstacle to *Z*-ing does not. Such a concept is clearly a negative concept of freedom: it is concerned with the opportunities available to me, here my opportunity to either *X* or *Y*. However, it is not concerned with obstacles to any possible action: obstacles to that which I do not want to do, such as *Z*-ing, do not restrict my freedom. I take it that this account not only reflects our normative intuitions, but partly explains these: insofar as we take desire-satisfaction to be a good, that which prevents us from doing so—and in this sense renders us unfree—is criticisable on this basis.

---

[29] MacCallum includes in this variable "doing, not doing, becoming, or not becoming something". I will focus only on doing, which I take to be primary.
[30] Mill, *On Liberty*, 146.
[31] Cited in Parijs, *Real Freedom for All*, 18.

There is, however, a popular objection to this answer.[32] Often referred to as the "contented slave" objection, this points out that if an agent is unfree when unable to do as they desire, then they can become free *either* by removing the obstacle *or* by removing the desire, and that this leads to counterintuitive results. The eponymous "contented slave" example provides one such result: a slave who no longer desires any of the things they are prevented from doing—perhaps as a result of indoctrination or adaptive preferences—, or who desires their continued enslavement, cannot be described by such a conception as unfree, and since enslavement is a textbook example of an unfree situation this conception necessarily fails. That is, it intuitively seems that I am sometimes not free even if I am only restricted from *Z*-ing: if, say, I once had a desire to *Z*, but when faced with the obstacle to *Z* simply abandoned this desire, our intuitions are that I am still rendered unfree by this restriction, and the account described above cannot account for this intuition.

This is a powerful objection, and any conception of freedom must have a response. There are two possible responses.

The first is to detach freedom from desire. For instance, Berlin's essay replaces desires with the obstruction of "what I could otherwise do":

> If I am prevented by others from doing what I could otherwise do, I am to that degree unfree; and if this area is contracted by other men beyond a certain minimum, I can be described as being coerced, or, it may be, enslaved.[33]

Similarly, Miller writes that "The extent of a person's freedom depends on the number of potential actions of his that are obstructed",[34] while Kristjánsson defines "freedom without any recourse to actual *wants* or their significance. A constraint on freedom is something that impairs a possible choice, whether we now (or ever) want to make the particular choice, and irrespective of how important it might seem to us".[35] On this account, an obstacle to any of acts *X*, *Y* or *Z* can restrict my freedom.

---

[32] See, for instance, Parijs, 18; Carter, 'The Concept of Freedom: Part I: Background, Methodology, and the "Negative-Positive" Debate'; Gray, 'On Negative and Positive Liberty', 521.

[33] Berlin, *Liberty*, 169; As is well documented by Arneson, '"Freedom and Desire"', Berlin is not consistent in this approach.

[34] Miller, *Market, State, and Community*, 29–30.

[35] Kristjánsson, *Social Freedom*, 99.

However, absent additional caveats, this response has serious problems. First, on this account it seems that any obstacles that leave me with the same number of available opportunities are neutral with respect to freedom. But we might wonder whether freedom is really a primarily quantitative concept, that it is the *number* of available opportunities that matters. As Nancy Hirschmann puts it, according to this approach "if there are only two options, one of which is the one I want, I would seem to be less free than if what I want is not available at all amidst dozens of other options",[36] which seems counterintuitive. Similarly, this response implies that an obstacle that closes down one opportunity but opens up an alternative one is neutral *vis-à-vis* freedom. Again this runs counter to our intuitions, as anyone offered a rail-replacement bus can attest. Finally, it seems this conception is unable to describe many intuitively unfree situations as unfree, and thus narrows the range of obstacles that restrict freedom beyond recognisability. Without appealing to an agent's preferences, it seems we cannot describe a threat of violence as restricting freedom: while your threat/offer to shoot me in the leg if I leave the room might obstruct my ability to leave the room while not suffering a severe pain in the leg, it replaces this with the opportunity to leave the room and suffer such a pain.[37] Without taking into account my desire not to have a sore leg it seems we are hard-pressed to account for our intuition that such threats restrict freedom.

The authors cited above as taking this first response by and large recognise these difficulties, and so caveat this approach with an additional argument, one which in fact takes us to the second possible response.[38]

The second response is to differentiate between agents' attitudes or desires in some way, such that, for instance, my lack of desire for Z, or the slave's desire for their continued enslavement, can be in some sense discounted. Discounting my lack of desire for Z would make the obstacles to my doing Z once again count as restricting freedom, thereby overcoming the contented slave objection.

---

[36] Hirschmann, *The Subject of Liberty*, 6.

[37] let us assume that you are the only person I know with a gun and that only by being shot can I experience this kind of sore leg.

[38] The second response differentiates between desires in order to discount some. Some proponents of the first response use similar arguments to discount particular options because it would be irrational to desire them. See Miller, *Market, State, and Community*, 40; Benn and Weinstein, 'Being Free to Act, and Being a Free Man', 195. Cf. Steiner, 'Individual Liberty', who does not take this approach.

There are several possible approaches available to writers who want to deal with the contented slave objection in this way. The first differentiates between desires with reference to some external or "objective" standard, such as rationality or human nature. For instance, John Gray writes that:

> Only by invoking some norm of human nature which is discriminatory as to the wants which are to be counted, and which includes evaluations of the agent's states of mind, can the intuition that the wholly contented slave remains unfree be supported.[39]

The second differentiates between desires with reference to a conception of *autonomy*. John Elster encourages this approach when he writes that "Being a free man is to be free to do all the things that one autonomously wants to do".[40] Elster, however, does not suggest a methodology for characterising autonomous desires. We might assume that he means to invoke some of the resources employed in the philosophical literature on autonomy. While this literature is primarily concerned with questions of positive freedom—with the question of whether, when an agent acts motivated by a desire, this is sufficient to say that this act was free—we might read Elster's claim as being that the resources used for differentiating between desires in this literature can be employed in questions concerning negative freedom.

There are at least two broad approaches suggested by the literature on autonomy. The first, advocated by Gerald Dworkin and Harry Frankfurt among others, identifies different *levels* of desires, or different ways agents might relate to their desires.[41] To take Taylor's example, consider a person who recognises that their spiteful feelings and reactions are undermining a valued relationship. Such an agent has desires motivated by their spiteful feelings, and a desire to not have such feelings, and these conflict. In their more reflective moments, Taylor argues, such an agent might recognise that the latter desires are "higher" or more significant for them, and as such recognise that they act freely when they act on these and not when they act on the spiteful desires. A second approach, advocated by John Christman, focuses our attention not on an agent's current relation to their desires, but on their origin.[42] This broad approach has been taken up in the literature on negative freedom by Kristján Kristjánsson, who suggests that when my desires are caused by

---

[39] Gray, 'On Negative and Positive Liberty', 521.
[40] Elster, 'Sour Grapes - Utilitarianism and the Genesis of Wants', 227–28.
[41] Dworkin, 'The Nature of Autonomy'; Frankfurt, 'Freedom of the Will and the Concept of a Person'; Taylor, 'What's Wrong with Negative Liberty'.
[42] Christman, 'Liberalism and Individual Positive Freedom', 346.

"brainwashing" or hypnosis, they are not "my own", and are thus not desires with respect to which I am free. On his account, characteristic of these phenomena are their external origin:

> If the bar is of a natural origin … it restricts the person's ability but not [their] freedom. If it has its origin in another agent, it is simply on a par with any other externally created bar, although it happens to be internally situated. Thus, there is no morally relevant difference between preventing a man from entering a room by locking it up or by hypnotising him into staying away from it.[43]

Each of these forms of this second approach to the contented slave objection raise concerns, and, while I cannot attempt to engage with any of them adequately, it is worth sketching some of these briefly. The first, objective-standard account raises a clear worry about paternalism: who are we to tell the slave of the example that their desire for the continuation of the *status quo* does not count because it is not in keeping with an objective standard of human nature or rationality? And once such a methodology has been adopted, can we not portray almost any situation as one in which agents are free once we impute to them certain rational or human desires of our own concoction?[44] The account that focuses on autonomous desires *qua* those the agent relates to in a particular way largely bypasses such concerns, being based on agents' own ranking of desires. But this internalism limits its applicability: what if the contented slave's contentment is so absolute that no "higher" desires exist to appeal to? It also raises problems of definition: how does an agent determine which desires are higher than others? Finally, the account that appeals to desires' origins has limitations of scope. Examples such as brainwashing and hypnosis are clear-cut cases of the external creation of desires. But is the suggestion that *only* such examples are cases of discountable desires? If so, the account would be particularly restricted, and would not respond to the force of the contented slave objection. Further, it assumes some methodology for distinguishing between externally- and internally-generated desires, but matters are rarely so simple. Some desires—such as for food—might seem clear examples of internally-generated desires. But what of my desire for a sandwich rather than horse meat? All our desires are formed in conversation with our social context, and are rarely free of external influence. The significance of this is highlighted particularly acutely by Hirschmann, who

---

[43] Kristjánsson, *Social Freedom*, 100. Kristjánsson is an author who takes the first response qualified by one akin to the second.

[44] See especially Day, 'On Liberty and the Real Will'.

emphasises the extent to which the "social construction of the self" complicates questions concerning freedom.[45]

I cannot here attempt to deal with these problems, nor to suggest an adequate methodology for addressing the contented slave objection. Indeed, it seems that resolving these questions requires considerably more than a neat philosophical distinction can bear: our mental lives and the range of influences on us seem too complex for any straightforward account. This is highlighted well by Hirschmann, whose discussion of freedom in the context of domestic abuse, welfare dependency and religious veiling emphasise the complexity of these questions. Indeed, Hirschmann's work also exemplifies what is surely the only plausible approach to such questions: in-depth empirical and theoretical analysis that addresses not only the detailed psychology of actual people living through such situations, but also the varied external factors that structure the environment in which they develop themselves, the options available to them, and how they learn about these options and interpret these through discourse. On this account, there is no philosophical quick-fix for the contented slave objection, for there is no prototypical contented slave: to address such questions adequately would involve not only speaking to actual enslaved people, considering their point of view with a seriousness that philosophical invocations of them rarely do, but also a broad-ranging analysis of the social structures in which slavery exists and how these shape and are shaped by the desires of enslaved people.

It seems, then, that we must maintain the connection between freedom and desire, given the very serious problems raised by its abandonment. And this raises considerable difficulties, which must be addressed head-on rather than with any simple philosophical distinction. However, for our purposes this need for further work need not stifle our attempts at developing a general conception of freedom. In a Searlean mood, we might say that this approach to borderline cases might reasonably be set aside to be dealt with once we have clarified our basic concepts. For surely there are more than enough obstacles in the contemporary world to the satisfaction of desires about which we feel no tension or ambiguity, and the social institutions that Marx describes as alienated are surely among them. I will therefore adopt the position that the concept of freedom concerns obstacles to individual agents

---

[45] Hirschmann, *The Subject of Liberty*.

doing *what they want*, with the caveat that this approach leaves much to be resolved at its edges.

4 (y) Restrictions to freedom

4.1 Contemporary philosophical accounts of restrictions to freedom

MacCallum's y variable concerns what kinds of obstacles to an agent's action constitute restrictions to their freedom. This is the question on which the philosophical literature is most divided. In what follows I will argue that there is a way of formulating the most intuitively satisfying answer to this question that this literature, while often appealing to it in passing or obliquely, has largely overlooked. I will also argue that it is this question to which the concept of the alienation of social institutions relates.

Before discussing what obstacles restrict agents' freedom, we must ask what an "obstacle" is. I take it that all external restrictions to freedom involve obstacles, but not all obstacles restrict freedom. If I am rendered unfree to do something (by something external), there is an obstacle to my doing it of a certain kind. Some writers on negative freedom, in particular Hillel Steiner,[46] hold that an obstacle in the relevant sense must make the action in question *impossible*. This definition corresponds well with our intuition that it seems strange to say of some agent that they were unfree to do something which they nevertheless, in fact, did, to which all competing definitions seem vulnerable. Nevertheless, this definition has significantly more counterintuitive consequences, and as such must be rejected. In particular, it suggests that threats are not obstacles and thus cannot restrict freedom: your threat to shoot me if I leave the room does not make it impossible to leave the room, and so cannot restrict my freedom.[47] Indeed, perhaps even imprisoning me does not restrict me freedom on this definition, if there is the faintest chance of escape.

