Branding and liberal autonomy
Terlaak, E.

Citation for published version (APA):
Are you manipulated by branding campaigns? This book addresses this question by drawing on work in both philosophy and psychology. It does so from the perspective of the liberal state. As such, manipulation is understood as the violation of liberal autonomy. The central claims are that, on this definition, branding does manipulate consumers, that it does so structurally, and that this gives the liberal state the obligation to take measures against it. Some parts of this argument are certainly more robust than others. But as techniques of persuasion are rapidly becoming more sophisticated, this book should at the very least give you good reasons to be concerned about the influence of branding.
Branding and Liberal Autonomy

ACADEMISCH PROEFSCHRIFT
ter verkrijging van de graad van doctor aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus prof. dr. D.C. van den Boom
ten overstaan van een door het college voor promoties ingestelde commissie, in het openbaar te verdedigen in de Agnietenkapel
op woensdag 18 Mei 2011, te 14:00 uur
door Edi Terlaak
tegen te Hemelum
Contents

Acknowledgements ......................................................................................................................... 5

Introduction ........................................................................................................................................ 7

1 Scope and Structure of the Argument ............................................................................................ 9
   1 Why the Liberal State Must Promote and Protect Autonomy .................................................. 9
   2 What an Autonomy-based Assessment of Branding Should Look Like .................................. 14

2 Minimal Liberal Autonomy ........................................................................................................ 20
   1 Intuitions about Autonomy ......................................................................................................... 21
   2 Self-control ................................................................................................................................... 23
      2.1 Self-controlled Action ............................................................................................................... 24
      2.2 Self-controlled Mental States .................................................................................................. 31
      2.3 The Shortcut from Self-control to Autonomy ........................................................................... 38
   3 Authenticity ...................................................................................................................................... 40
      3.1 Authentic Action ....................................................................................................................... 41
      3.2 Authentic Mental States ............................................................................................................ 47

4 Practical Thresholds of Liberal Autonomy .................................................................................. 50
   4.1 Thresholds of Self-Control ........................................................................................................ 51
   4.2 Thresholds of Authenticity ......................................................................................................... 55

3 The Structure of Branding .......................................................................................................... 59
   1 When Branding Leads to Sales .................................................................................................... 63
      1.1 Brand Equity and Consumer Attitudes .................................................................................... 63
      1.2 Advertising Effectiveness Research ......................................................................................... 64
   2 Human Motivations ....................................................................................................................... 65
      2.1 Theoretical Arguments for Evolved Higher-order Human Motivations ................................. 66
      2.2 Evidence for Higher-Order Human Motivations ................................................................ 67
   3 Motivations and Consumption ..................................................................................................... 74
      3.1 Consuming Community ............................................................................................................. 75
      3.2 Good Mate Value for Money ................................................................................................... 76
      3.3 Selling Power to the Powerless ............................................................................................... 78
   4 Social Motivations, Consumption and Branding ......................................................................... 79
      4.1 Affiliation, Advertising and Existential Threats .................................................................... 80
      4.2 Mating, Advertising and the Sex Ratio in Society ................................................................. 80
      4.3 Status Rank, Advertising and Foreign Occupations ................................................................. 81
   5 Managerial Implications of the Motivational Structure of Branding ........................................ 82
      5.1 Widening the Brand Community ............................................................................................. 82
      5.2 Prudent Displays of Fitness ..................................................................................................... 84
      5.3 Status Appeals That Fit ............................................................................................................ 85
How Branding Structurally Violates Minimal Liberal Autonomy

1 Branding, Autonomy and Consumer Action
   1.1 Unconscious Point of Purchase Influence
   1.2 Unconscious Point of Consumption Influence

2 How Branding Changes Mental States
   2.1 The Mechanics of Autonomy Violating Attitude Change in Branding
   2.2 How Autonomy Violating Brand Attitudes Cause Sales
   2.3 Historical Evidence of Autonomy Violation in Brand Attitude Change

3 The Centrality of Autonomy Violation in Branding
   3.1 The Centrality of Autonomous Action Violation in Branding
   3.2 The Centrality of Autonomy Violation Brand-related Mental State Formation

4 The Retrievability of Brand Managerial Intentions
   4.1 Managerial Knowledge about the Autonomy Violation Involved in Using Persuasion Techniques in Branding
   4.2 Managerial Intentions to Influence Behavior by Means of Autonomy Violating Techniques

5 Consumer Consent to Covert Brand Influence
   5.1 Tacit Consent to Autonomy Violation in Branding
   5.2 Explicit Consent to Autonomy Violation in Branding

Why Branding is Special in Structurally Violating Liberal Autonomy

1 Liberal Autonomy and Religion
   1.1 Structural Autonomy Violation in Religion
   1.2 Practical Thresholds of Autonomy Violation in Religion

2 Liberal Autonomy and Political Promotion
   2.1 Structural Autonomy Violation in Political Promotion
   2.2 Practical Thresholds of Liberal Autonomy in Political Promotion

Conclusion

English Summary

Dutch Summary

Notes

References
I want to thank first of all my advisor Veit Bader for his prompt and incisive commentary as well as his encouragements. His work has moreover been a model for me throughout the process of writing this thesis. Admittedly, the result ended up looking quite a bit different from the model, but I like to think that this is due mainly to the nature of my subject matter. I want to thank Geoffrey Levey for the opportunity to present parts of this thesis at the workshop ‘Authenticity, Autonomy and Culture’, and Joel Anderson, Alison Dundes Renteln and Beate Rässler for their thoughtful comments on that occasion. Jan Hein Furnée, Jaap den Hollander and Homme Wedman commented constructively on my work when it was in its early stages, while John Christman, Tabe Bergman, Agnes Nairn and Jan Bransen did so when it was almost finished. I also want to thank Georg Toepfer and Catrin Misselhorn for granting me access to their seminars at Humboldt University on, respectively, culture and evolution and political liberalism. I benefited in many ways from discussions with them and other participants of the seminar. Furthermore, Eva Walther and Jaap van Ginneken generously sent me drafts of their work, while Dorothee Horstköter kindly sent me her PhD thesis. I am grateful also to Hans Wubs for designing the wonderful cover and for doing the lay out. I would not have been able to write this thesis without the fellowship granted to me by ASCA. I thank its staff, and Eloë Kingma in particular, for the trust they placed in me. Privately, I received special support from my roommate in Berlin, Florian Noto, my good friend Wybo Wiersma and my parents.
In all contemporary industrialized countries a large cluster of organizations is dedicated to influencing how people feel, think and ultimately act in relation to branded products. But most industrialized countries are also liberal states and as such – on a common understanding – committed to protecting autonomy. If there is indeed such a commitment, then it seems relevant to ask whether it implies that the liberal state has the obligation to protect its citizens from branding. In this thesis I contribute to answering this question in two ways. I both outline the structure that such an argument must have for it to be convincing, and – on the basis of this design – make a case for thinking that branding undermines liberal autonomy.

Outlining the structure of my argument – my first contribution – is the task of chapter 1. In this chapter, I first lay out the case for thinking that liberal legitimacy implies a commitment to autonomy and then explore what an assessment of branding with regard to this standard should look like. The guiding idea of my design is to isolate a core of minimal thresholds of liberal autonomy and to render these empirical with the help of psychology. If the psychological processes by which branding influences consumers are known, then it can be established with regard to a branded message whether it violates any one threshold of autonomy. My second contribution consists in building an argument on the basis of this design. The first stage of this argument, set out in chapter 2, is to spell out – in sufficient empirical detail – what the minimal core among theories of autonomy consists in. Only in some instances, such as in my discussion of authentic mental state formation, do I have to add minimal principles of my own that follow from a commitment to central intuitions about autonomy. Admittedly, my account is more contentious in these places, but since I present a differentiated set of minimal standards, my assessment need not be contingent on the plausibility of these principles (and this will indeed turn out not to be the case).

Since I will argue that autonomy violation is relevant to the liberal state only if it is sufficiently structural, I need a theory about structural features of branding. Unfortunately, such a theory does not yet exist. What I will therefore do in chapter 3 is to apply to branding the epidemiological approach to culture. Obviously, I cannot hope to develop a comprehensive theory of branding of my own in the current context. Hence, I focus on a limited number of key features of branding that can be readily identified on such an approach, and that are also a priori relevant to the normative assessment of branding.

Chapter 4 then turns to consumer research in order to establish whether the use of techniques for which the previous chapter postulated structural pressures are effective in enhancing sales. Although there is no research that speaks to this issue directly, a suggestive case is made for thinking that autonomy violating techniques are effective under the conditions specified in chapter 3. In chapter 5 I go on to ask whether my conclusion with regard to autonomy violation in branding ought to be extended to traditional religions and political
communication. If that were the case, then the framework for my analysis would contradict very strong intuitions. Because I am committed to political justification on the basis of some form of reflective equilibrium, this would count as an argument against the framework. However, I claim that traditional religions and political communication are different from branding in that they do not violate all of the relevant minimal conditions of liberal autonomy. Thus, my argument in chapter 5 is that branding is special in structurally violating liberal autonomy and that this gives the liberal state a pro tanto reason to protect citizens against branding.