---

[46] Steiner, 'Individual Liberty'.

[47] Day, 'Threats, Offers, Law, Opinion and Liberty', 258 counters by suggesting that such a threat makes it impossible to conjunctively leave the room and not be shot. However, all possible actions make some such conjunction impossible. As such taken seriously Day's position ends up much like the irrelevance-of-size view described below. See Kristjánsson, *Social Freedom*, 44.

In opposition to this view, Kristján Kristjánsson proposes—with an account that he also attributes to Miller (and Oppenheim)—the *irrelevance-of-size view*. This view is not spelled out in great detail, but can be seen as the claim that anything external that makes the performance of an act more difficult or costly can be classified as an obstacle to that action. On this account, the key question is not so much what counts as an obstacle, but rather which of the many obstacles that agents face restrict their freedom. I will adopt this approach. Let us, then, turn to this latter question.

The broadest possible account holds that *any* obstacle to an agent's action restricts their freedom. The account that comes closest to this might be that of van Parijs, who takes

> the broadest possible characterization of freedom-restricting obstacles consistent with the view that lacking freedom is being prevented from doing some of the things one might want to do. Abstracting for the time being from the time dimension, any restriction of the opportunity-set is relevant to the assessment of freedom.[48]

As I suggested above, this broad characterisation of freedom-restricting obstacles is perfectly valid insofar as it captures the broad, evaluative use of the word "freedom". However, as I also suggested, this broad characterisation is incompatible with the normative sense of freedom invoked by Marx, amongst many others. There are many obstacles to me doing what I want that do not restrict my freedom in a normative sense: consider, for instance, the mountain range blocking my path, or my love's indifference towards me. On this normative conception, the term "unfree" entails that an obstacle is criticisable or requires normative justification, and making the distinction between such obstacles and others—even others which the agent would be better off without—is this conception's very *point*. As Benn and Weinstein put it, this "moral and political concept of freedom cannot be stretched to cover every case in which it is linguistically appropriate to speak of 'being free', without hopelessly attenuating it".[49] Another way of putting this is that this conception delineates those obstacles from which we demand to be liberated, and as Berlin dramatically puts it, with a definition of freedom-restricting obstacles this broad "Total liberation in this sense (as Schopenhauer correctly perceived) is conferred only by death".[50]

---

[48] Parijs, *Real Freedom for All*, 23.
[49] Benn and Weinstein, 'Being Free to Act, and Being a Free Man', 198.
[50] Berlin, 'Two Concepts of Liberty', 186.

Many accounts therefore attempt to differentiate between obstacles in some way. One common approach holds that obstacles that are human-made—for which other human beings are causally responsible[51]—restrict freedom, while those that are not (and which are frequently called "natural"[52]) do not. For instance, Berlin tells us that "You lack political liberty or freedom only if you are prevented from attaining a goal by human beings".[53] Similar positions are held by Nozick,[54] Miller,[55] Oppenheim,[56] and Day.[57] These authors generally recognise the insufficiency of this account on its own and supplement it in various ways, but it is worth considering in its own right before adding additionally complexity.

There is an appeal to this way of distinguishing between obstacles. First, it draws attention to our hopes for a normative conception of freedom, indeed for any normative concept at all. By directing our focus towards human action and obstacles created thereby, this distinction highlights the extent to which normative concepts can be effective when addressed to phenomena that are responsive to them: describing a mountain as freedom-restricting in a normative sense would appear to be a waste of breath, unless one takes there to be a reason-responsive mechanism governing the universe. However, this does not tell quite the whole story, for a normative conception might address itself not only to the obstacle, but to the agent obstructed by it: an agent confronted with a natural obstacle which they had not previously seen as restricting them might, by means of the description of their situation as one of unfreedom, come to see the obstacle differently and be motivated to overcome it. Second, this distinction divides the various examples introduced above broadly according to our intuitions: the slave's enslavement or somebody's abusive partner—as obstacles for which human beings are causally responsible—are deemed to restrict their freedom, while the mountain that blocks my path or the gravity that prevents me from jumping over the moon—as "natural" obstacles—are not.

However, this distinction is inadequate, for two reasons. First, it is clearly too broad: employing this distinction, every possible human action could be portrayed as

---

[51] Oppenheim, '"Constraints on Freedom" as a Descriptive Concept', 306.
[52] e.g. Oppenheim, 305; Miller, *Market, State, and Community*, 31.
[53] Berlin, *Liberty*, 169.
[54] Nozick, *Anarchy, State, and Utopia*, 262.
[55] Miller, 'Constraints on Freedom', 68.
[56] Oppenheim, '"Constraints on Freedom" as a Descriptive Concept', 305–6.
[57] Day, 'On Liberty and the Real Will', 180.

restricting the freedom of another agent. My eating a sandwich restricts the freedom of others to eat that sandwich. And—using an example adopted from Robert Nozick—my choosing a husband would restrict the freedom of other suitors to marry me. Second, it is not obvious that no naturally-occurring obstacles can restrict my freedom in a normative sense. Consider the following scenarios, drawing loosely from Miller's discussion.[58]

(1) I cut down a tree which falls and blocks a road.

(2) The wind knocks down a tree which blocks a road.

On the account described above (1) clearly restricts drivers' freedom, while, as a merely "natural" event, (2) does not. However, our intuitions on the matter are rather more flexible than this account would suggest. Consider the following situations:

(3) The wind could knock the tree down only because I, whose land the tree grew on, did not adequately maintain the soil, diminishing the strength of its roots. I have not caused the tree to fall by my actions, but I have not prevented it from falling.

It at least seems more plausible in this situation than in (2) to say that the drivers' freedom is, in an intuitively normative sense, restricted by my negligence.

(4) The wind knocks the tree down, and I am parked next to it in a tow truck. Feeling particularly antisocial this morning, I refuse to move the tree.

Here, the feeling that the drivers' freedom is restricted reaches something like the strength of that for case (1).

Scenario (3) suggests that not only my causal responsibility in the sense of my action having caused an obstacle to exist, but also in the sense of my omission enabling it to exist, is sufficient to restrict others' freedom.[59] Further, scenario (4) suggests that not only the *origins* of an obstacle, but also the reasons for its persistence, are pertinent to whether it restricts freedom.[60] The problem with the inclusion of these criteria, however, is that it exacerbates the first problem discussed above: our definition of freedom-restricting obstacles becomes so broad as to exclude almost

---

[58] Miller, *Market, State, and Community*, 32–35.

[59] Miller, 32–33.

[60] Miller, 'Constraints on Freedom', 71–72; Kristjánsson, *Social Freedom*, 89.

nothing, indeed nothing other than natural events that could neither have been prevented nor could have been removed.

Matters are complicated further by Marx's observation that what are merely "natural" obstacles under a certain level of technological development become decidedly less "natural"—and closer to the situation described in example (4)—once new technology becomes available. More will be said of this below. For now, I will focus on how to distinguish between human-made obstacles that do and do not restrict freedom, for it is human-made phenomena with which the concept of the alienation of social institutions is concerned.

A common way to address the objection that a conception of freedom that focuses on human-caused obstacles results in too broad a definition of freedom-restricting obstacles is by specifying that freedom-restricting obstacles are those that are *intentionally* caused by human beings: freedom-restricting obstacles are those that were caused by an intentional action in which (part of) the intention was to cause that obstacle. Miller explicates this position as the view that "I am unfree when and only when the obstacle that prevents or deters me from doing X was placed there deliberately by some other person or persons".[61] Felix Oppenheim is perhaps the most explicit advocate of this position,[62] but is far from alone in endorsing it. For instance, Berlin states that, amongst other things, "The extent of my freedom seems to depend on … how far they [the possibilities open to me] are closed and opened by deliberate human acts"[63] (although he elsewhere makes a slightly different argument, discussed below). Similarly, Friedrich Hayek defines liberty with reference to coercion, and this in terms of being subject to the will or ends of another agent:

> Our definition of liberty depends upon the meaning of the concept of coercion…. By "coercion" we mean such control of the environment or circumstances of a person by another that, in order to avoid greater evil, he is forced to act not according to a coherent plan of his own but to serve the ends of another.[64]

---

[61] Miller, 'Constraints on Freedom', 31–32.
[62] Oppenheim, '"Constraints on Freedom" as a Descriptive Concept'; Oppenheim, 'Social Freedom'.
[63] Berlin, 'Two Concepts of Liberty', 177 footnote 1.
[64] Hayek, *The Constitution of Liberty*, 71.

Following Hayek, Robert Nozick also emphasises coercion, in which "the will of another is operating or predominant".[65]

This definition drastically reduces the number of obstacles that are defined as freedom-restricting. For instance, on this account, my eating a sandwich is no restriction to your freedom, unless I ate the sandwich with an intention to prevent you from doing so. And similarly for my choosing of a husband or my inadequate arboreal care. It also has a convenient consequence for these authors. As Oppenheim puts it:

> It leaves outside the range of constraints on freedom not only physical or psychological inabilities and natural obstacles but also incapacities caused by anonymous demographic or economic or institutional conditions.[66]

And as Nozick and Hayek famously argue, other ills stemming from such "economic conditions"—including the exploitation of workers by capitalists—are similarly excluded as possible targets of criticism as freedom-restricting. These implications appear, of course, counterintuitive to many. And it is not only these exclusions that are counterintuitive: Miller, for instance, emphasises the counterintuitive result that somebody obstructed by a fence is said to be either unfree or free depending on the intentions of whoever built it.[67] However, our methodology in this chapter prevents us from appealing to these intuitions: whether such obstacles as the market restrict freedom belongs to those difficult borderline cases to which the more refined conception of freedom might be applied once derived from cases about which there is broader agreement. What else, then, can be said with respect to this distinction?

If the objection that this definition of freedom-restricting obstacles is too narrow is unavailable to us, the opposite objection—that it is, as with the conceptions outlined above, too broad to capture our intuitions—appears clear. Does anything somebody does that intentionally limits my being able to do as I please constitute a normative restriction to my freedom? Does my locking of the bathroom door, carried out with the intention of preventing you from walking in on me, limit your freedom in a normative sense? It would seem counterintuitive to answer these questions in the affirmative.

---

[65] Nozick, 'Coercion', 38.
[66] Oppenheim, '"Constraints on Freedom" as a Descriptive Concept', 306.
[67] Miller, *Market, State, and Community*, 29.

Hayek and Nozick clearly accept this objection, for both attempt to narrow their accounts of freedom-restricting obstacles even further to accommodate it, in various ways. In Hayek's work, agents' actions that intentionally create obstacles for others but which are in keeping with (certain kinds of) the law do not restrict others' freedom: only intentional acts that are counter to law do so. In Nozick's *Anarchy, State and Utopia*, this narrowing instead involves natural rights: intentionally created obstacles that are in keeping with their creators' natural rights are not restrictions to others' freedom. In Nozick's earlier paper 'Coercion', the distinction is spelled out instead with reference to acts that constitute *offers* and those which are *threats*: threats restrict my freedom, whereas offers do not. Threats and offers are both conditional proposals: if you do (do not) *x*, I will (will not) *y*. The distinction, on Nozick's account, relies on how *y* relates to certain standards, of which he suggests two. According to one, my conditional proposal is a threat if *y* makes you worse off relative to "the normal or natural or expected course of events"[68]; according to the other, it is a threat if *y* makes you worse off relative to the "morally expected course of events".[69] (This latter standard can be seen as mirroring the emphasis on natural rights in his later work: indeed, Nozick apparently described 'Coercion' as an intellectual ancestor of that book.[70]) He argues that our intuitions are that each of these standards accords better with our intuitions in different situations.