I do not claim to offer a knockdown argument for state intervention against branding. That would be presumptuous given that my argument covers a lot of ground, much of which is either hotly contested or largely unexplored in philosophy and consumer research. My claim is therefore rather that I render plausible the possibility that autonomy is structurally violated in branding to such an extent that it should urgently become the object of more intense philosophical and psychological scrutiny.
1 Scope and Structure of the Argument

In this chapter I prepare the ground for the argument that will be developed in this thesis. In a first step I argue for the political relevance of autonomy as my standard of normative evaluation. I do so by presenting a number of arguments that have recently been developed by Ben Colburn for thinking that the liberal state has the obligation to protect and promote autonomy. As good reasons for this position will count that it follows from commitments to widely shared principles in liberal political philosophy. As such, I will present an argument to the effect that those who are committed to anti-perfectionism (but are not anarchists) must accept that the liberal state is to promote and protect autonomy. Moreover, perfectionist theories of the liberal state that claim to be able to do without autonomy will be shown to either implicitly rely on a commitment to autonomy or to be seriously unattractive. If this argument is correct, and if branding were shown to undermine autonomy, then the liberal state would have a pro tanto reason for acting against it.

Whether branding in fact undermines autonomy has been the subject of a protracted debate in business ethics. I will, and this is the second step, briefly run through the arguments for and against this position in order to learn not so much from the arguments themselves as from the strategies that have been pursued. On the basis of an evaluation of the appropriateness of these respective approaches I make a case for a more comprehensive design that will structure my argument in the remainder of this thesis.

1 Why the Liberal State Must Promote and Protect Autonomy

If we want to know whether autonomy violation in branding calls for action by the liberal state, then we must first know whether autonomy violation in general is a reason for state action. For autonomy violation to be a reason for state action it must be true that the state has an obligation to protect autonomy. There are two ways to make such a case. First, it can be argued that it is intrinsically valuable to promote and protect autonomy, and that the liberal state should act on the basis of this value. However, this approach is suboptimal because intuitions about the value of autonomy are ambivalent. A more convincing type of argument is one that would show that the obligation to promote and protect autonomy arises from a commitment to uncontroversial liberal values. In this section I will present an argument to this effect that has been developed by Ben Colburn. Although I will argue that most
steps in his argument are valid, I also claim that it is incomplete and suggest how it can be completed.

**Why Autonomy is Compatible with Anti-Perfectionism**

A common objection against the demand for the promotion of autonomy by the liberal state is that this demand would be perfectionist (i.e. that it would impose upon citizens a version of the good life). Because anti-perfectionism is one of the central tenets of liberal political philosophy this is a serious objection. However, as Colburn demonstrates, a commitment to both anti-perfectionism and autonomy is not only coherent, it is also the case that anti-perfectionism follows from a commitment to autonomy. In order to demonstrate why this is so, Colburn makes a distinction between first-order values and second-order values.

First-order values can be made sense of by discovering states of affairs in the world. Such values can be either content specific or content neutral. Content specific first-order values tell us what is valuable for someone in a straightforward way, in the sense that there is no need to relate the prescription to properties of the individual whose life is assessed by the value. To say that it is valuable to be able to play the piano sonatas by Schubert is an example of such a value. If you know what it means to play the piano and are familiar with the piano sonatas by Schubert, then you know what is meant by this prescription and you can say whether your life is valuable by this standard. Content neutral first-order values, on the other hand, tell us what is valuable but require us to apply it to an individual's life. If I say that it is valuable to satisfy desires, then even if you know what is meant by desires and by satisfaction, you still do not know whether your life is valuable. To this end, you must find out what your desires are (or, if you want to know whether the life of another person is valuable, you must know what her desires are). The ultimate test for establishing whether you are dealing with a content neutral value or a content specific value, Colburn maintains, is to fill in the details that are required to be able to say whether the life of an individual is valuable. If by doing so the truth value or meaning of the value changes, then you are dealing with a content neutral value. If nothing of the sort occurs, then you are dealing with a content specific value. Thus, if my desire is to play computer games, and the value under investigation is the satisfaction of desires, then I cannot change the latter prescription into ‘playing computer games is valuable’. If not everybody desires to play computer games, then the truth value changes. And even if everybody desired to play computer games, then the meaning of the prescription would change. For what is valuable here about me playing computer games is not contingent on some intrinsic property of computer games, but derivative of the satisfaction of desires that I bring about by playing computer games.

Second-order values are a special subset of content neutral values. They are different from first-order values in that in order to know what is meant by them it is of no use to learn anything about states of affairs in the world. Instead, one is referred to other values. The example Colburn gives is the difference between the content neutral prescriptions ‘What is valuable is to satisfy your desires’ and ‘What is valuable is to do what your parents value’. In the first case, if I want to know what is meant by the prescription I must know about states of affairs (i.e. my desires), whereas in the second case I must know about values (i.e. the values that my parents entertain). The value of my parents can be both content specific (e.g. what is valuable is to play Schubert’s piano sonatas) or content neutral (e.g. what is valuable is to satisfy desires). If the latter is the case, then the second-order value (i.e. ‘do what your parents value’) contains within it a content neutral value (i.e. ‘satisfy your desires’). Thus, according to Colburn, it is characteristic of second-order values that they can have content neutral values inside of them.
It can be contested whether the nesting of content-neutral values is really what is driving the argument though, for nothing seems to prevent a desire from referring to a content-neutral value (such as a desire to act ethically). If that is true, then the first-order value that desires must be satisfied can refer to another content-neutral value (e.g. ‘it is valuable to act on the basis of ethical rules’) and as such has the relevant nested structure as well. What makes second-order values different may be rather that they prescribe not what to do (e.g. play Schubert’s piano sonatas) or where to look for prescriptions about what to do (e.g. one’s desires), but that they prescribe by which type of processing the prescription should be generated (i.e. the valuing process).

If either one of these versions of the distinction between first and second-order values is correct, then it seems reasonable to say – with Colburn – that anti-perfectionism means only that no first-order values should be promoted. That is, anti-perfectionism prescribes that the state should not tell citizens which first-order values they are to pursue. Autonomy, by contrast, would be a second-order value. That is, if I say that it is valuable to be autonomous (i.e. to make decisions for yourself), then you are referred not to a state of affairs but to the specification of values by your valuing process. If so, then it is coherent to say that it is true both that the state should promote (the second-order value) autonomy, and that it should not promote any (first-order) values. What is more, if the liberal state were to promote autonomy, then this is a reason for it to be anti-perfectionist. After all, if one wants to stay true to a second-order value (such as autonomy) one cannot fill in the first-order value to which it refers, since that would change its meaning and possibly its truth value. Hence it is true that if autonomy is valuable, the state should not promote any first-order values.

Why Anti-Perfectionism Requires Autonomy

A much stronger claim than saying that autonomy justifies anti-perfectionism is to say that it is the only value that can do so. One can of course deny that anti-perfectionism is to be justified. However, intuitions on the value of anti-perfectionism are too ambivalent to allow for such an assertion. If it is true that anti-perfectionism must be justified, then the least controversial way of doing so is to construct an inductive argument for thinking that autonomy is the only serious candidate. Such an argument would have to show that alternative justifications fail, that they are unattractive or that they implicitly recur on autonomy. Colburn has made precisely such an argument with regard to the values pluralism, skepticism and neutrality.

With regard to pluralism, the only variant that can hope to justify anti-perfectionism is strong pluralism. It holds that there can be choices between mutually exclusive values that are both worth pursuing. Such a justification would be that one option should not be pursued if it diminishes another option because that other option is valuable, and that the state should hence be anti-perfectionist. However, in that case anti-perfectionism is justified by virtue of the valuing of a first-order value, and such valuing of first-order values is precisely what the anti-perfectionist state should not do. In other words, if anti-perfectionism were justified by strong pluralism, then we would have to say that the state should not value any first-order values because first-order values are valuable. And that is a contradiction.

The skeptical justification for anti-perfectionism is that uncertainty about whether a value is really valuable should hold the state back from promoting first-order values. However, because uncertainty in itself is not an intrinsic reason for or against engaging in an action, the skeptic must point to such a reason in the form of negative consequences that follow from the uncertainty about the value promotion. One possibility is to claim that if the state promotes a first-order value that it wrongly believes to be valuable some other value may
be impaired. But then anti-perfectionism would be justified on the basis of a first-order value and that was shown to be a contradiction. Another possibility for the skeptic is to claim that it is uncertain that anything is valuable at all. But that is an argument for indifference, and indifference is not an argument against promoting first-order values.

Finally, neutrality towards comprehensive doctrines (as advocated by the late John Rawls) can be said to justify anti-perfectionism. Neutrality with regard to comprehensive doctrines is required, on this account, because the state is only legitimate if it is accepted by the governed. If it is correct to suppose that differences between comprehensive doctrines are not going to disappear (because of the burdens of judgment), then if the state wants to be accepted it should be neutral with regard such comprehensive doctrines. And if the state is to be neutral towards comprehensive doctrines, then it could be argued that it ought not promote any first-order values. The problem with this argument, Colburn notes, is that it is incomplete. That is, it is not clear why the state is deficient in legitimacy if it is guided by comprehensive doctrines that some do not accept. In the case of Rawls for example it is not explicitly explained why actual consent is necessary for legitimacy, but the only relevant argument from within his theory, Colburn claims, would be to say that it is out of respect for the two moral powers. One is the sense of justice and the other the ability to form and revise a conception of the good. However, the latter is but a conception of autonomy. Hence, the concern for neutrality understood as respect for the consent of the governed recurs on the value of autonomy.

If Colburn is correct in supposing that these are the only serious alternative justifications for anti-perfectionism, and if he is also correct in claiming that they all fail as independent justifications, then this will count as inductive evidence for thinking that only autonomy can justify anti-perfectionism. If that is true, then all who think that the state is to be anti-perfectionist in its policies should accept that the state is to promote autonomy. One can of course think that anti-perfectionism is important but still hold that the state is not to promote autonomy if one is against any state action. In that case the state is to promote nothing at all. In this regard I point out merely that this anarchist position is seriously unattractive in a great number of respects that would take me too far afield here.

Why Liberal Perfectionism Without Autonomy is Unattractive

So far Colburn’s argument has shown that all liberal anti-perfectionists must accept that the state is to promote autonomy. If that claim is correct, my overall argument would already have a considerable scope. However, not all political theorists are anti-perfectionists (and among them not all support a brand of perfectionist autonomy-minded liberalism), so that the scope would be broadened further still if it were shown that no first-order perfectionist value can convincingly support the framework of liberal values. To this end, Colburn has identified diversity as the main candidate and given arguments for thinking that it is significantly less attractive than autonomy.

Philosophers who want to ground liberalism on the value of diversity, Colburn maintains, have not yet made an explicit argument for thinking that diversity is valuable and appear to assume that intuition would inform us that it is. It is not clear though that people do indeed universally share a concern for diversity for its own sake. An example of Colburn illustrates this point. He describes how the Aran Islanders lost their distinctive culture with the arrival of industrial society in the course of the 20th century. If we suppose that something of value was lost, then it is not clear what drives this intuition. Diversity liberals would say that we value diversity in itself and therefore are concerned about the loss of a culture. An autonomy-minded liberal would be more sympathetic to the interpretation that the loss
we feel is due to the realization that the culture was valuable to islanders themselves (and thus, from the perspective of autonomous agents). On the face of it, it seems impossible to say which theory best explains the intuition. However, Colburn argues that when we tweak the case somewhat and suppose that the Islanders themselves wanted to give up their culture in favor of modernity we can see that autonomy liberalism is substantially more attractive than diversity liberalism. For diversity liberals would have to say that the loss of a culture is at least a \textit{pro tanto} reason to ignore the wishes of the Islanders. Such a reason can of course be overturned by other reasons, but the point that Colburn makes is that if diversity liberalism is to be distinguished from autonomy liberalism is some respect, then it is in cases such as these.\textsuperscript{13} And here the \textit{pro tanto} reason to ignore people’s judgments must count against diversity as a basis for liberal political theory.\textsuperscript{14}

\textbf{Why Colburn’s Argument For Autonomy-Minded Liberalism is Incomplete}

Opponents of autonomy-minded liberalism will retort that placing autonomy at the center of the liberal value structure has unattractive consequences of its own. Some such claims are not convincing. The tyranny argument of Isaiah Berlin for example holds that supporters of autonomy (or “positive freedom”) adhere to a notion of “rational self-direction” which they with the best of intentions want to impose – if must be with force – on other individuals.\textsuperscript{15} That is, if we accept the idea of autonomy as rational self-direction, we would by implication hold that this is the “one true purpose” of man; that “the ends of all rational beings must … fit into a single universal, harmonious pattern”; and that all men must be “made rational” so that they “obey the rational laws of their own natures”. Thus, the result of a commitment to autonomy would be that “despotism … turns out to be identical with freedom”.\textsuperscript{16} In response to this argument John Christman distinguishes between internalist and externalist accounts of rationality in relation to autonomy. The internalist account focuses solely on the subjective side of autonomy, where rationality refers only to the coherence of beliefs and desires. The externalist or objective account of rationality refers instead to a relation of fit with the external world, which can apply to beliefs (weak externalism) or to values (strong externalism).\textsuperscript{17} On the basis of this distinction it follows that Berlin’s attack on the despotism of autonomy only hurts strong externalist versions. Because most of those who advocate autonomy-minded liberalism hold that the rationality required for autonomy is confined to internalist or weak externalist accounts, Berlin’s tyranny argument applies only to a small subset of accounts of autonomy-minded liberalism.

A more serious objection is that there is no agreement on what autonomy is about and that deciding which aspects of the valuing process matter (and to what extent they matter) would therefore involve either arbitrariness or first-order values. In other words, there must be a standard for assessing rival conceptions of autonomy and that standard cannot be an idea about the good life (i.e. a first-order value). Note that there is a difference between a second-order value being biased towards a first-order order value with regard to its outcome (which is unproblematic from the perspective of anti-perfectionism) and a second-order value being defined on the basis of a first-order value (which is troubling from the perspective of anti-perfectionism).\textsuperscript{18}

In order to avoid this apparent spill-over of an ideal of the good life into a theory of autonomy, we must have credible guarantees that the second-order value autonomy is not defined from the perspective of an idea of the good life.\textsuperscript{19} One such guarantee would take the form of objective facts about the mental process that leads to values being valued by an agent. However, there is disagreement about which features of this process are relevant and the extent to which such features must be exercised for a person to be autonomous. Although
some such accounts can be said to be objectively misguided (because they are psychologically impossible) there remains disagreement within the realm of the possible. And it is not clear what an objective standard for, say, the amount of reflection that is required for autonomy would look like. Another type of guarantee would be evidence for a minimal common ground among theories and intuitions about liberal autonomy. If there were such a common ground, then the second-order value of autonomy can be said to be sufficiently independent of a version of the good life. What is more, because finding a common ground means that one does not have to choose among theories of autonomy, the concern for arbitrariness in selecting a theory of liberal autonomy would be no longer relevant.

Let me sum up. The argument of Colburn shows that valuing autonomy is not the same as valuing a conception of the good life, because autonomy is a second-order value that, as such, specifies how ideas about the good life are to be generated. On this basis a case can be made for saying that anti-perfectionism concerns only first-order values. However, different theories may emphasize different aspects of the valuing process, while the selection of a theory of autonomy should not be informed by an idea about the good life. As a guarantee against the suspicion of this spill-over, a concept of autonomy must be sufficiently independent from ideas of the good life, where such independence is contingent on there being a nontrivial minimal common ground among liberal conceptions of autonomy. In chapter 2 I will argue that this is the case.

2 What an Autonomy-based Assessment of Branding Should Look Like

For an adequate assessment of autonomy violation in branding we must know not only the political status of the normative standard of the assessment (i.e. autonomy) but also which structure the assessment must have for it to be convincing. To this end it will prove informative to revisit the debate about the relation between advertising (i.e. the most prominent case of branding) and autonomy in the Journal of Business Ethics.20 My main claim will be that no decisive arguments for or against the claim that advertising undermines autonomy have been made so far, and that this is due mainly to deficiencies of the respective argumentative strategies. On the basis of this analysis I present a comprehensive argumentative strategy that will be pursued in the remainder of this thesis.

The First Case for Advertising in the Business Ethics Debate

The debate on advertising and autonomy in the Journal of Business Ethics was opened by Robert Arrington, who comes down “tentatively in favor of advertising.”21 He arrives at this conclusion by applying to advertising a framework for thinking about autonomy that consists of four components. I will here briefly run through Arrington’s argument and show that it is at the very least not a knock-down case against thinking that advertising undermines autonomy.

First, Arrington follows the philosopher Harry Frankfurt in defining the formation of autonomous desires as the second-order reflective approval of a first-order desire. Thus, a desire to buy steak sauce is autonomous if a person upon reflection wants to have that desire. By this standard, Arrington claims, consumers will have predominantly autonomous desires for products because they “often return to purchase the same product over and over again,
without regret or remorse”. For this fact to be an argument, Arrington must assume that if a behavior is repeated, this is evidence of it being informed by a desire that has been approved of upon higher-order reflection. But this assumption seems unwarranted. For it is possible that Arrington’s consumer simply failed to form any second-order desire with regard to the desire to buy the product or that she acted against her second-order desires.