Clearly, the argument that pre-existing circumstances—what happens to be the law, or what course of events, given our experiences, we have come to take as "normal"—or a "natural" set of rights, which determine a fixed "morally expected course of events", are capable of playing the role of reference point to which our intuitions concerning freedom can refer is not one to which Marx could appeal: it is precisely these pre-existing circumstances—and most centrally those property rights that Hayek and Nozick are primarily concerned with—that Marx's invocation of freedom targets, and the notion of natural rights—again, in particular the natural right of property which is Nozick's central concern—is one which Marx explicitly pillories.[71] Further, Marx's objection to these notions is based on arguments that many would accept, and even for those who do not, the idea that specifically the concept of freedom and the intuitions we have concerning it depend on such controversial

---

[68] Nozick, 'Coercion', 24.
[69] Nozick, 27.
[70] Haworth, 'What's so Special about Coercion?', 377.
[71] Marx, 'On the Jewish Question', 233–34.

notions might appear strange: must I really have worked out the legality of every action that led to my being unable to do what I wish before I can have an opinion on whether I am unfree? And must I take an interlocutor to have in mind an identical set of natural rights to those I believe in if I am to share with them a conception of freedom?

To highlight this approach's counterintuitive nature, consider highway robbery, which literature on coercion often takes for its intuitive baseline. In being held up for "your money or your life", our intuitions strongly suggest that you are rendered unfree. In standard conditions this intuitive result is reproduced by each of Hayek's and Nozick's conceptions. First, the robber's act runs counter both to the law and to the "natural rights" that Nozick specifies. Second, the robber's threatened action—taking your life—is, relative both to the normal course of events and to what is morally expected, worse for you. The obstacle that their act presents to your keeping your money is thus said to restrict your freedom. Nevertheless, given a different context, each of these conceptions can be made to produce counterintuitive results. I will here focus on Nozick's account in 'Coercion', for it is the most clearly explicated and the most influential. Let us first consider his definition of threats that invokes the "normal course of events". As Alan Haworth has suggested, in certain places and historical periods highway robbery was a commonplace. We might imagine, indeed, a time when such robbery was so commonplace that the normal course of events was that at least one robbery would occur on any journey. So commonplace might such an event be that its quasi-inevitability comes to be built into the plans of any traveller. Given this baseline, according to Nozick's conception the highway robber's intervention can no longer be interpreted as a threat—and thus as a restriction on freedom—but instead simply as an offer, akin to the offer of employment which it is Nozick's interest to portray as compatible with freedom. Similarly with the version involving the *morally* expected course of events: as Haworth puts it, "it is no doubt possible to advance an arguable case that highway robbery was a legitimate response to poverty and deprivation, and to that extent not morally unexpected".[72] In such not unimaginable scenarios, Nozick's distinction produces importantly counterintuitive results.

---

[72] Haworth, 'What's so Special about Coercion?', 387.

"Responsibility theorists" of freedom such as Miller offer an alternative way of narrowing down the range of human-caused obstacles that restrict agents' freedom.[73] For Miller, an obstacle to my acting restricts my freedom if and only if it is something for which another agent is *morally responsible* (whether they are morally responsible for having created the obstacle or for it remaining in place). On this view, the range of obstacles that restrict freedom is far wider than it is for the likes of Nozick. Not only obstacles caused by acts performed with the intention of so causing them, but also obstacles caused by negligent activity or obstacles that an agent has failed to remove despite having a moral obligation to do so, all count as restrictions to freedom on Miller's view.[74]

I take it that again this conception is not one available to Marx, for whom such fixed notions of moral responsibility would seem to be both rather abstract and likely to reflect the standards of dominant social relations. Further, this account is somewhat awkward for our purposes, for while we have a fairly clear intuitive conception of what acts individuals are morally responsible for, when it comes to social institutions' obstruction of our desired activity, matters become considerably more complex. Miller is keen to emphasise that his account leaves space for collectively caused obstacles, for which no one individual is entirely causally responsible, to be described as freedom-restricting.[75] However, this still requires the identification of particular "individuals who contributed to the outcome intentionally or in dereliction of duty",[76] and when it comes to the social institutions that are our central concern it seems that our intuitions are divided concerning whether anybody, and if so whom, bears such moral responsibility. Insofar as our aim is to derive a conception of freedom from those areas in which there is relative intuitive unanimity which we can then apply to more controversial questions, it seems that this account will not take us very far, for its application requires an account of moral responsibility concerning which there is at least as much controversy.

Taking a step back, however, the overall conception that responsibility theorists such as Miller defend does broadly reflect many of our intuitions concerning freedom. If

---

[73] Miller, 'Constraints on Freedom'; Miller, *Market, State, and Community*. Cf. Kristjánsson, *Social Freedom*.
[74] Miller, 'Constraints on Freedom', 72.
[75] Miller, 81.
[76] Miller, 80.

this conception is inadequate for our purposes, an alternative account would do well to explain the intuitive appeal of this conception.

4.2 Freedom and alterability

I would like to propose an alternative account of how to differentiate between obstacles that are freedom-restricting in the normative sense and those which are not. This account centres on the extent to which an obstacle is *alterable by human agency*: when an individual (*x*) is prevented from doing as they desire (*z*) by an obstacle that is alterable by human agency (*y*), they are unfree in the normative sense. Obstacles to an agent's desired activity which are inalterable are ineligible for criticism by means of this conception.

The obvious ambiguities in the key term "alterable by human agency" in this definition will be explored below. First, let us consider its virtues. Perhaps the most unusual feature of this conception of freedom is its *forward-looking* nature. We saw above that many conceptions of freedom differentiate between freedom-restricting and non-freedom-restricting obstacles according to their *origin*, for example whether they originate in intentional human action or arose "naturally". The "responsibility theorists" such as Miller and Kristjánsson partly distance themselves from this position insofar as what counts is the cause of an obstacle's origin *or* of its persistence. Nevertheless, both approaches are fundamentally *backward-looking*. When Miller, for example, is concerned with the possibility of removal of an obstacle, this is at some point in the past: what matters is that some agent's past act is responsible for the obstacle preventing me from doing something, whether this be a positive act that created the obstacle or an omission that failed to remove it.[77] In contrast, the conception that I am suggesting is concerned not with an obstacle's origin or past possibilities of removal, but with the extent to which it is—here and now—alterable.

The forward-looking nature of this conception perhaps explains its normativity. As I suggested above, the normativity of the concept perhaps derives from its connection to desires: insofar as we take desire-satisfaction to be a good, that which prevents it is criticisable on this basis. Yet this character is shared with the evaluative sense of

---

[77] Miller, *Market, State, and Community*, 71–72.

freedom invoked by Van Parijs. What we can now add is that the increased strength of the normative sense of freedom, according to which the naming of some obstacle as freedom-restricting is necessarily to criticise it or demand justification for it, derives from its concern with only those obstacles that there is any point in criticising: criticism in this sense can be seen as a demand for change, and such a demand can be made only of that which is changeable.

While, as I suggested above, there is no clearly defined conception of freedom in Marx's work, he does seem to invoke something like this conception of freedom in some passages. For instance, in the *German Ideology* he and Engels say:

> Thus, in imagination, individuals seem freer under the dominance of the bourgeoisie than before, because their conditions of life seem *accidental* [*Zufaellig*]; in reality, of course, they are less free, because they are more subjected to the violence of things.[78]

In the context of this passage, *Zufaellig* refers to the fact that these conditions seem to be the result of mere chance as opposed to being determined by "historical" and "social relationships". Here, then, Marx and Engels are saying that it *seems*, "in imagination", that agents are free because their obstacles appear to be the product of mere chance and thus as inalterable by human agency, whereas in fact they are unfree because these obstacles are the product of social relations which are alterable.

While this distinction between alterable and unalterable obstacles has not been incorporated into the explicit accounts of freedom examined above, many of these authors do allude to it occasionally. For instance, Berlin, who we saw above in his 'Two Conceptions of Liberty' defined freedom as depending on, amongst other things, "how far they [the possibilities open to me] are closed and opened by deliberate human acts",[79] in a later paper came to recategorize this as a sub-type of unfreedom (named "oppression"), with the broader category being defined with reference to the obstacle's alterability:

> absence of such freedom is due to the closing of such doors or failure to open them, as a result, intended or unintended, of alterable human practices[80]

John Gray emphasises this theme in interpreting Berlin, for whom, he says, "the relevance to questions of liberty of social arrangements no-one has designed and

---

[78] Marx and Engels, 'The German Ideology', 78–79.
[79] Berlin, 'Two Concepts of Liberty', 177 footnote 1.
[80] Berlin, *Liberty*, 32.

which accord with no-one's intentions is determined by their alterability or remediability".[81] Benn and Weinstein, too, invoke this idea when they argue that:

> It is a general feature of events or conditions that can be accounted reasons for saying that a person is not free in respect of a certain kind of action, that unlike the tides they are not natural, unalterable, given, but rather that some rational being (or beings) can be held responsible for them, for bringing them about, or for allowing them to continue…. By extending the range of restrictive conditions judged capable of alteration, the concept of freedom can itself be extended.[82]

(Note, however, that here Benn and Weinstein present a false contrast between that which is unalterable and that for which "some rational being (or beings) can be held responsible", rather than the more adequate pair "unalterable"/"alterable".) Similarly, Connolly in one passage defines constraints in terms of being "limited with respect to important actions or goals by circumstances potentially alterable at a less than prohibitive cost through individual or social action".[83] Finally, J.P. Day's discussion of the connection between freedom and ability, according to which I can be rendered unfree to do only that which I am otherwise able to do, can perhaps be interpreted in this light. For instance, Day argues that someone's freedom to walk somewhere is restricted when they are imprisoned or have their ankle broken, but not when their legs are cut off, for in the latter case walking is no longer something the agent is able to do, "the difference resid[ing] in whether the inability is irretrievable or not".[84]

While these authors, then, occasionally invoke the distinction between alterable and unalterable obstacles to give plausibility to their conceptions of freedom, they do not incorporate it into their "official" accounts. Rather, these accounts focus on backward-looking causes of obstacles that more or less closely track their alterability. However, an obstacle's origin or the fact that it could, in the past, have been prevented or removed never corresponds perfectly to whether it can, now, be altered.

I would like to suggest that the intuitive appeal of the conceptions of freedom examined above stems from their coincidence with the alterability distinction, while

---

[81] Gray, 'On Negative and Positive Liberty', 522.
[82] Benn and Weinstein, 'Being Free to Act, and Being a Free Man', 199.
[83] Connolly, *The Terms of Political Discourse*, 167.
[84] Day, 'Threats, Offers, Law, Opinion and Liberty', 264.

their implausibility stems from their divergence from it. I will focus on Miller's account, which I take to be the most plausible from the accounts reviewed above.

As we saw above, Miller takes an obstacle to restrict my freedom if it is one for which another agent is morally responsible. To demonstrate that this account captures our intuitions, Miller encourages us to consider a thought experiment, in which somebody is left trapped in a room by a range of unfortunate events, against which we are asked to test our intuitions regarding whether they are rendered unfree. Most of the situations he describes only with reference to the cause or origin of this entrapment: in some, a strong wind blows the door closed; in others it is closed by an agent Y with the intention of entrapping them; and in others, the wind blows the door closed and Y, whose job it is to check that nobody is trapped in any of the rooms, neglects to do so. From the point of view of the account I am developing, these descriptions do not give us the most relevant information (namely, the extent to which A's entrapment is alterable), and so comparing the two accounts is difficult. However it seems that in most situations, portrayed appropriately, the two accounts give similar results. When I am trapped in a room because the wind blew the door shut, Miller suggests I am not unfree, but simply unable, to leave, because nobody is morally responsible for my entrapment; and with the information provided in this example, it seems too that we have no reason to suppose that in such a situation my entrapment is alterable—there is no mention of other agents who might open the door—so on my account, too, this would not seem like a situation in which I am unfree in a normative sense. Where I am trapped because Y intentionally closed the door, Miller argues that I am unfree because my entrapment is something for which Y is morally responsible; and on my account, it seems that similarly I am rendered unfree because it is alterable: Y can as easily open the door as close it.

However, one of Miller's scenarios does explicitly provide more information about the alterability of the situation, and by focusing on this scenario we can see the counterintuitive nature of Miller's analysis. The scenario is the following.