Arrington goes on to maintain, secondly, that reliable information is a necessary condition for making autonomous choices. In response to concerns by cultural critics about false information and puffery he claims that these types of advertising need not imply that advertising in general is deficient in providing relevant information. At least puffery could according to Arrington inform the consumer about feelings associated with a product, and this information may be deliberatively processed as a reason to buy the product. The factual assumption of this argument may be correct, but it does not show that advertising does not undermine autonomy in typical cases. Consumers may of course sometimes process feelings about a product deliberatively, but there is no good reason for assuming that this will always happen. For this to be the case it must be true that only deliberative processing informs consumer decisions, so that the market mechanism will weed out information that is not relevant for such decision making. However, Arrington does not give evidence for thinking that this assumption is true (and in chapter 4 we will see that it is in fact false).

The third component of Arrington’s view on autonomy consists in acting for a reason or being open to reasons, were they to present themselves. That is, if we have a reason for action but fail to act on it, then we act impulsively and this reduces our autonomy. In this regard Arrington acknowledges that consumers sometimes act impulsively and that they may do so on the basis of desires that are the result of advertising. However, he denies that advertisers can be blamed for this, because they do not intend to implant desires. This argument is closely associated with the fourth component of autonomy, which concerns manipulation.

For manipulation to occur, Arrington claims, three conditions must be met. Uncontroversially, a controller must intend that P acts in a certain way A, and – secondly – her intention must be causally effective in bringing about A. Less obvious is the third part of the definition, which holds that the controller must intend to make sure that all the necessary conditions for A are satisfied. Arrington’s rationale for adding this condition is that is allows one to distinguish a controller from a teacher. A teacher also wants to influence persons, but she makes such influence contingent on the student wanting to learn something. A controller in front of a class would be different in that she would not want the success of her influence to depend on the desire of students to learn. This contingency on existing desires (e.g. the desire to learn) is then generalized to all desires. Hence, if an influencer produces persuasive messages that build on, say, an existing desire to be powerful, then no manipulation occurs. By contrast, Arrington continues, if her messages appeal to a desire for power that she herself implanted in a person, then she manipulates. On this basis Arrington concludes that advertisers typically do not manipulate consumers because they normally target segments of a market and hence intend messages to have an effect when consumers already have certain desires.

Something clearly goes wrong here. After all, the fact that a message appeals to an existing desire does not seem to guarantee freedom from manipulation. If I subliminally associate an attitude object (or a goal) with one of your desires and do so in order to make you positively evaluative the object (or to pursue the goal), then this seems to be a case of manipulation regardless of whether the desire with which the object (or goal) was associated was already there. If this is correct, then it is rather the type of processing by which the relation to an existing desire is established that seems to matter. The intuition in the teaching exam-
ple would on this reading be pumped by the fact that the desire to learn will result in deliberative processing. If this is correct, then teaching is not different from control because any one desire is respected, but because respect for the desire to learn means that messages are crafted in a way that facilitates reasoning.  

Arrington concludes on the basis of his assessment of these four components that advertising has the “potential to harm a few poor souls” in that it incidentally violates the autonomy in all four respects, but that it is “innocent of the charge of intrinsically or necessarily” violating autonomy. If my assessment of his arguments is correct, then this is an overhasty conclusion at best. This conclusion is shared by a number of critics of Arrington who have attempted to argue in favor of the proposition that advertising undermines autonomy.

The Critics of Advertising in the Business Ethics Debate

The first reply to Arrington’s argument was the contribution of Paul Santilli, who focused exclusively on the issue of rational choice (Arrington’s third point). Santilli makes a distinction between real needs (arising from an “objectively real deficiency”) and emotional desires (arising from a “felt deficiency”). Specifically, he thinks that a desire is only legitimate if it is “preceded by a knowledge of actual needs” where such knowledge is the result of a “rational understanding”. The problem with advertising, on this view, is that it appeals to emotions on a day to day basis, and undermines the capacity of consumers to understand their real needs (and so to form desires on the basis of these needs). Therefore Santilli wants to abolish all persuasive advertising (which appeals to emotions) and allow only for informative advertising. Although Santilli does not spell out the notion of autonomy on which his critique of advertising is based, it in any case relies on a shaky psychological foundation. For it is not clear whether desires can be formed on the basis of rational understanding in the first place, let alone that all of one’s desires could be formed in such a way.

In a second critical reply to the Arrington article Roger Crisp maintains that the former has set the bar for autonomy violation too low if we take seriously our common sense understanding of autonomy. On such an understanding, Crisp continues, autonomous desire formation requires at least three things. First, we must be aware of the attempt to influence our desires, so that subliminal attitude change would be a pro tanto case of autonomy violation. However, and this is the second point, if the influence is wielded for a good reason, then it does not violate autonomy. Thirdly, desires must not be transmitted in a way that leads them to trigger action while bypassing “ordinary processes of decision-making”. On the basis of these criteria, Crisp revisits the cases brought up by Arrington and argues that persuasive advertising – insofar as it has unconscious effects – leads to desires and decisions that are not autonomous, because there is no good reason that justifies such influence.

For this conclusion to be convincing, Crisp must indicate what counts as a good reason for unconscious influence. His proposal is to say that a good reason is one that people accept as such. He illustrates this position with a distinction between behavior in the realm of aesthetics and economics. An actor, Crisp notes, uses “unconscious body-language” to improve his performance, but if the audience were informed afterwards, they are unlikely to resent this fact. Instead, they will see it as a relevant quality of the performance. By contrast, if consumers are told that their desire to purchase an item was manipulated, they will resent this fact. However, for this criterion to be practically useful it is not good enough to know after the fact whether manipulation occurred. That is, for a theory of autonomy in advertising to be relevant, both practitioners and state regulators need to know when unconscious influence is justified in advance. And because a manipulator cannot know in advance whether people will regret unconscious influence, Crisp needs to introduce some standard for a good
reason. As such he could propose to make the definition of a good reason dependent on empirical facts about regret with regard to manipulation in specific domains. If it were to turn out that people typically agree to manipulation in art, then this would give artist a license to wield such influence. A problem with this approach though is not only that there may be a minority that does mind the manipulation, but also that one and the same individual may reach different judgments about the appropriateness of the manipulation on a case by case basis within a domain.34

In a third reply to Arrington’s contribution, Andrew Sneddon claims that the former’s assessment is incomplete because it only focuses on choices. For a comprehensive assessment, Sneddon maintains, one must also take the autonomy of a life as a whole into account. The latter approach is referred to as ‘deep autonomy’ and consists of two components. First, a person must assess whether a first-order desire is consistent with her values. That is, she must be able to relate such a desire to the values that represent the “structure of her life” and check whether it is coherent with it.35 Second, it must be assessed whether this value structure (with which desires are checked for coherence) is valuable. For the latter assessment, a person must compare her own value structure with value structures of other people. Sneddon goes on to suggest that for the second type of evaluation to succeed, an agent must know about a wide variety of possible ways of living, because the assessment of our values will be more significant if we compare them with more (and more varied) value structures. Furthermore, knowledge of different values must be acquired in interaction with people who live their life according to them, because this would be how our ‘horizons’ of values are formed.36

Sneddon next argues that if this ‘strong evaluation’ is a requirement for autonomy, then in order to evaluate the value of our values we must meet three conditions. We must know about different values, we must be motivated to reflect about them and different values must inform the lives of real agents. It is then argued that advertising compromises all of them. With regard to our knowledge of values it is argued that advertisements collectively transmit the value that the consumerist way of life is the good life and that this would diminish the opportunity for learning about “socialist, agricultural and hunter-gatherer ways of living”.37 Because the transmission of the consumerist way of life is supposedly so pervasive – aided by processes of social pressure to conform – it would moreover diminish our motivation to reflect about values in the first place. Finally, Sneddon thinks that the transmission of advertising messages will cause those people who lead a nonconsumerist way of life to give it up, so that there will be less variety of life forms.

Because Sneddon gives no empirical evidence to back up any of these claims it is not clear whether these processes do in fact occur, and if so, to what extent. But even if they do occur to some extent, then it is implausible that this would compromise autonomy. First of all, it is not clear why we could only learn about other values in interaction. Learning about them in history book, novels or movies would seem to deliver sufficient exemplary information. Second, there seems currently to be a staggering variety of values that inform people’s lives in societies with advertising, so that even if Sneddon could indicate some plausible threshold of variety in values, such societies are unlikely to fall short in that regard.

The Second Case for Advertising in the Business Ethics Debate
In response to these critiques of advertising, Anne Cunningham has made a case for the compatibility of advertising and autonomy. She does so by rejecting both the theories of Harry Frankfurt (on which Arrington and Crisp rely to some extent) and Sneddon as inappropriate for understanding the relation between advertising and autonomy. Frankfurt’s
theory would be deficient because of its often noted inability to explain why second-order desires would speak for the agent.\textsuperscript{38} The theory of Sneddon is said to be unattractive because it relies on the ideal of the reflective life – in which decisions in life must be the outcome of a reflective process – which is a contested account of what autonomy is about. As Cunningham has it, “given that we individuals function within a socially constructed world and tend to abide by, often without questioning, its socially constructed norms and values, it seems none of us would, according to Sneddon, be autonomous”.\textsuperscript{39} Moreover, Cunningham notes that the framework of Sneddon cannot distinguish advertising from other institutions such as religion, which has the same capacity for shutting out different perspectives.

As an alternative, Cunningham presents a theory of autonomy developed by Robbert Noggle to account for autonomous desire formation.\textsuperscript{40} I will discuss Noggle’s theory in greater detail later on.\textsuperscript{41} At this point I note only that for him a desire is not autonomous if it is based on a belief that is not properly related to or concordant with the integrated set of belief of an agent (as the result of conditioning for example). On the basis of this theory, Cunningham asserts that “advertising, though powerful enough to shape our beliefs, does not create the kind of discordant quasi-beliefs that violate one’s autonomy”.\textsuperscript{42} This is true if advertising either does not involve any conditioning, or if all effects of conditioning in advertising result in quasi-beliefs that are concordant with the integrated beliefs of the agent. The first of these assumptions is false (as I will show in chapter 4) while the second is difficult to prove because it will depend on the nature of the belief system of an individual.\textsuperscript{43} Cunningham’s still makes a case for the truth of the second assumption in that she maintains that according to many of its critics advertising is an ideology and as such a highly integrated belief system itself. As such, being more heavily influenced by advertising would supposedly lead one to have a more integrated set of beliefs, which would in turn make one more autonomous. Setting doubts about the plausibility of this standard of autonomy aside for the moment, the fact that some critics of advertising assert that advertising is an integrated belief system does not make that assertion true.\textsuperscript{44} Moreover, the wide variety of brand images and lifestyle advertising that are in fierce competition on the market strongly suggests that advertising is by no means an integrated belief system. If this is correct, then Cunningham’s case for the absence of autonomy-violation in advertising is also not convincing.

Lessons From the Business Ethics Debate

It is safe to say then that the debate on advertising and autonomy in the \textit{Journal of Business Ethics} has not settled the matter. Not only is the way in which theories of autonomy have been applied to advertising not convincing, but participants have also failed to give reasons for thinking that their standard for the assessment of advertising (i.e. their theory of autonomy) is to be accepted. However, although we may not learn much about autonomy-violation in advertising from this debate, there is a lot to learn with regard to the strategy by which such an argument should be constructed. As such I will here discuss three limitations of current argumentative strategies and on that basis present an alternative strategy which remedies these. This alternative strategy will function as the blueprint for the remainder of this thesis.

A first limitation is that participants in the business ethics debate have failed to justify the concept of autonomy on which their analysis is based. This is true not just for Crisp and Santilli, who rely on common-sense notions of autonomy, but also for Arrington, Sneddon and Cunningham, who simply pick a theory of autonomy and do as if that is all there is to autonomy. Admittedly, Cunningham has made a start in criticizing rival conceptions of autonomy, but she does not explain why the theory of autonomy on which her analysis is
based is to be preferred. Moreover, if one starts to compare theories of autonomy one must specify criteria with an eye to the context in which autonomy is to function as a normative standard. In my case, I am interested in the political consequences of the relation of autonomy to commercial promotion from the perspective of the liberal state. Hence, I will in chapter 2 make an independent case for a comprehensive minimal concept of liberal autonomy that is to serve as the standard of my subsequent assessment. It will furthermore be true of such a concept of autonomy that it is of value for the purpose of the assessment of branding to the extent that it defines precise empirical thresholds.

A second limitation of the contributions to the business ethics debate is that their scope is too restricted. Insofar as advertising is discussed by Arrington, Crisp and Santilli, they discuss individual cases such as the advertisement for ‘A1 stake sauce’ or the one for the ‘newest Mazda’. It is therefore unclear to what extent their assessments can be generalized. Arrington and Sneddon purport to analyze the relation between autonomy and structural features of advertising, but they do not spell out what these structural features are. Granted, Santilli, Crisp and Sneddon distinguish (implicitly or explicitly) between informative and associative advertising, but they do not offer empirical criteria for this distinction. By way of an alternative, I will in chapter 3 define what will count as a structural feature of branding, indicate the type of structural features that are normatively relevant (on the basis of the minimal concept of autonomy developed in chapter 2) and determine whether branding has any such features.

Third, if commercial promotion were found to (structurally) undermine autonomy, we must ask whether it is special in this regard. This concern was raised by Cunningham in response to Sneddon’s position. If advertising were to undermine autonomy because of properties that also pertain to institutions such as religion, then the argument contradicts very strong intuitions. If the goal of political philosophy is to strive for reflective equilibrium, then such a conflict might point one back to the drawing board. In chapter 5 I will therefore apply the conceptual apparatus developed in chapters 2 and 3 to religious and political communication in order to test whether it leads to intuitively implausible conclusions.

For now we must take a step back from this panoramic view of the argument that lies ahead and proceed with the difficult first step of identifying the minimal components of liberal autonomy.
2 Minimal Liberal Autonomy

The primary aim of this chapter is to construct a minimal theory of autonomy that is maximally inclusive of intuitions and theories about autonomy. A subsidiary aim is to formulate such a theory in the language of psychology. If I were to meet these aims, then I would have rendered plausible autonomy-minded liberalism and have at my disposal empirically precise standards by which to assess autonomy violation in branding. In order to arrive at such a theory we must know at least two things: we must know what requirements such a theory must meet (i.e. we must know its design specifications) and we must know which arguments can move us beyond the current controversies and make us meet these requirements (i.e. we must have an argumentative strategy).

With regard to the design specifications I propose that three uncontroversial conditions must be met: maximal intuitive plausibility, consistency and realism. The first principle (which was introduced in chapter 1) simply states that the theory that is able to incorporate most intuitions – provided subsequent conditions are met – is the best theory.1 The additional two specifications are limiting principles on the expansive drive of the first.2 The second principle of consistency also hardly requires elaboration. Intuitions about autonomy often clash and at least a political theory must explain with which intuition to side in prototypical situations. Political theory must give such explanations because these generate political legitimacy, and the explanations that are best at generating legitimacy are (at least minimally) rational ones (and hence consistent ones). The third principle of realism stipulates that the elements of a minimal theory of autonomy must be at least compatible with what we know about the workings of the human mind. There is an important reason for this requirement in addition to my demand for empirically precise standards for the analysis of branding. Because our intuitions about autonomy have not been shaped on the basis of a solid understanding of psychology, they may make demands on us that we cannot actually meet. A minimal liberal theory of autonomy must reject such intuitions because it makes no sense for the state to protect or promote mental states that citizens cannot attain. In order to avoid such a result, a minimal theory of autonomy must be conservative with regard to the principles it accepts and require evidence from psychology for thinking that they are realistic.

My argumentative strategy is to first select from among the intuitions about autonomy the ones that are minimally plausible and relevant to liberal autonomy. Where possible, I integrate intuitions, so that I end up with a set of seven core intuitions (section 1). Because these intuitions are interrelated, assessing which theories of autonomy are best in realistically and consistently making sense of them should preferably be done in some higher-order context. For this purpose I propose that the seven intuitions coalesce into two clusters. The first consists of agentic properties that make no reference to a self. I call this set self-control (where this term is understood in the way that Daniel Dennett uses it, so that the word ‘self’ does not refer to an element that causes anything) and discuss how they fit together in sec-
tion 2. The second cluster of intuitions goes under the header of authenticity and adds to the framework of self-control intuitions about the self (section 3). In the final section I discuss a number of factors that moderate the need for liberal state action in defending autonomy and so establish minimal practical thresholds that are in line with intuitions about appropriate political action.

1 Intuitions about Autonomy

In order to find a common ground among conceptions of autonomy it is important that all relevant intuitions and theories are considered. Otherwise my selection might be biased and that would undermine the credibility of the guarantee of anti-perfectionism that this chapter aims to offer. Because the philosophical literature on autonomy is substantial, it is most fortunate that I have been able to rely for this purpose on earlier surveys of the literature. From these surveys one can deduce two types of lists. One can make a list of different approaches to autonomy and one can make a list of the intuitions of what autonomy is about.

With regard to the first list, discussions about the correct approach focus on whether the history of mental states matters for autonomy and whether autonomy is a local or a global phenomenon. A number of authors maintain that the history of mental states is irrelevant insofar as this history does not strengthen capacities that presently advance the autonomy of an agent. However, this position is hardly tenable in light of an intuition that has been brought out by Alfred Mele in a thought experiment. Imagine two persons who have exactly the same psychological make-up so that they are just as open-minded, flexible and rational. Let us call them Ann and Beth. Suppose that Ann is autonomous, whatever that may turn out to be. Suppose that Beth is different in that she has been subjected to brainwashing so that her entire memory system has been wiped out and replaced by a new coherent set (which – remember – is identical to Ann’s). If this were the case, then we would clearly not think of Beth as autonomous. And because her psychological states are identical to those of Ann, the only difference that can drive this intuition is the difference in the history of the states in question.