> The wind blows the door shut. At 6.30 P.M. I call to a passerby to unlock the door, but the passerby, who knows about Y's duties, is busy and pays no attention.[85]

On Miller's analysis, since the passerby is not morally responsible for people trapped in rooms, in this situation my freedom is not restricted: I am merely rendered unable

---

[85] Miller, 'Constraints on Freedom', 71.

to leave the room. But this seems odd. Certainly, if I were trapped in a room and a stranger was outside with the ability to open the door but was refusing to do so, I would be likely to feel that my freedom was being restricted in a distinctly normative sense. Indeed, it seems like it is precisely situations like this in which the normative sense of freedom is appropriate. Here, it seems that the specification that the passerby is not morally responsible makes little difference to our intuitions concerning whether or not I am unfree.

Of course, Miller might respond by suggesting that this difference in intuition is simply due to different intuitions concerning moral responsibility: perhaps I, in contrast to Miller, simply think that people are morally responsible for helping strangers trapped in rooms. But this only highlights the problem with his approach discussed above: our intuitions about moral responsibility in many scenarios are significantly more divided than are our intuitions about freedom.

It seems, then, that we might account for the persuasiveness of Miller's account by pointing to the extent to which it often runs in parallel with the alterability criteria: alterable obstacles are often those for which we hold others morally responsible. This coincidence is particularly close where the obstacle we are discussing is some agent's behaviour itself: if your tapdancing is stopping me from sleeping, then my freedom to sleep is restricted. You could easily stop tapdancing, and yet you do not. And, *because* you could stop at any time, I hold you morally responsible for it. Contrast this with a situation in which your heavy breathing is what is keeping me awake: I do not hold you morally responsible precisely because there is nothing you can do about it. The appeal of Miller's account, then, seems to derive from our intuition that agents are morally responsible for those actions over which they have some choice or intentional control. Acts that agents carry out without any such choice are, according to most accepted conceptions of moral responsibility, those for which they are not morally responsible. This is captured in the legal notion of duress. Thus, where my activity is blocked by the ongoing activity of another for which that other is morally responsible, we take this continued activity to be alterable, and thus to be a restriction on my freedom. But where the alterability of an obstacle and an agent's moral responsibility for that obstacle come apart, as in the example above, it seems that our intuitions are on the side of the alterability account.

(This is not to say that we do not have normative intuitions concerning cases in which somebody is morally responsible for an obstacle which is now unalterable. Here,

however, it seems our intuitions derive from the fact that somebody is morally responsible for my unfreedom in an *evaluative* sense: I would be better off without this restriction, and you are morally responsible for my situation. This is comparable to the intuitions we have concerning the fact that somebody is morally responsible for me being in pain. But that is different to saying that the normative sense of freedom with which we are concerned—according to which to label an obstacle as a restriction to my freedom is to criticise and demand change of it—hangs on this being the case.)

One problem with my proposed account, of course, is that the notion of "alterability by human agency" is multiply ambiguous. I will here address two crucial ambiguities and a possible objection, before moving on to explaining the role that the concept of the alienation of social institutions can play in relation to this conception.

The first ambiguity pertains to one particular kind of obstacle that these terms can be applied to, namely those created or constituted by human actions. We saw above that underlying a number of conceptions of freedom is the idea that human actions, in contrast to natural phenomena, are not determined or fixed, and that it is this alterability that not only enables conceptions such as moral responsibility to come into play, but also which gives the plausibility to the intuitive idea that obstacles created or constituted by them restrict freedom. However, it seems that some human actions and the attitudes upon which they are based are not as fluid and alterable as others, and it seems that where such attitudes are found, action motivated by them no longer intuitively restricts freedom. A hungry person arriving late to a restaurant and joining the back of a very long queue of other hungry people does not, it seems, have their freedom restricted in a normative sense by these people. In such a situation, the hungry person is more likely to criticise their own lateness than the fact that other hungry people are queuing. If, in contrast, the queue consists not of other hungry people but of the cast of an eccentric piece of performance art, it seems plausible that the hungry person *would* consider their freedom to be restricted, and this is so precisely because we consider the motivations behind the queue to be more alterable in the latter case than the former.

Second, the terms I have employed above are ambiguous regarding temporality. What is inalterable today may not be so in the future, and what was inalterable in the past is not so today. Some obstacles are currently inalterable and to that extent

"necessary", but—as Marx put it—this represents "a merely *historical* necessity".[86] John Elster attempts to capture this point by distinguishing between "historical feasibility" and "physical feasibility", where the latter captures what is necessary across all time, and employs the latter in his normative conception of exploitation.[87] This distinction, however, does not get us very far for our purposes, for the list of phenomena which are ahistorically, "physically" inalterable is very short indeed: there are few that we have not imagined a world without in science fiction literature, and that we cannot imagine being overcome in some distant future. Rather, I take it that when we employ the concept of freedom we are always operating with a relative, "historical" notion of inalterability.

This temporal relativising, of course, has precedent in Marx's work, most obviously in his account of history. Under a certain mode of production—i.e. given a certain level of technological development—certain ways of living are said to be unavoidable, and certain economic structures therefore portrayed as inalterable. For instance, given conditions of scarcity, an economy constituted by competition is unavoidable. As such, criticism of the institutions of a competitive economy as restricting freedom in a normative sense is implausible in such periods: one might as well criticise the 19th Century British Government for not providing sufficiently fast broadband.

> so long as the productive forces are still insufficiently developed to make competition superfluous, and therefore would give rise to competition over and over again, for so long the classes which are ruled would be wanting the impossible if they had the "will" to abolish competition and with it the state and the law.[88]

Once these productive forces have changed, however, the historically relative inalterability of such institutions changes. While Marx does not apply this idea explicitly to a conception of freedom—again, Marx gives relatively little by way of an explicit account of freedom—it does not seem far-fetched to suggest that these ideas can be combined such that the change through history in whether a particular institution is alterable or inalterable changes our ability to criticise it as restricting freedom. That is, we might suggest that for Marx capitalism only restricts freedom in a normative sense at the point at which it becomes humanly possible to replace it.

---

[86] Marx, *Grundrisse*, 831.
[87] Elster, *Making Sense of Marx*, 200–201.
[88] Marx and Engels, 'The German Ideology', 329–30.

One likely objection to the account I have presented is the following. According to it, I am presently not unfree to travel to Mars in a normative sense (assuming that doing so were something I desire), for we are not able to alter my inability to do so. However, were technological innovations to remove the physical barriers to my doing so, such that I could feasibly—at very great expense—travel to mars, it would seem that I am instantly rendered unfree in a normative sense. Further, this would be true of everyone who similarly desired to travel to mars. This result appears counterintuitive: we cannot seriously think that any society that does not allow for universal space travel as soon as this is technologically possible restricts our freedom in this sense. (I have heard no complaints that NASA restricts most Americans' freedom by only enabling a tiny minority of citizens to fly to the moon, which is presently technologically possible.)

This objection can be addressed as follows. Travelling to Mars requires more than the necessary technological knowledge. In addition, it requires others' cooperation, and—at least in our society—social institutions that govern the distribution of labour and material resources. As such, once the technological obstacles are overcome, the obstacles to my journey to Mars that remain are these social institutions and the lack of others' cooperation. As such, to determine whether my inability to travel to Mars constitutes a restriction to my freedom we need to know whether these social institutions and these others' cooperation are alterable or not. It is to this topic that we now turn.

5 Freedom, social institutions, and alienation

The purpose of developing a conception of freedom in this chapter has been to enable us to reconstruct the normative significance of the concept of the alienation of social institutions. This concept describes a process that occurs to social institutions, and its normative significance too is directed at these: to describe a social institution as alienated is to criticise it, the social institution, and in particular—I have claimed— to criticise it as restricting freedom. We need, then, to consider how social institutions themselves can restrict individuals' freedom. Given the conception of freedom described above, this requires answering two questions.

First, exactly how do social institutions present obstacles to agents doing what they want? Certainly, many social institutions do seem to present such obstacles. Yet

precisely how they do so is rather unclear. Social institutions primarily affect our activity either through the activity of other individuals or through our expectations of such activity: why, then, would we talk of freedom being restricted by the social institution as opposed to by these individuals themselves? To properly understand how social institutions can restrict an agent's freedom, we must be more specific about how they might prevent individuals from acting as they want, and how this prevention relates to the role played by other individuals within these institutions. I will explore these questions in section 5.1.

Second, we must address the relevant sense of alterability by human agency with which such a conception can operate. There appears to be a sense in which all social institutions are alterable: certainly, many that once appeared solid have subsequently melted into air. Yet there is another sense in which social institutions seem necessarily inalterable *vis-à-vis* any particular individual. How, then, are we to assess social institutions' alterability? I will explore this question in section 5.2. I will claim that it is in relation to this hurdle that the connection between freedom and the concept of the alienation of social institutions lies: this latter concept provides a means for differentiating between (relatively and historically) alterable and inalterable social institutions. Since according to this account all and only alienated social institutions are (relatively and historically) alterable, all and only alienated social institutions that stop individuals from doing what they want to do restrict their freedom in the relevant sense.

5.1 Social institutions as obstacles to individuals' desired activity

To understand how social institutions might present obstacles to individuals doing what they want, we must again translate our discussion to the level of individual agents and their relations, using the framework developed in chapters 3 and 4.

Take a situation in which a social institution exists such that (a significant proportion of) people in a group, G, are acting in accordance with the perceived deontic powers constitutive of a perceived status, S, motivated by attitudes that represent S in some way. Let us focus on agent X, who is part of group G and who desires to D. Say that perceived status S involves a perceived obligation for X to perform some act that is incompatible with D-ing. This act might simply be the act of not D-ing, or might be some act the performance of which is incompatible with D-

ing. For simplicity's sake let us assume the former: X has a perceived obligation to not D.

Now, that some social institution exists amongst G such that, according to the perceived status involved, X has a perceived obligation to not D does not by itself present any obstacle to X's D-ing. As Searle's repeated example about a group of Australian chancers suggests, that a significant proportion of some group collectively accepts (in any sense) some status has no necessary relevance to me and the options available to me: I might simply not care that this institution exists. In the Searlean terminology outlined in chapter 3, this social institution might give me no *reason* to act in any particular way, because I might lack any relevant *motivator*. It is only in situations in which I have some relevant motivator that facts concerning social institutions can be reasons for me to act or not act in particular ways, and thus affect whether or not I can, given my attitudinal make-up, do what I want to do. In asking how the social institution in question might affect, and thus present an obstacle to, X's ability to D, then, we must take into account not only the social institution, but also X's (current and past) attitudes, including their relevant motivators.

As I suggested in chapters 3 and 4, agents can have different kinds of reasons for acting, in particular *desire-dependent* and *desire-independent* reasons, differentiated by the different kinds of motivators involved. Let us consider situations in which X has a reason to not D of these different kinds separately.

As we have seen, to say that X has a desire-independent reason to not D is to say that they actually have a rational obligation to not D, i.e. they have in the past rationally committed themself to not D-ing by means of performing some speech act of committing acceptance. This obligation is itself a motivator. If we suppose that a rational agent would find it psychologically difficult or costly to act counter to their rational commitments—and I take it that the Searlean framework that we have adopted involves such a supposition in the claim that agents are motivationally responsive to their rational commitments—then it seems that in such a situation there is an obstacle to X D-ing, namely the fact that they have a reason not to D by virtue of their obligation.

However, while X thus has an obstacle to D-ing, in the circumstances just described it seems that this obstacle is *not* the social institution. Rather, the social institution simply coincides with X's own self-created obligation, and it is this obligation that presents the obstacle. (As an aside, it seems that according to the conception of

freedom outlined above, *X* is not rendered unfree by this obstacle, for—as I suggested in chapter 3—obligations, in contrast to other kinds of rational commitment, are not alterable: *X* cannot change the fact that they, in the past, performed a speech act that rationally committed them to not *D*-ing in the future.)

Of course, the fact that a social institution exists might have been *involved* in some sense in *X*'s self-creation of such an obligation. As I argued previously, *X*'s performance of the relevant speech act might have been socially grounded, i.e. might have been motivated by *X*'s belief that others were similarly undertaking such obligations, and this belief might have been grounded in *X*'s descriptive recognition that this social institution exists: *X* might have interpreted others' institutional acts as acts of committing acceptance, and this might have grounded *X*'s own such acceptance. We might here be tempted to ask whether in such circumstances *X*'s freedom might be restricted where this interpretation of others' acts is mistaken, however this question is not one I can address here.