Another disagreement concerns the relation between local autonomy, which is a property of mental states and actions, and global autonomy, which relates to life as a whole. Some think that the autonomy of a life can be known simply by adding up instances of local autonomy. Others disagree and maintain that the autonomy of a life depends on it being lived according to life-guiding values that have been endorsed (in an autonomy-conferring fashion). Fortunately, I can sidestep this controversy because taking a side on this issue does not make a practical difference from the point of view of the liberal state. That is, because – as I will argue later on in this chapter – the liberal state cannot (ethically) keep track of whether people live in accordance with properly endorsed values, it can only foster the critical skills that citizens need for evaluating their values. And the same skills (which on most accounts involve some type of reflection) are required for evaluating values on the local and the global account of autonomy.

The second list concerns conceptions of autonomy. Some of these are clearly not relevant to liberal autonomy and include obscure perfectionist conceptions of autonomy, non-liberal conceptions of political autonomy and conceptions of moral autonomy.
to the remaining conceptions it can be established that they are driven by seven core intuitions which cannot be reduced to each other. The first intuition is about the input into the reasoning process and requires that the agent has accurate beliefs about the world. Intuitions two to four are about the reasoning process itself and concern the point of view of the agent, the type of reasoning and its coherence with self-related states. Distinct from these intuitions is the notion of rationality, which is (typically) a standard by which the output of such reasoning is evaluated. Whether this output causes behavior depends, and this is the sixth intuition, on impulse control (i.e. the opposite of weakness of will). Finally, in all of these respects the agent can be more or less autonomous depending on her independence from the interference by other agents.9

If it is true that these intuitions are relevant to liberal autonomy as well as not reducible to each other – which I take to be uncontroversial – it can still be objected that some important intuitions are missing. The most prominent such omissions – espoused mostly by feminist philosophers – are intuitions which hold that autonomy is about having certain intellectual and physical skills, self-perceptions and positions in social relations. There are two ways in which these elements can be said to be relevant to liberal autonomy. The first is uncontroversial and concerns the relevance of skills, self-esteem and social relations for the exercise of any one of the seven components of autonomy that I have outlined above. It can for example be argued that certain intellectual skills are necessary to assess options, which can be argued to be a necessary component of the type of reasoning that confers autonomy.10 Similarly, it can be maintained that endorsement (as a type of reasoning or as coherence) only occurs when the agent has a certain level of self-esteem, so that such self-esteem is derivatively relevant to autonomy.13 And because self-esteem is contingent on certain social relations, it can by extension be argued that these social relations are a condition for the development of a condition for autonomy (i.e. self-esteem) and in that sense relevant to autonomy.12 I accept all of this, but because theorists who push this line of argument have yet to connect psychological work on self-esteem and skills to empirical thresholds of components of autonomy I am not able to rely on this work in the remainder of my argument.13

The other and more radical case for the importance of skills, self-esteem and positions in social relations to autonomy maintains that these phenomena are constitutive of autonomy.14 Even if we suppose that this position is intuitively plausible, which is very debatable, it is not well suited to support a theory of liberal autonomy, since there is no non-perfectionist way to determine which skills, level of self-esteem or positions in social relations are conducive to autonomy. If one wanted to do this, one would either have to outline a set of goals that foster autonomy and then determine which skills, level of self-esteem and positions in social relations facilitate the realization of these goals. But such a move would introduce first-order values (for the selection of the goals) into the theory of liberal autonomy, which is irreconcilable with anti-perfectionism.16 Or one would have to demand that an autonomous agent must possess all the skills and self-esteem, and be in all of the types of social relations that the proficent pursuit of autonomously chosen goals requires. But such a conception is surely absurdly demanding.17 If these approaches are indeed both implausible (and exhaustive), then having skills, self-esteem and positions in social relations cannot – realistically speaking – be constitutive of liberal autonomy. Because these phenomena are not only hard to define in terms of thresholds but also clearly something different from psychological autonomy I will call them freedoms.18 I agree that on the basis of a minimal theory of autonomy a good case can be made for the protection of such freedoms in terms of rights and the distribution of freedoms in terms of primary social goods. But that is not to say that having
more freedoms increases one’s autonomy (once we disregard their effect on fostering the seven components of autonomy).

If it is agreed that the seven intuitions on my list are relevant and irreducible while it is also granted that the list is comprehensive, then the hard questions are which conceptions best make sense of the intuitions and how intuitions relate to each other.\textsuperscript{19} I take it that these questions cannot be answered independently. That is, if it is true that the meaning of intuitions is contingent on their relation to other intuitions, then one should discuss conceptions of intuitions in the context of an integrative framework that at least roughly indicates how intuitions fit together. In this regard I follow John Christman in thinking that attempts at such integration have so far predominantly taken one of two forms.\textsuperscript{20} Some think of autonomy as a special type of control without seeing any role for the self. In terms of my taxonomy, autonomy as control is made up of a cluster of four core intuitions: epistemic accuracy, type of reasoning, rationality and impulse control. Autonomy thus conceived involves flexibly and effectively responding to the world on the basis of accurate knowledge. Daniel Dennett has offered an influential outline of such an account and I will follow this outline as well as his terminology in my discussion of autonomy as self-control (see section 2). Others have conceived of autonomy as a special relation to the self on the basis of the two intuitions coherence and point of view. I refer to this cluster as authenticity and argue that it is a legitimate part of liberal autonomy if we conceive of it much more minimally than is common (see section 3). The two types of autonomy imply different conceptions of independence. On the self-control account, independence refers to the absence of another agent taking over control of behavior by interfering with any one component of self-control. On the authenticity account, independence refers to the absence of interference in the formation of a point of view or the process of maintaining coherence.

I need to add one more aspect to this framework. If it is plausible that a person who is hospitalized and unable to communicate can still be autonomous, as Alfred Mele argues, then autonomy must be about more than just actions.\textsuperscript{21} Hence, I am going to say that autonomy can be a property of mental states as well as a property of an action. But what about influence on mental states which in turn issue in action? In this regard I will say that influence affects autonomy as a property of mental states insofar as the changed mental states are stored in memory before issuing in an action, and that influence affects autonomy as a property of actions insofar as the changed mental states issue in an action before being stored in memory (if at all). On the basis of these distinctions my aim in the upcoming two sections will be to find a minimal common ground on self-controlled actions (2.1) and mental states (2.2) as well as authentic actions (3.1) and mental states (3.2).

\section{Self-control}

A system that controls something must meet three minimal conditions. It must have a desire about the state into which the controlee is to be driven. It must have accurate relevant knowledge of the world. And it must be able to deliberate, so that the desire and the beliefs can be adjusted. By corollary, if a system drives itself into a different state, we may speak of it exercising self-control. And were we to ascribe subsequent learning capacities and higher order thinking to it, then we would have something approaching the self-control that humans typically exercise. There is of course the possibility that other agents take over control, but as
long as this influence is known by the influenced agent and enters into the deliberation process, self-control is preserved. This, at least, is the analysis of Daniel Dennett. Pace skeptics on the prospects of agreement on a minimal theory of autonomy I will argue that such agreement is possible for self-control, as understood by Dennett, in terms of action (2.1) and attitude formation (2.2). I will come to disagree however with the view, espoused by Dennett as well as Bernard Berofsky, that self-control (if sufficiently well-informed, open-minded and flexible) would be all there is to autonomy (2.3).

2.1 Self-controlled Action

A liberal account of self-controlled action must specify four intuitions. It must specify the type and amount of reasoning (2.1.1), the rationality of this reasoning or its outcome (2.1.2), strength of will (2.1.3) and the independence from other agents (2.1.4). In the context of this thesis I cannot hope to give an exhaustive account of all these properties of self-controlled agents. I will therefore bring the criteria of assessment (as stipulated by my design specification) to bear only on intuitions with regard to which there exists disagreement that threatens the claim to non-perfectionism of a minimal theory of liberal autonomy (e.g. responsiveness to reasoning), or ones that are prima facie relevant to the assessment of branding and thus require empirical specification (e.g. independence).

2.1.1 Responsiveness to Reasoning

Although all theories of autonomy agree on the importance of reasoning, there is disagreement on the type and amount of reasoning that is to be required. Obviously, the amount of reasoning that a minimal theory can require will depend on what the reasoning should look like. In this regard two theories give very different accounts of both the mode of reasoning and its object. On the evaluative account, reasoning must involve normative judgment and is typically about (but need not be restricted to) courses of action. On the hierarchical account, reasoning need not be normative, but its object must be desires (courses of action will not do). Because a liberal theory of autonomy must meet demanding standards of realism, I will assess both approaches with an eye mainly to their psychological feasibility. In this regard it will prove instructive to point out in advance that there is a growing consensus among psychologists for making a distinction between system 1 and system 2 processes. System 1 processes are for the most part unconscious, show little individual variability and operate on the basis of parallel processing, which allows them to incorporate a lot of information and to process it fast. The system 2 process is sequential, conscious, slow, rule-based and accurate. Moreover, whereas system 2 is thought to consist of a single system, system 1 is likely made up of many systems. I am going to assume from here on that some distinction of this sort is correct, and that what theorists of autonomy are concerned with when they demand that people reason and reflect is that they use system 2.