It seems, then, that in cases where the perceived obligations involved in a social institution give *X* a desire-independent reason not to *D*, we cannot talk of the social institution being an obstacle to *X*'s *D*-ing: the only such obstacle is their own self-created obligation. However, in some situations, we can imagine an agent who has such an obligation being prevented from *D*-ing not only, or not even primarily, by their obligation, but *also* because this obligation is reinforced by a social institution. I may see my option of not repaying my debt to you to be curtailed by my obligation to do so; but in cases where my rational recognition of this commitment is insufficient to foreclose this option (say, by being in tension with a very strong desire to spend the money), I might still see this option as limited also by the social institution surrounding such debts, as spelled out below. *X*'s having of a desire-independent reason not to *D* therefore does not imply that the coincident social institution presents no obstacle at all to their *D*-ing: such desire-independent reasons can exist alongside other ways in which *D* is made more difficult or costly. This suggests that whenever we can talk of a social institution itself being an obstacle to *X*'s *D*-ing, we are primarily concerned with desire-dependent reasons, even in cases where these exist alongside desire-independent ones.

Let us now turn to cases in which *X* has a desire-dependent reason to not *D* which renders *D*-ing more difficult or costly. What kind of motivators—beyond, obviously, being desires—are we here concerned with? Since *X* desires to *D*, it seems the

relevant motivator cannot simply be a desire to not D. (Agents might, of course, have such conflicting desires. I am here presupposing a level of rational consistency that excludes such possibilities for the purpose of analysis.) What other relevant motivators might X have that would render D-ing more difficult or costly?

In chapter 4, I suggested a few such motivators in passing. For instance, I mentioned a desire not to stand out making the fact that others all act in a particular way into a desire-dependent reason to act similarly. And I mentioned a desire not to be rebuked or sanctioned making the fact that acting in certain ways will lead to me being rebuked or sanctioned into a desire-dependent reason not to act in those ways. There are likely many more such motivators, however I will here focus on these two.

First, let us consider a situation in which X's relevant motivator is a desire not to stand out. Given this additional desire not to stand out, we can see how the fact that the social institution exists in group G makes it costly for X to not D, and is thus an obstacle to D-ing. This is because it is entailed by the fact that the relevant social institution exists that a significant proportion of group G are acting in accordance with the perceived deontic powers constitutive of perceived status S, such that "not standing out"—and thus not suffering the cost of their desire not to stand out going unmet—requires X to similarly so act, i.e. to act in accordance with these perceived deontic powers, i.e. to not D. (With respect to our discussion in chapter 4 concerning the necessary proportion of members of a group that must be acting in accordance with a perceived status for us to say that the institution exists in that group in the sense that it gives members of G reason to act in accordance with its perceived deontic powers, we can here see that where the kind of motivator involved is a desire not to stand out this proportion must be roughly a majority. If less than a majority of people are so acting, then a desire not to stand out would motivate me to not act accordingly. Of course, the framing of what the relevant group is is here important.)

The example of social institutions preventing X from D-ing by virtue of X's desire not to stand out is perhaps a rather abstract example, because it does not describe any direct relations between X and any other specific individuals.[89] While Marx often talks about agents' unfreedom *vis-à-vis* social institutions, it is also clear that this unfreedom is primarily manifested in particular kinds of power relationships between individuals, most obviously—in the case of the social institutions of

---

[89] Cf. Searle, *Making the Social World*, 155–60.

capitalism—that between workers and capitalists. To begin to get a more concrete idea of how such relationships fit into my account, let us consider the second kind of motivator described above, the desire not to be rebuked or punished by others.

The perceived deontic powers involved in some social institutions are such that agents are thought to have obligations not only to act in certain ways, but also to perform acts of rebuke or punishment against other members of the group. For instance, let us imagine a particularly obligation-heavy social institution of queuing, according to which not only does each agent have a perceived obligation to join the back of the queue rather than the front, but also has a perceived obligation to rebuke others for trying to push in front of them. If I am in such a queue and somebody pushes in front of me, I have a perceived obligation to say "no, join the back of the queue". If I instead accept payment from them for keeping quiet, I would be in breach of the perceived deontic powers of this social institution (for which there might be third-level obligations on agents to rebuke me, and so on).[90]

Let us consider a situation in which the social institution that exists in group G is of this kind. Thus, the perceived deontic powers constitutive of status S are such that X has a perceived obligation to not D, and every group member has a perceived obligation to rebuke X should they D. From the presupposed fact that the social institution exists in G, we know that (a significant proportion of) members of G are acting in accordance with the perceived deontic powers constitutive of the status, and we suppose that X knows this. As such, X has good reason to believe that (a significant proportion of) people in group G will continue to so act, and thus to believe that if X Ds, they will be rebuked (because rebuking X for D-ing is one of the acts to which the other members are perceived to be obligated). In such a situation, given X's desire not to be rebuked, the social institution's existence makes it costly for X to D—D-ing would cost them the satisfaction of their desire not to be rebuked—, and is therefore an obstacle to X's D-ing.

In such a situation, we can consider the relation between the social institution as an obstacle to X's D-ing, and the acts of the other individual group members. As we have seen, the social institution makes D-ing costly for X by virtue of X's desire not to be rebuked and the fact that a significant proportion of members of G would rebuke X for D-ing. So, are we not better off saying that it is the dispositions of other members

---

[90] Mechanisms of this kind are discussed in Elster, 'Rationality and Social Norms', 119.

of *G*, and not the social institution *per se*, that presents an obstacle to *X*'s *D*-ing? Indeed, there might be a particular individual, *Y*, who *X* takes to be the person that will rebuke them: can we not then say that it is *Y*, and not the social institution, that makes *D*-ing costly for *X*?

To see why it is often more accurate to say that the social institution makes *D*-ing costly, rather than that *Y* does so, let us return to the example of the queue. I might be tempted to push in front of Dave. Knowing the perceived obligations involved in the queue, I know that if anyone will rebuke me for pushing in, it will be Dave. And I might not care at all what Dave thinks of me, and be quite willing to do things that displease him or lead to him scolding me verbally. Nevertheless, I take it that there is a significant difference between, on the one hand, displeasing Dave and being scolded by him, and, on the other, being *rebuked* by him. When we say that Dave *rebukes* me, we are saying more than that he scolds me: we imply that in saying "no, join the back of the queue" he is in some sense acting as a representative of the social institution or of the group as a whole, that he is invoking the perceived deontic powers involved in the social institution. While there might be cases in which I do not push in simply because I wish not to be scolded by Dave, it seems that in many situations my desire is not of this kind, but is instead the quite different desire not to be *rebuked* by him (or others). We might compare this difference to that between a police officer giving an order *qua* police officer or *qua* friend: depending on what role we take them to be performing in the giving of an order, we either understand it as an order of a representative of *the police force*, or as a (significantly less persuasive) order given by a friend.

We cannot here explore the further question of the kinds of reasons for which an agent might desire specifically not to be rebuked as opposed to being, say, scolded by Dave. As I suggested in chapter 3, in some situations this might be to do with wanting to be recognised by others—by the members of the group as a whole—as a rationally consistent agent: if *X* actually has an obligation to not *D*, then they might not want their *D*-ing to be drawn attention to for this kind of reason. In others, it might be that some act I desire to perform can be performed only assuming the cooperation of the group or some proportion of it, but which would be little affected by annoying Dave alone: if the queue were to enter a tango competition with randomly assigned partners, and I know that nobody will tango with me if they know I jumped the queue, then I will desire not to be rebuked considerably more than I

will desire not to annoy Dave alone. In yet other situations, it might be to do with the combined physical power of the group as opposed to the physical power of the individual rebuke-giver: I might take Dave's rebuke as a threat of physical coercion by the combined force of the group, knowing that others will physically enforce the perceived deontic powers of the institution, as opposed to the significantly smaller threat of physical coercion by Dave himself. Whatever the reason, I take it that desires of this kind—desires that are specifically concerned with the group—are often part of the motivation for agents to act in accordance with the perceived deontic powers involved in social institutions, even in cases where the obstacle in question involves an act carried out by an individual that represents this group.

I take it that something like this underlies our intuitions about the extent to which individual capitalists obstruct workers' ability to do what they want to do, as well as Marx's focus on social institutions over individual capitalists (discussed below). Marx famously does not focus his normative criticism on individual capitalists for their actions. Indeed, their actions are said to be restricted by the social institutions of capitalism just as are those of workers.[91] Nevertheless, it seems that in threatening a worker with unemployment, demanding overtime, etc., capitalists do obstruct workers. Here, we can see more clearly what is going on. Capitalists draw their ability to make workers do what they do not want to do from the social institutions of capitalism: just as Dave's rebuke is a rebuke rather than an individual's expression of dissatisfaction, so are capitalist's actions invocations of the social institutions of which they are a part, breaches of which will have consequences for workers that go significantly beyond the ire of the capitalist alone. If, then, we are to determine whether such actions restrict workers' freedom, we need to ask *not* whether these actions themselves are alterable (of course, any individual capitalist might decide at any time to give away their capital to workers), but rather whether the social institutions of which they are a part are alterable.

5.2 Alterability

According to the conception of freedom I have given above, an individual's freedom is restricted when they are rendered unable to do what they want to do by an obstacle

---

[91] e.g. Marx, 'Economic and Philosophical Manuscripts', 301.

that is alterable by human agency. In section 5.1, I described how social institutions can present obstacles to an agent's doing what they want. I will now explore the sense in which social institutions might be said to be alterable by human agency, and thus the extent to which they make such agents unfree. I will argue that social institutions that are alienated, and only these social institutions, restrict agents' freedom.

Our task in discussing the alterability of social institutions is particularly complex, for our relevant intuitions are heavily affected by our methodological point of view. On the one hand, we can interpret social institutions from a macro point of view, from which we view the enormous range of possible alternatives to the social institutions that we have, and ask what is stopping us from adopting any of these instead if we so wished. Here our question is "could *we* alter this institution?", and the answer invariably "yes". On the other hand, we can interpret social institutions from a micro point of view, from the standpoint of any particular individual. Here our question is "could *I* alter this institution?", and the answer (almost) always "no".

Clearly, neither point of view will help us apply the conception of freedom I have developed to social institutions, because depending on which point of view we take it appears that either all social institutions are alterable or none are, and this defeats the very purpose of talking of freedom *vis-à-vis* social institutions. If, as per the macro view, all social institutions are alterable and thus freedom-restricting, then either we had better try living without social institutions or abandon the goal of being free and find some other way of evaluating social institutions such that we can promote some over others. And if, as per the micro view, social institutions are all inalterable and thus no restriction to freedom, we would again require some alternative means of evaluating them.

However, from the point of view of a Searlean social ontology that translates social institutions to the level of individual agents, their motivations, and their relations, we can develop a more useful way of discussing social institutions' alterability. I would like to suggest that the concept of the alienation of social institutions that I have reconstructed in previous chapters can be seen as one particular way in which social institutions can be categorised as alterable or inalterable, and thus as freedom-restricting or not. That is, alienated social institutions are—in a sense to be elaborated below—alterable, and are thus able to restrict freedom, while non-alienated social institutions are—in this sense—not.

Of course, this concept is a rather blunt tool, and given the complexity of what is involved in social institutions and the number of variables involved, many alternative ways of making this distinction might be found. Indeed, we have already seen that any social institution can be described as alterable from at least one point of view. I will therefore talk of the concept of alienation as differentiating between social institutions according to their *relative* alterability: in contrast to non-alienated social institutions, alienated ones are relatively alterable (because of that by virtue of which they are said to be alienated). Nevertheless, I take it that this distinction is sufficient to explain many of our intuitions *vis-à-vis* the freedom-restricting nature of certain social institutions, and in particular to explain Marx's connection between the concepts of alienation and freedom.

In chapter 4, the translation of social institutions to the level of individual agents and their relations involved focusing on the motivating reasons for which each member of a group that acted according to its relevant perceived deontic powers did so such that their activity together constituted the social institution in question. From this point of view, I argued that the distinction between alienated and non-alienated social institutions can best be portrayed as a distinction between social institutions that exist only by virtue of activity motivated by certain kinds of *mistaken beliefs* and those that do not. These mistaken beliefs, I argued, constitute cases of agents mistakenly treating the social institution in question as independent, either of their own activity, of a sub-group of which they are potentially a part, or of the activity of the group as a whole. As I put it there, a social institution is alienated if a) a proportion of group members are motivated to act according to the relevant perceived deontic powers only because of mistaken beliefs of a certain kind, and b) the fact that this proportion so act makes it the case that a sufficient proportion of group members so act that the institution is able to affect others' behaviour in a particular way.

Understood in this way, I will now argue, we have intuitively plausible grounds for the claim that alienated social institutions are relatively alterable, and therefore restrict individuals' freedom, whereas non-alienated social institutions are (relatively) inalterable and thus do not.

The intuitive ground for the relevance of the concept of alienation to the question of the alterability of social institutions follows from my discussion above concerning the extent to which other individuals' actions restrict my freedom. There I suggested

that we intuitively differentiate between those attitudes of our neighbours—such as their desires for food—which are (again, relatively) inalterable, and others which we think of as alterable. I take it that amongst this second category the least controversial are *mistaken beliefs*. It is a widespread assumption that our neighbours are oriented to the truth. Certainly, this is a presupposition of the Searlean rational framework with which we have been operating, and which in the introduction to section 2 I argued broadly in favour of. Given this widespread assumption, we commonly expect that demonstrating the falsity of others' beliefs will lead them to amend them. Regardless of the truth of this assumption, its widespread acceptance explains the intuitive sense that mistaken beliefs are alterable. The concept of alienation, then, which describes certain social institutions as existing only by virtue of certain kinds of mistaken beliefs, draws on these intuitions to portray alienated social institutions as alterable, and thereby as restrictions on the freedom of those agents whose desires are restricted by them.

Of course, it *is* sometimes possible to change others' desires, too. (Indeed, it is also possible to change others' *true* beliefs through manipulative falsehoods). And in practice many mistaken beliefs prove to be intractable. Complexities of this kind reinforce our caveat that we are concerned with *relative* alterability. Furthermore, certain relatively inalterable motivations will motivate different kinds of activity in different historical periods with different technologies and cultural contexts. For instance, to use Marx's example mentioned above, in conditions of scarcity my inalterable desire for survival might necessarily motivate me to act in a competitive manner in which I need not in conditions of surplus. As such, we must also talk of social institutions as merely *historically* alterable or inalterable: they are designated as such only here and now. Nevertheless, this relative and historically-specific conception captures an important intuition concerning the alterability of social institutions. When social institutions exist only because of others' mistaken beliefs, it seems that these social institutions are unnecessary and alterable, preventing me from doing what I want to do for no good reason. It is this intuition that I take the concept of the alienation of social institutions' connection to a conception of freedom to be appealing to.

5.3 Clarifications and objections

The role that I have attributed to the concept of alienation in enabling a conception of freedom to be applied to social institutions is somewhat unusual, and will initially present the reader with a number of concerns which I will attempt to address in this section.

First, the reader may find curious the claim that some social institutions are described as non-alienated, and thus as being no restriction to freedom, by virtue of not depending for their existence on certain kinds of mistaken beliefs. Does not this claim imply that non-alienated social institutions require an unrealistic and unobtainable degree of perfect knowledge? And does this not paint the concept of non-alienation as hopelessly utopian?

To respond to this objection we must note that, according to the account of alienation given in the previous chapter, non-alienated social institutions need not be completely free of mistaken beliefs. This is so in two senses. First, on our definition an alienated social institution *exists because of* the activity motivated by the relevant mistaken beliefs, i.e. were we to correct these mistaken beliefs the institution would no longer exist. This leaves room for a social institution to not be alienated despite being partly constituted by mistaken-belief-based activity, i.e. in cases in which the number of agents whose institution-constituting activity is motivated by mistaken beliefs is small enough that correcting these would leave the institution intact amongst the group. Second, the concept of alienation I outlined is quite specific about the *kinds* of mistaken beliefs involved in alienated social institutions. In contrast to these particular kinds of mistaken beliefs, there are many other beliefs agents may have which are mistaken but which do not *per se* result in a social institution constituted by activity motivated by them being alienated. For example, take an ancient society in which most people believe the earth to be flat, and who therefore create social institutions devoted to ensuring that people do not fall off the edge of the planet. This belief, while clearly—we now know—mistaken, may nevertheless be described as unavoidable or inalterable given the existing level of scientific knowledge: as such, while such beliefs may not be absolutely inalterable, they are reasonably described as historically inalterable. Further, though, a similar institution created by a group of people living in *our* society who share this same belief—one which now appears distinctly avoidable—may not be criticised as *alienated* according to the concept I have developed (though their beliefs may, of

course, be described as irrational), for such a belief is not of the kind described in the previous chapter as an instance of mistakenly treating an institution as independent. This result may seem intuitively justifiable given that we assume this kind of mistaken belief, maintained in the face of such strong evidence to the contrary, to have deep social or psychological causes which render them more difficult to amend: as such, the criteria of rationality appears a more suitable model of critique than does that of alienation or freedom.

Second, the reader might object to the account I have outlined as follows (which is in effect a translation of a common objection to writers such as Miller and Nozick who attempt to provide a normative conception of freedom[92]). The concept of alienation described in the previous chapters leaves open the possibility of a huge range of social institutions that we intuitively consider to restrict freedom to nevertheless be formed in a non-alienated way, and hence to be described as not restricting freedom. For instance, we might imagine a social institution of slavery to exist without any mistaken beliefs of the kind I have described at all. Does not such a social institution still, in a thoroughly normative sense, restrict the freedom of those it enslaves?

The short response to this objection is that yes, of course, such a social institution involves the restriction of the freedom of those it enslaves. However, this is so in a quite different sense to that in which agents might be made unfree by a social institution in the sense I have described above. As I suggested in section 5.1, when we are concerned with agents' freedom *vis-à-vis* social institutions, rather than with their freedom in relation to the acts of any particular individuals, we focus on a particular kind of the desire-dependent reasons that make doing what they want to do difficult or costly, namely those that are such reasons by virtue of desires concerning the group as a whole such as not standing out or not being rebuked. In our intuitive characterisation of situations of slavery,[93] however, the desires relative to which enslaved agents have reason to do what they are told are characteristically not of this kind, but are rather more direct desires concerning the avoidance of physical force by their enslavers. Therefore in such situations, our question is not whether the social institution is alterable or inalterable, but whether that individual enslaver's behaviour is so alterable – and our strong intuition that enslaved people

---

[92] Oppenheim, '"Constraints on Freedom" as a Descriptive Concept', 305; Cohen, *Self-Ownership, Freedom, and Equality*, 60.

[93] I am not here making any historical claims about the attitudes of enslaved people in actual situations of slavery.

are unfree suggests that we take such behaviour to be distinctly alterable. It seems, then, that in prototypical situations of slavery, enslaved agents are unfree *vis-à-vis* their enslavers *qua* individuals or sub-groups of individuals, rather than by the institution of slavery *per se*. (Of course, in actual social institutions of slavery, slave owners often have the physical force necessary for their practices only by virtue of collective acceptance of their status as owner. The argument here is meant to counter the specific objection outlined above—that in a hypothetical non-alienated social institution of slavery, slaves are still unfree—rather than to make any claims about a fundamental difference between slavery and other social institutions.)

That is not, of course, to say that the threat of physical force cannot be involved in individual agents' unfreedom *vis-à-vis* social institutions, too. As I suggested above, if I desire not to be rebuked by an individual because I take this rebuke to be backed by the potential physical force of the group as a whole rather than by my individual rebuke-giver, it is the social institution which creates an obstacle, and not necessarily the individual (whose isolated physical force might be insufficient to sway my decision). For instance, we might portray the social institution of private property as being underpinned by the threat of force in precisely this way. I might refrain from taking what I want from a shop not because of a fear that the shopkeeper will hurt me if I do so, but rather because of a fear that the social institution of private property as a whole is such that I will be punished for doing so, including potentially by the physical force employed by the police whose ability to employ such force in turn relies on public acceptance. While physical force might thus still be central to the prevention of me doing what I want to do, in comparison to the example of slavery this physical force is at a remove.

This distinction, I take it, can be clearly seen in Marx's critique of capitalism. For Marx, capitalism replaced a mode of social organisation that was based on direct, often physical, coercion. Marx did not develop a detailed normative critique of such direct coercion—perhaps, as many of us now do, he saw such direct coercion as so intuitively problematic that such an account was unnecessary, or perhaps it was simply not of interest because he saw it as no longer definitive of his social world— but instead focused his critique on what replaced it. With capitalism, such direct, physical, and easily-identifiable coercion was replaced by a different mode of social organisation in which the labour of many was organised for the benefit of a few in much the same way as before, but in which this was achieved without direct physical

coercion—indeed, which was achieved on the basis of market mechanisms which its advocates described as based on freedom. The purpose, then, of Marx's concept of alienation was to enable criticism of social institutions as freedom-restricting *despite* not being based on—or, perhaps better, *even if* they were not based on—such direct and physical coercion.

In Marx's words, then, his work—and, I claim, his concept of the alienation of social institutions—criticises not primarily the individuals who benefit from particular social relations, but these "relations" themselves, in which such individuals act.

> In the present epoch, the domination of thinglike [*sachlichen*] relations over individuals, and the suppression of individuality by fortuitous circumstances, has assumed its sharpest and most universal form, thereby setting existing individuals a very definite task. It has set them the task of replacing the domination of relations and chance over individuals by the domination of individuals over chance and relations. It has … called for liberation from a quite definite mode of development. This task, dictated by present-day relations, coincides with the task of organizing society in the Communist way.[94]

This comes out perhaps even more clearly in his later work, for instance:

> [Capitalism] is distinct from the preceding form in that surplus-labour needs no longer be performed in its natural form, under the direct supervision and compulsion of the landlord and his representative; the direct producer is driven rather by the force of relations than by direct coercion, through legal enactment rather than the whip, to perform it on his own responsibility.[95]

Of course, the fact that Marx-inspired criticisms of capitalism criticise the institutions themselves rather than the behaviour of individuals within them— criticise "the game itself" rather than how particular participants play it—is, perhaps, a truism.[96] However, in the account I have given above I hope to have concretised this claim, and to have shown *how* such a criticism can plausibly be portrayed as based on a concept of negative freedom.

Finally, the account I have given above is unusual in particular because, according to it, the freedom of an agent with respect to a social institution is dependent partly on the actions and attitudes of the other individuals whose activity constitutes that

---

94 Marx and Engels, 'The German Ideology', 438, translation amended.
95 Marx, *Capital: Volume III*, 930–31, translation amended.
96 For instance, see Jaeggi, 'What (If Anything) Is Wrong with Capitalism?', 59; Cohen, *Self-Ownership, Freedom, and Equality*, 50; Haworth, 'What's so Special about Coercion?', 388; Benn and Weinstein, 'Being Free to Act, and Being a Free Man', 203.

institution, and whose activity is in turn affected by it. Whereas my being forced to work by an institution of slavery can be described as a situation in which I am unfree simply by virtue of the behaviour of my individual enslaver, whether or not a social institution which presents obstacles to my desired action renders me unfree or not cannot be determined simply by looking at me and my direct environment alone. Rather, to determine whether or not the existence of a social institution renders me unfree or not, we must first determine whether or not that institution is (historically and relatively) alterable in the sense described above, and to do so requires us to examine the kinds of attitudes that motivate other agents' relevant activity. In this sense, we might say that we are here concerned with a form of "social" freedom. Nevertheless, unlike some other conceptions of freedom that are described as "social", this conception—as we saw above—is one in which the agent of freedom remains an individual agent.