The Evaluative Account of Reasoning and Autonomy

The most influential account of the idea that autonomy is conferred by evaluative reasoning has been developed by Gary Watson. He claims that autonomous action results from judgments about what it is good to do, all things considered. Such judgments are the result of reasoning that combines beliefs (and probability estimates) with considerations about the good (the latter set of considerations makes up what Watson calls the ‘valuational system’). Actions that lack autonomy are ones that are caused by motivational states that do not result
from reasoning or do not result from reasoning that relies on input from the valuational system. These three categories – unreflective motivations, non-normative reflective motivations and normative reflective motivations – together make up the ‘motivational system’. An agent is perfectly autonomous if the valuational and the motivational system are identical.26

There are three types of incrementally serious objections to this account that are presently relevant. The first concerns Watson’s assumption that the output of normative system 2 reasoning is motivational. In this regard he explicitly sides with Plato (and against Hume) in assuming that when we think that something is good we thereby desire to do it.27 However, there are reasons to doubt such a view in light of evidence from patients with intact system 2 reasoning skills but damage to parts of the brain that are involved in processing emotions. It turns out that there are among such patients individuals who cannot settle on what to do in the real world (such as deciding which pencil to pick), whereas they have no problem in answering hypothetical questions about what it would be best to do.28 This suggests that affect plays a decisive role in practical reasoning. On the basis of such evidence several models of practical reasoning have been constructed that consist of system 1 processes that send signals of affect or pleasure (depending on the model) which are monitored unconsciously by the practical reasoning system.29 If this picture is correct, then acting on the basis of evaluative reflection might look something like this.30 A thought arises in the conscious system 2 reasoning system with the content ‘I shall do x’ (e.g. choose the blue pencil). This globally broadcast inner speech message is then picked up by a specialized module, which infers that a commitment is made to do x. Combined with a desire to act consistent with commitments, the belief that the person committed herself to doing x issues in a motivational state that may lead to the selection of a motor instruction in the practical reasoning system (if it outcompetes other motivations).31 If this is a credible type of account of practical reasoning, then Watson would have to adapt his theory in such a way that autonomous actions are indirectly caused by best judgments (which are beliefs) that are the output of the system 2 reasoning system. That is, evaluative reflection can influence the selection of an action scheme, but it can do so only if it is paired with a desire.

A second objection against Watson’s account, raised by Michael Bratman, is that evaluative judgments underdetermine what is to be done. Such underdetermination applies to cases in which two options are equally valuable (and where there is no value that would enable a comparison) and cases in which the agent is uncertain about what value to ascribe to a course of action.32 Furthermore, if the definition of an evaluative judgment is said to depend on some intersubjective standard, then uncertainty mushrooms, because there are many cases in which it will be uncertain whether rational agents will agree on the value to be given to a course of action.33 Bratman claims that in such situations people still act on the basis of reflection – in which they draw not on value judgments, but on feelings and inclinations – and that they do so in a way that we think of as autonomous.34 If so, then autonomy would be about more than just evaluative reflection.

Bernard Berofsky has gone yet further in claiming that we can think of an entire life in which people act against their judgment of what it is good to do and yet say that such a life is lived autonomously. This alternative conception sees values not as the result of reflection on what it would be good to do, but as stable preferences that reveal themselves in a wide range of (hypothetical) choices that are made in an informed and dispassionate way.35 A person may for example find herself spending most of the time gambling because of the thrill it gives her and the interesting people she meets. She may think her life is not up to her ethical standards, that it is not in line with her ideal self and perhaps even that it is utterly
worthless. Although such a situation would surely lead to a lot of conflict, it need according
to Berofsky not make her gambling life any less autonomous. If this is plausible, and I think
it is, then it would be wrong to tell such people that they are not living an autonomous life.
This is not to deny that autonomous persons have the capacity for reflection on the good, or
that they sometimes use this capacity, but it is to deny that they must live their life in accord-
ance with such judgments. I take it then that on a liberal conception of autonomy there
should be no requirement that people act on the basis of reflections that draw on an idea of
the good.

The Hierarchical Account of Reasoning and Autonomy

Hierarchical theories maintain that autonomy consists in a hierarchy between two conative
states that is established in reflection (e.g. an agent reflectively comes to desire to have a
desire). The intuitive impetus that drives such theories is the notion that it is crucial for
autonomy to reflectively step back from a situation and to reflect on it. In hierarchical theo-
ries this takes the form of a person desiring, upon reflection, to be driven by one or more of
her desires. When she wants such second-order desires to be effective, then she is said – in
the classical formulation by Harry Frankfurt – to have second-order volitions.

A difficulty that Frankfurt notes himself is that in case of conflict the higher-order desires
are to win out, and that there is thus a danger that a chain of ever higher-order desires may
lead to “the destruction of the person”. This chain can be terminated though, Frankfurt
maintains, when the person “identifies himself decisively with one of his first-order desires”
whereupon “this commitment ‘resounds’ throughout the potentially endless array of higher
orders”. This is anything but a satisfying solution though, because it is not clear what such
decisive identification consists in. But if we suppose with Gerald Dworkin that for all practi-
cal intents and purposes people do not carry on reflecting at ever higher levels we can set
aside the matter for the moment. A more pertinent difficulty is that it is not clear from
Frankfurt’s account whether hierarchy does any autonomy-conferring work. An alternative
explanation would be that it is the reflective endorsement (with may or may not involve
conative hierarchy) which pumps the intuition.

In one response to this alternative explanation, Bratman points to underdetermination of
evaluative judgments. In such cases, to repeat, agents may come to a proper reflective deci-
sion that does not rely on evaluative judgments. But although these cases may show that a
non-evaluative commitment has a role to play in a theory of autonomy, they do not show that
such a commitment should be hierarchical. Another strategy of Bratman is to show that
there are cases of higher-order ‘identification’ with a desire that have an autonomy-conferr-
ing effect which can only be explained with reference to hierarchy. He concedes that proper
identification with a desire involves treating it as a justified reason for action. But according
to Bratman there is a subset of cases in which autonomously treating a desire as a reason for
action can only be explained with reference to hierarchy. In support of this claim he notes
two hierarchical accounts of treating a desire as a reason that do not work, and presents a
middle way as the solution.

One dysfunctional account says that a policy must support treating a desire as a justified
reason in deliberation. But then the identification appears to occur at the first-order level in
the desire being treated as a justifying reason in deliberation. Hence, higher-order attitude
support does not seem to add anything. If so, then hierarchy does not add anything. The
other dysfunctional account says that the second-order attitude must support the desire’s
functioning as an effective motive. But then the desire issues in an action (due to the support
by a higher-order attitude) without being treated as a justified reason.
As a middle way Bratman rolls out an ingenious argument to the effect that there can be cases in which we engage in attenuated reasoning and where higher-order attitudes can give the decisive little push. As an example he mentions the behavior of participants in the Milgram experiments. These subjects went along with authority to give electric shocks to a supposed examinee and are assumed by Bratman to have not fully endorsed their action. In such cases, he maintains, the agent engages in practical reasoning and comes to see the desire for something as a “justifying end” without endorsing that the desire function in this way. At this point Bratman introduces his higher-order attitudes to suggest that the gap between treating a desire as end-setting for practical reasoning and true endorsement of it as a justifying reason would be bridged by a higher-order attitude in support of the desire functioning in this way (i.e. as end-setting for practical reasoning). In the Milgram example this would mean that a participant who decides to push the electric shock button after a process of attenuated deliberation that treats the desire to punish as a justified end (simply because she goes along with authority) would not yet be autonomous. But the action would be autonomous if she had a self-governing policy that treated the desire to punish as a justified end in deliberation. In this way the agent still does some deliberation (unlike the case in which the higher-order attitude merely supports a desire functioning as an effective motive), while the higher-order attitude does seem to add something (i.e. treating the desire as a legitimate reason). However, what this argument leaves open is the possibility that what is driving the intuition is that autonomy depends on reflection, and that the higher-order policy support drives the intuition by virtue of that policy being formed on the basis of reflection. On this interpretation, the fact that the relation between the reflection and the desire is temporally mediated by a higher-order attitude would be accidental. In sum, the absence of evaluative reflection in intuitive cases of autonomy does not force us to accept hierarchy, and cases of autonomy in which hierarchy is involved do not rule out the possibility that evaluative reasoning is pumping the intuition.