In providing an account of the normative basis underlying Marx's concept of the alienation of social institutions, I have attempted to show that this basis both appeals to the intuitions of members of the contemporary world and can be derived from the conceptions propounded by its most sophisticated and explicit defenders. This concept claims that certain social institutions restrict our freedom, and thus ought to be changed, and does so not on the basis of an idiosyncratic and uniquely Marxist normative conception, but on the very concept of negative freedom that animates so much of contemporary political discourse. Further, by showing the structure of this critique, I hope to have shown that we can see the concept of the alienation of social institutions as able to characterise and so evaluate a far broader range of social institutions than those to which Marx applied it.

Conclusion

I have argued that underlying many of Marx's invocations of *Entäußerung* and *Entfremdung* is the idea that many of the social institutions of the contemporary world are alienated. This is, at base, the idea that, while these social institutions are constituted by our activity, many of them have become independent of us—in particular, of our decision or control—and to this extent restrict our freedom. I have attempted to clarify this basic idea by translating the social-level terms employed in its expression to the level of individual agents and their relations, using John Searle's work on social ontology as a starting point. The concept of the alienation of social institutions can be spelled out in three steps:

(1) Group of agents *X* act together, thereby constituting social institution *Y*. This means that the proportion of members of *X* who are acting in accordance with the relevant perceived deontic powers is large enough to matter to the members of *X*.

This social institution is *alienated* when:

(2) *Y* becomes causally independent of the decision of *X*. This means that *Y* exists in the sense outlined in (1) only by virtue of being partially constituted by the activity of members of *X* who so act only because of a mistaken belief concerning the independence of the social institution *vis-à-vis* either (i) themself as an individual, (ii) a cooperative sub-group of which they take themself to be a part, (iii) *X* as a whole, or (iv) are mistaken about the available opportunity for cooperative activity with others.

When this is the case:

(1) As a result of (2) being the case *Y* restricts the freedom of those members of *X* who (i) do not want to act in accordance with the relevant perceived deontic powers and (ii) have desires such that if *Y* exists in the sense outlined in (1) it would be costly not to so act.

This is a restriction of these agents' freedom because *Y* is said to exist only by virtue of being partially constituted by the activity of members of *X* whose activity is motivated by mistaken beliefs, which activity is intuitively thought to be *alterable*—

a necessary condition for an obstacle's ability to restrict freedom—on the assumption that agents are able to overcome mistaken beliefs.

While this account of course loses the simplicity and the common-sense feel of Marx's original discussion, I hope that it makes up for at least some of this by being more explicit and clear. There are, however, several aspects of my account that I have not been able to explore adequately, and I would like to here sketch a few of these as areas for further research.

First, I have put certain kinds of false beliefs at the heart of my account of the alienation of social institutions. But, we might ask, how and why might agents develop such false beliefs? And of the different kinds of false beliefs I have sketched, are there some that are more widespread, or more amenable to change, than others? How might we go about amending such false beliefs? Further, it seems that certain social institutions themselves are causally involved in the production of such beliefs: certainly, the social institutions around the dissemination of information and data seem likely culprits. How are we to theorise these connections?

Second, and relatedly, it seems that more than mere beliefs are involved in some of the descriptions of the motivations of members of *X*. In particular, in the kind of mistaken belief mentioned in (iv) above, it seems that a particular way of viewing the social world—of viewing oneself and others as isolated monads rather than as potential cooperators—is involved in agents' acting so as to constitute an alienated social institution. How are we to understand the role of such ways of viewing the world and the extent to which these are alterable? Here, we might see the early work of the members of the Frankfurt School, and in particular Horkheimer and Adorno's *Dialectic of Enlightenment*, as opening an interesting way forward: in this work, it seems that something like the concept of alienation I have described is at work, but is applied not primarily to social institutions but to modes of thought. For these writers, certain ways of thinking about the world, such as those characteristic of the enlightenment, were first developed in order to deal with particular problems that agents face,[1] but are said to have become fixed,[2] ossified[3] or objectified,[4] akin to the becoming-independent of social institutions that I analysed in chapter 4. Might we

---

[1] Adorno and Horkheimer, *Dialectic of Enlightenment*, 83.
[2] Adorno and Horkheimer, 37, 74, 78, 105.
[3] Adorno and Horkheimer, 41.
[4] Adorno and Horkheimer, 28.

articulate this insight using the tools developed in this thesis, and how else might this process interact with that I have described?

Third, as mentioned in chapter 4, my analysis of agents' behaviour has been carried out using a relatively thin conception of human agency, one which treats agents' desires and preferences as fixed, focusing primarily on the effect of mistaken beliefs for their behaviour. But agents' desires are not as stable or fixed as this account implies. And, indeed, agents' desires are frequently affected by other agents, and by the social institutions in which they live. How might we incorporate this into our account? In particular, how does this affect our discussion of freedom? A recent study by Rahel Jaeggi[5] tackles these questions with great perspicuity, however—as I have argued elsewhere,[6] and as she acknowledges towards the end of the book[7]—her account largely ignores the important differences between what I have called the modes of existence of different social institutions. A study that attempted to combine Jaeggi's discussion of these questions with an account of the alienation of social institutions might enable us to develop a richer understanding of both sets of problems.

---

[5] Jaeggi, *Alienation*.
[6] Ferguson, 'Review of Rahel Jaeggi, "Alienation"'.
[7] Jaeggi, *Alienation*, 220.

Bibliography

'Äußerung - English translation in English - Langenscheidt dictionary German-English'. Accessed 5 August 2019. https://en.langenscheidt.com/german-english/aeusserung.

Adorno, Theodor, and Max Horkheimer. *Dialectic of Enlightenment*. London: Blackwell Verso, 1997.

Althusser, Louis. *For Marx*. London: Verso, 2005.

Alvarez, Maria. 'Reasons for Action: Justification, Motivation, Explanation'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Winter 2017. Metaphysics Research Lab, Stanford University, 2017. https://plato.stanford.edu/archives/win2017/entries/reasons-just-vs-expl/.

Arneson, Richard J. '"Freedom and Desire"'. *Canadian Journal of Philosophy* 15, no. 3 (January 1985): 425–48.

Arthur, Chris. *Dialectics of Labour: Marx and His Relation to Hegel*. Oxford: Blackwell, 1986.

Bacharach, Michael. *Beyond Individual Choice: Teams and Frames in Game Theory*. Edited by Natalie Gold and Robert Sugden. Princeton Univ. Press, 2006.

Baker, Katie. 'UC Berkeley Was Warned About Its Star Professor Years Before Sexual Harassment Lawsuit'. *BuzzFeed News*. Accessed 2 August 2019. https://www.buzzfeednews.com/article/katiejmbaker/john-searle-complaints-uc-berkeley.

Benn, Stanley and William Weinstein. 'Being Free to Act, and Being a Free Man'. *Mind* LXXX, no. 318 (April 1971): 194–211.

Berlin, Isaiah. *Liberty*. Edited by Henry Hardy. Oxford Univ. Press, 2002.

———. 'Two Concepts of Liberty'. In *Liberty*, edited by Henry Hardy, 166–217. Oxford Univ. Press, 2002.

Brennan, Geoffrey, Lina Eriksson, Robert E. Goodin, and Nicholas Southwood. *Explaining Norms*. Oxford University Press, 2013.

Brudney, Daniel. *Marx's Attempt to Leave Philosophy*. Cambridge, Mass: Harvard University Press, 1998.

Carter, Ian. 'The Concept of Freedom: Part I: Background, Methodology, and the "Negative-Positive" Debate'. *Humane Studies Review* 10, no. 3 (1996).

———. 'Positive and Negative Liberty'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Summer 2018. Metaphysics Research Lab, Stanford University, 2018. https://plato.stanford.edu/archives/sum2018/entries/liberty-positive-negative/.

Chitty, Andrew. 'Hegel and Marx'. In *A Companion to Hegel*, edited by Stephen Houlgate and Michael Baur, 477–500. Malden, Mass.: Wiley-Blackwell, 2011.

———. 'Recognition and Property in Hegel and the Early Marx'. *Ethical Theory and Moral Practice* 16, no. 4 (August 2013): 685–97.

Christman, John. 'Liberalism and Individual Positive Freedom'. *Ethics* 101, no. 2 (January 1991): 343–59.

Cohen, G. A. *Self-Ownership, Freedom, and Equality*. Cambridge: Cambridge University Press, 1995.

Connolly, William E. *The Terms of Political Discourse*. Oxford: M. Robertson, 1983.

Day, J. P. 'On Liberty and the Real Will'. *Philosophy* 45, no. 173 (1970): 177–92.

———. 'Threats, Offers, Law, Opinion and Liberty'. *American Philosophical Quarterly* 14, no. 4 (1977): 257–72.

Dworkin, Gerald. 'The Nature of Autonomy'. *Nordic Journal of Studies in Educational Policy* 2015, no. 2

Elster, Jon. 'Sour Grapes - Utilitarianism and the Genesis of Wants'. In *Utilitarianism and Beyond*, edited by Amartya Sen and Bernard Williams. Cambridge: Cambridge University Press, 1982.

———. *Making Sense of Marx*. Cambridge: Cambridge University Press, 1985.

———. 'Rationality and Social Norms'. *Arch. Europ. Sociol.* 32, no. 1 (1991): 109–29.

Engels, Friedrich. 'Ludwig Feuerbach and the End of Classical German Philosophy'. In *Collected Works 26*, by Karl Marx and Friedrich Engels. London: Lawrence & Wishart, 1975.

'English Translation of "Äußerung" | Collins German-English Dictionary'. Accessed 5 August 2019. https://www.collinsdictionary.com/dictionary/german-english/ausserung.

Epstein, Brian. *The Ant Trap: Rebuilding the Foundations of the Social Sciences*. New York: Oxford University Press, 2015.

———. 'Précis of The Ant Trap'. *Journal of Social Ontology* 2, no. 1 (January 2016)

Faigenbaum, Gustavo, and John Searle. *Conversations with John Searle*. LibrosenRed, 2003.

Ferguson, Chris. 'Review of Rahel Jaeggi, "Alienation"'. Accessed 13 September 2019. https://marxandphilosophy.org.uk/reviews/7971_alienation-review-by-chris-ferguson/.

Feuerbach, Ludwig. '"The Essence Of Christianity" in Relation to "The Ego And Its Own"'. Translated by Frederick Gordon. *The Philosophical Forum* viii, no. 2-4 (1978): 81–91.

———. *The Essence of Christianity*. Mineola, N.Y.: Dover Publications, 2008.

———. 'Towards a Critique of Hegel's Philosophy'. In *The Fiery Brook: Selected Writings*, edited by Zawar Hanfi. London: Verso, 2012.

———. 'Preliminary Theses on the Reform of Philosophy'. In *The Fiery Brook: Selected Writings*, edited by Zawar Hanfi. London: Verso, 2012.

———. 'Principles of the Philosophy of the Future'. In *The Fiery Brook: Selected Writings*, edited by Zawar Hanfi. London: Verso, 2012.

Fichte, Johann Gottlieb. *Attempt at a Critique of All Revelation*. Edited by Allen W. Wood. Translated by Garrett Green. Cambridge: Cambridge University Press, 2010.

Frankfurt, Harry G. 'Freedom of the Will and the Concept of a Person'. *The Journal of Philosophy* 68, no. 1 (January 1971).

Geras, Norman. *Marx and Human Nature: Refutation of a Legend*. London: Verso, 1983.

Gilbert, Margaret. *A Theory of Political Obligation: Membership, Commitment, and the Bonds of Society*. Oxford: Oxford University Press, 2006.

———. 'Searle and Collective Intentions'. In *Intentional Acts and Institutional Facts: Essays on John Searle's Social Ontology*, edited by Savas L Tsohatzidis, 31–48. Dordrecht: Springer, 2007.

Gray, John . 'On Negative and Positive Liberty'. *Political Studies* 28, no. 4 (December 1980): 507–26.

———. 'Against Cohen On Proletarian Unfreedom'. *Social Philosophy and Policy* 6, no. 1 (September 1988).

Hacking, Ian. 'Searle, Rationality and the Social'. *History of the Human Sciences* 10, no. 4 (November 1997): 83–92.

Hanfi, Zawar. 'Introduction'. In *The Fiery Brook: Selected Writings*, by Ludwig Feuerbach, edited by Zawar Hanfi. London: Verso, 2012.

Hart, H.L.A. 'Are There Any Natural Rights?' *The Philosophical Review* 64, no. 2 (April 1955).

———. *The Concept of Law*. Oxford: Oxford University Press, 2012.

Haworth, Alan. 'What's so Special about Coercion?' *Economy and Society* 19, no. 3 (August 1990): 376–89.

Hayek, Friedrich A. von. *The Constitution of Liberty: The Definitive Edition*. Edited by Ronald Hamowy. The Collected Works of F. A. Hayek, v. 17. Chicago: University of Chicago Press, 2011.

Hegel, Georg Wilhelm Friedrich. *Phenomenology of Spirit*. Translated by A. V. Miller. Oxford: Oxford Univ. Press, 2013.

Hess, Moses. *The Holy History of Mankind and Other Writings*. Edited by Shlomo Avineri. Cambridge: Cambridge University Press, 2004.

Hindriks, Frank. 'The New Role of the Constitutive Rule'. *The American Journal of Economics and Sociology* 62, no. 1 (2003): 185–208.

———. 'The Freedom of Collective Agents'. *Journal of Political Philosophy* 16, no. 2 (2008): 165–83. https://doi.org/10.1111/j.1467-9760.2007.00292.x.

———. 'Collective Acceptance and the Is-Ought Argument'. *Ethical Theory and Moral Practice* 16, no. 3 (June 2013): 465–80.

Hirschmann, Nancy J. *The Subject of Liberty: Toward a Feminist Theory of Freedom*. Princeton, NJ: Princeton University Press, 2003.

Hobbes, Thomas. *Leviathan*. Edited by J. C. A. Gaskin. Oxford: Oxford University Press, 1998.

Hohfeld, Wesley Newcomb. *Fundamental Legal Conceptions as Applied in Judicial Reasoning and Other Legal Essays*. Aldershot: Ashgate/Dartmouth, 2001.

Hook, Sidney. 'Marx and Feuerbach'. *New International* 3, no. 2 (April 1936).

Hume, David. *A Treatise of Human Nature: A Critical Edition*. Edited by David Fate Norton and Mary J. Norton. Oxford: Oxford University Press, 2007.

Hurka, Thomas. *Perfectionism*. New York: Oxford University Press, 1993.

'Inseparable Prefixes (Feste Vorsilben)'. Accessed 5 August 2019. http://www.dartmouth.edu/~deutsch/Grammatik/Wortbildung/Inseparables.html.

'Interview with Michael Gove, 25/09/2019'. *Today*. BBC Radio 4, 25 September 2019. https://www.bbc.co.uk/sounds/play/m0008nwz.

Jaeggi, Rahel. *Alienation*. Translated by Frederick Neuhouser and Alan E Smith. New York: Columbia Univ. Press, 2016.

———. 'What (If Anything) Is Wrong with Capitalism? Dysfunctionality, Exploitation and Alienation: Three Approaches to the Critique of Capitalism'. *The Southern Journal of Philosophy* 54 (September 2016): 44–65.

Johnston, Larry. *Between Transcendence and Nihilism: Species-Ontology in the Philosophy of Ludwig Feuerbach*. New York: P. Lang, 1995.

Kamenka, Eugene. *The Philosophy of Ludwig Feuerbach*. London: Routledge & K. Paul, 1970.

Kant, Immanuel. *Religion within the Boundaries of Mere Reason and Other Writings*. Translated by Allen W. Wood and George Di Giovanni. Cambridge: Cambridge University Press, 1998.

Kristjánsson, Kristján. *Social Freedom: The Responsibility View*. Cambridge: Cambridge University Press, 1996.

Laitinen, Arto. 'Recognition, Acknowledgement, and Acceptance'. In *Recognition and Social Ontology*, edited by Heikki Ikäheimo and Arto Laitinen, 309–47. Leiden: Brill, 2011.

Lawrence, Rachel V Kohout. 'Foreword'. In *The Essence of Christianity*, by Ludwig Andreas Feuerbach. New York: Prometheus Books, 1989.

Leezenberg, Michiel. 'The Contract of Language: John Searle's Philosophy of Society'. *International Review of Pragmatics* 3, no. 1 (2011): 98–112.

Leopold, David. *The Young Karl Marx: German Philosophy, Modern Politics, and Human Flourishing*. Cambridge: Cambridge University Press, 2007.

Lewis, David. *Convention: A Philosophical Study*. Oxford: Blackwell, 2011.

Locke, John. *The Second Treatise of Government: And, A Letter Concerning Toleration*. Mineola, N.Y: Dover Publications, 2002.

Luckmann, Thomas, and Peter Berger. *The Social Construction of Reality. A Treatise in the Sociology of Knowledge*. London: Allen Lane, 1971.

Lukes, Stephen. 'Emancipation'. In *A Dictionary of Marxist Thought*, edited by Thomas B. Bottomore. Oxford: Blackwell Reference, 1983.

MacCallum, Gerald. 'Negative and Positive Freedom'. *The Philosophical Review* 76, no. 3 (1967): 312–34.

Margolis, Joseph. 'Praxis and Meaning: Marx's Species Being and Aristotle's Political Animal'. In *Marx and Aristotle: Nineteenth-Century German Social Theory and Classical Antiquity*, edited by George E. McCarthy. Savage, Md: Rowman & Littlefield Publishers, 1992.

Marx, Karl. *Grundrisse: Foundations of the Critique of Political Economy (Rough Draft)*. Translated by Martin Nicolaus. Harmondsworth: Penguin, 1973

———. 'Critique of Hegel's Doctrine of the State'. In *Karl Marx: Early Writings*, 57–198. Harmondsworth: Penguin, 1975.

———. 'On the Jewish Question'. In *Karl Marx: Early Writings*. Harmondsworth: Penguin, 1975.

———. 'A Contribution to the Critique of Hegel's Philosophy of Right. Introduction'. In *Karl Marx: Early Writings*. Harmondsworth: Penguin, 1975.

———. 'Excerpts from James Mill's *Elements of Political Economy*'. In *Karl Marx: Early Writings*, 259–78. Harmondsworth: Penguin, 1975.

———. 'Economic and Philosophical Manuscripts'. In *Karl Marx: Early Writings*, 279–400. Harmondsworth: Penguin, 1975.

———. 'Critical Notes on the Article "The King of Prussia and Social Reform. By a Prussian."' In *Karl Marx: Early Writings*. Harmondsworth: Penguin, 1975.

———. 'Concerning Feuerbach'. In *Karl Marx: Early Writings*. Harmondsworth: Penguin, 1975.

———. *Capital: Volume III*. Translated by David Fernbach. Penguin Classics, 1993.

———. 'Letter to Arnold Ruge, May 1843'. In *Collected Works 3*. London: Lawrence & Wishart, 2010.

Marx, Karl, and Friedrich Engels. 'The German Ideology'. In *Collected Works 5*. London: Lawrence & Wishart, 1976.

———. 'The Holy Family'. In *Collected Works 4,* 5–211. London: Lawrence & Wishart , 2010.

Mészáros, István. *Marx's Theory of Alienation*. Delhi: Aakar Books, 2006.

Mill, John Stuart. *On Liberty*. Edited by Edward Alexander. Peterborough, Ontario: Broadview, 1999.

Miller, David. 'Constraints on Freedom'. *Ethics* 94, no. 1 (1983): 66–86.

———. *Market, State, and Community: Theoretical Foundations of Market Socialism*. Oxford: Clarendon Press, 1989.

Moggach, Douglas. *The Philosophy and Politics of Bruno Bauer*. Cambridge: Cambridge University Press, 2003.

Nozick, Robert. 'Coercion'. In *Socratic Puzzles*, 15–44. Cambridge, Mass: Harvard University Press, 1997.

———. *Anarchy, State, and Utopia*. Malden, MA: Blackwell, 2012.

Ollman, Bertell. *Alienation: Marx's Conception of Man in Capitalist Society*. Cambridge: Cambridge Univ. Press, 1998.

Oppenheim, Felix. '"Constraints on Freedom" as a Descriptive Concept'. *Ethics* 95, no. 2 (January 1985): 305–9.

———. 'Social Freedom: Definition, Measurability, Valuation'. *Social Choice and Welfare* 22, no. 1 (February 2004): 175–85.

Parijs, Philippe van. *Real Freedom for All: What (If Anything) Can Justify Capitalism?* Oxford: Oxford University Press, 1995.

Pinkard, Terry P. 'Translator's Note'. In *The Phenomenology of Spirit*, by Georg Wilhelm Friedrich Hegel, translated by Terry P. Pinkard. New York: Cambridge University Press, 2017.

Rosen, Zvi. 'The Influence of Bruno Bauer on Marx's Concept of Alienation'. *Social Theory and Practice* 1, no. 2 (1970): 50–68.

———. *Bruno Bauer and Karl Marx: The Influence of Bruno Bauer on Marx's Thought*. The Hague: Nijhoff, 1977.

Rousseau, Jean-Jacques. *The Social Contract*. Translated by Christopher Betts. Oxford: Oxford University Press, 1999.

Schacht, Richard. *Alienation*. London: Allen & Unwin, 1971.

Schmitt, Richard. 'Marx's Concept of Alienation'. *Topoi* 15, no. 2 (September 1996): 163–76.

Schutz, Alfred. *The Phenomenology of the Social World*. Evanston, Ill: Northwestern University Press, 1972.

Searle, John R. 'How to Derive "Ought" from "Is"'. *Philosophical Review* 73, no. 1 (1964): 43–58.

———. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge Univ. Press, 1969.

———. *Intentionality, an Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press, 1983.

———. 'Collective Intentions and Actions'. In *Intentions in Communication*, edited by Philip R. Cohen, Jerry L. Morgan, and Martha E. Pollack, 401–15. Cambridge, Mass: MIT Press, 1990.

———. *The Construction of Social Reality*. London: Lane, Penguin, 1995.

———. *Rationality in Action*. Cambridge, Mass: MIT Press, 2001.

———. 'Language and Social Ontology'. *Theory and Society* 37, no. 5 (October 2008): 443–59.

———. *Making the Social World: The Structure of Human Civilization*. Oxford: Oxford University Press, 2010.

Sensat, Julius. *The Logic of Estrangement: Reason in an Unreasonable Form*, New York: Palgrave Macmillan, 2016.

Simmons, A. John. *Moral Principles and Political Obligations*. Princeton, N.J: Princeton University Press, 1979.

Stahl, Titus. 'Institutional Power, Collective Acceptance, and Recognition'. In *Recognition and Social Ontology*, edited by Heikki Ikäheimo and Arto Laitinen. Leiden: Brill, 2011.

Steiner, Hillel. 'Individual Liberty'. *Proceedings of the Aristotelian Society* 75 (1974): 33–50.

Stirner, Max. *The Ego and Its Own*. Edited by David Leopold. Cambridge: Cambridge University Press, 1995.

Taylor, Charles. 'What's Wrong with Negative Liberty'. In *Philosophical Papers 2*. Cambridge: Cambridge University Press, 1985.

Thomason, Burke C. *Making Sense of Reification*. London: Macmillan, 1982.

Ullmann-Margalit, Edna. *The Emergence of Norms*. Oxford: Clarendon Press, 1977.

Walicki, Andrzej. 'Marx and Freedom'. *The New York Review of Books* 30, no. 18 (November 1983). https://www.nybooks.com/issues/1983/11/24/.

———. *Marxism and the Leap to the Kingdom of Freedom: The Rise and Fall of the Communist Utopia*. Stanford, Calif: Stanford Univ. Press, 1996.

Wartofsky, Marx W. *Feuerbach*. Cambridge: Cambridge University Press, 1982.

Weinberg, Justin. 'Searle Found to Have Violated Sexual Harassment Policies'. *Daily Nous*. Accessed 1 August 2019. http://dailynous.com/2019/06/21/searle-found-violated-sexual-harassment-policies/.

'Wesen | Translate German to English: Cambridge Dictionary'. Accessed 23 August 2019. https://dictionary.cambridge.org/dictionary/german-english/wesen.

Wood, Allen W. *Karl Marx*. London: Routledge & Kegan Paul, 1981.