What is more, even if we grant that hierarchy does somehow confer autonomy it would seem, as Watson noted, that reflection on desires is too rare a phenomenon to be the basis of autonomy. If we only rarely reflect on our desires, then we are only rarely autonomous. At least for a theory of liberal autonomy, that will not do. In response to this concern, Bratman proposes to think of higher-order states in hierarchical theory as plans and policies rather than desires. Plans are generalized intentions that can settle courses of action and coordinate them over time (e.g. keep my teeth clean), while policies specify what such plans amount to in practice (e.g. brush my teeth after breakfast and before going to bed). According to Bratman, policies not only settle actions, but also settle which desires should be considered justified reasons for action in practical reasoning and roughly fix the weight that is to be given to desires in practical reasoning. Such policies are called self-governing policies. In this capacity, self-governing policies (which have the relevant hierarchical structure) co-regulate which desires activate motor instructions. Because such policies operate in the background, there is no need to reflect on every desire that moves us to act. Hence, the fact that we only rarely reflect on desires must not count against our lives being autonomous.

However, we earlier considered evidence for thinking that the weighing of options involved in decision-making occurs automatically and is thus not under conscious control. Hence, it is unlikely that policies can fix how much weight is to be given to a desire or how a desire is to be treated (i.e. as a justified reason) in affective practical reasoning. This is not to say that policies in general can have no influence on practical reasoning. They can have influence if they are about a course of action, and are supported by an innate desire to do what one has committed oneself to doing (by way of forming the policy). But if policies are
about courses of actions, then they do not support a theory of conative hierarchy. It is conceivable that there are self-governing policies that inform normative system 2 reasoning about which desires can count as justifying considerations. And system 2 judgments about courses of actions can in turn – when coupled with an innate desire to do what it is best to do – influence the practical reasoning system. But if this is how self-governing policies work, then the objection that they do not have much influence seems to stand.

In response to this concern it must be considered that Bratman has noted two structural pressures that supposedly drive agents in the direction of relying on second-order desires. One is the need for self-management, understood as the need to suppress desires that compete with reflective judgments for activating motor instructions. According to Bratman, the fact of such competing desires generates pressure to reinforce the (presumed) conative force of reflective judgments with higher-order desires in support of the reflective judgment. The way this is thought to occur is, again, by means of self-governing policies that contain the instruction to “give (refuse) justifying significance to consideration x in motivationally effective practical reasoning, in part by giving (refusing) such significance to relevant first-order desires”.49 But I just argued that such policies can only influence system 2 reasoning, which is not itself motivationally effective. If that is true, then hierarchical policies apparently do not help in solving self-management problems, because these problems seem to be located not within system 2 (i.e. the formation of the judgment of what it is best to do) but between system 2 and practical reasoning (i.e. ensuring that the best judgment activates a motor instruction).

A second pressure for hierarchy is supposed to be exerted by the underdetermination of value judgments that we discussed earlier. However, we saw that such cases only demonstrate that there is pressure to make some commitment in favor of a course of action (that can be said to speak for the agent) but that underdetermination exerts no pressure on such commitments for being hierarchical ones.50 Taken together then, the evidence for thinking that conative hierarchy is a common feature of human action is not convincing.

Absent decisive arguments for the evaluative or hierarchical accounts of autonomy, what remains is the strong intuition that autonomy involves system 2 reasoning that produces a belief about what it is best to do (which is different from a belief about what it is to act in accordance with the good). Although this intuition is uncontentious, it is not clear how much of it an autonomous agent should display. Given the fact that much of our everyday behavior consist of habits, and given the indirect nature of system 2 influence on behavior, it would seem that the role of reflection in a theory of autonomy should be modest. In this spirit Ben Colburn proposes to say, as we saw earlier, that a person is autonomous if she would still have carried out an action had she reflected on it. The problem with this approach – implicit in my earlier remarks about this idea – is that there is no way for the liberal state to know whether people would have reflectively endorsed an action of theirs. As a result, it is impossible for the state to accurately foster or protect autonomy, since it cannot measure either progress or transgressions of thresholds.

If it is agreed that some reflection matters for liberal autonomy, but that it is impossible to determine a minimal amount of reflection (e.g. in the form of hypothetical reflection), then the best we can hope for may be to require that people have developed the capacity to reflect. This is the approach advocated by Gerald Dworkin.51 In practical terms this means that there is a role for the state to foster via education basic reasoning skills, but that monitoring the amount of reflection citizens engage in is (absent violations of independence) not the business of the liberal state.
2.1.2 Rationality
I can be very brief with regard to autonomy and rationality. There is wide agreement that Kantian views of autonomy as rationality are too controversial to serve as conceptions of liberal autonomy and there is no need for me to reiterate these discussions.52 Besides concerns about rationalistic perfectionism there are worries about abuse of state power in the name of rationality. This was the point of Berlin’s tyranny argument discussed in chapter 1.1. In line with such worries it was noted with Christman that liberals are reluctant to employ anything but the most minimal standards of rationality in fleshing out concepts of autonomy. These minimal standards include some level of coherence of desires (understood negatively as the absence of clinically defined levels of incoherence) and the capacity for basic means-end reasoning that allow a person to pursue goals.53

2.1.3 Strength of Will
Whether there can in principle be such a thing as weakness of will has traditionally puzzled many philosophers. Apparently there are cases in which we judge that it would be best to do x but in the end fail to actually do x. And this is an everyday experience rather than an exception: we think it best to write a section but find ourselves surfing the internet. This is a puzzling state of affairs if we assume that a best judgment always settles a decision. However, there is no good reason to think that this must be so. One plausible explanation of what is going on is that system 2 reasoning does not settle what is to be done, but that such decisions are made by system 1 processes. If so, then the system 2 conclusion that it is best to do x must be tagged with a – possibly innate – desire to do what is best, and this desire may lose out against other desires in the practical reasoning process.54 On this account, weakness of will occurs when the desire to do what is judged best loses out against rival desires.55

If this is somewhat plausible, then the serious question is not whether weakness of will occurs, but how much of it a self-controlled person can be allowed to display.56 In this regard Christopher Suhler and Patricia Churchland have offered a neurobiological account of self-control, which consists of two components.57 The first is anatomical and concerns the unimpeded functioning of areas of the brain involved in “prototypical” cases of “good control”.58 In this context the authors refer to the central role of the prefrontal cortex in maintaining goal-related patterns of activity and sending biasing signals to several parts of the brain to bring the goal in question about. The second component is about neurochemicals, which must be “calibrated to the spectrum of values that the brain evolved to maintain in response to environmental demands typical of the species’ evolutionary past”.59 If serotonin levels are low, for example, people are more likely to engage in aggressive and impulsive behavior, and if there is a delay in the return to baseline of the corticotrophin releasing factor, recidivism of addicts is more likely.60 If we assume that our evolutionary past was stable and relatively uniform (which is not uncontroversial), this approach should potentially be able to offer the kind of verifiable criteria for self-control that a minimal theory of autonomy demands.61 Specifying such criteria is certainly not a task for the present occasion however.

2.1.4 Independence
Self-control requires that our actions are not determined by others outside of our control.62 In this section I discuss interference insofar as controllers influence behavior directly (influence over behavior that is mediated by attitude change is discussed in the next section). Controllers can do so in three ways. They can activate a goal outside of awareness, they can
disturb a controlee’s reasoning process and they can deplete her resources for exercising self-control.

*Unconscious Goal and Trait Activation*

Activating a goal or a trait outside of awareness can occur at two levels. In subliminal influence the stimulus itself that activates the goal or trait is not perceived, whereas in instances of supraliminal priming only the process by which a presented stimulus influences behavior is not perceived. The first route – of subliminal influence – is the clearest method of taking control outside of awareness.\(^53\) The acceptance of this phenomenon relies on the acceptance of a distinction between an objective threshold for perception (i.e. when a stimulus is discriminated by the senses) and a subjective threshold (i.e. when subjects report awareness of perceiving the stimulus). One can of course say that all subjective thresholds are unreliable and so define away the possibility of subliminal influence.\(^54\) But at least for a normative theory of self-control reasoning about influence attempts is relevant, and such reasoning must have access to information that is consciously accessible. If one does not know that another agent has exerted influence (because it did not pass the subjective threshold), then one can also not take the intention of the agent into account. Subliminal priming procedures have been shown to affect drinking behavior, as in an experiment conducted by Ap Dijksterhuis and colleagues. When they flashed the words ‘drink’ or ‘cola’ for 15 milliseconds and then immediately masked the words with a string of letters (as part of an unrelated task) subjects were significantly more likely to accept the offer of a beverage by the experimenter during a pause in the experiment (than subjects in a control condition in which random letters had been subliminally presented).\(^65\)

The best evidence for influence by means of supraliminal priming comes from experiments in which subjects are presented with scrambled sentence tasks in which they are to make correct sentences out of a collection of words. John Bargh and colleagues in particular have achieved remarkable results by using this method. In one experiment some of the words were unbeknownst to the subjects related to aging (e.g. bingo, wrinkles and Florida) while subjects likewise did not know that when the experiment was supposedly over, the experimenters measured the time it took subjects to walk from the experiment room to the one of the conductor of the experiment. It turned out that it took those who had been primed with old age on average well over eight seconds, while the control group needed only seven.\(^66\) In a similar experiment the words in the scrambled sentence test were related either to rudeness or politeness. The response was now measured in terms of the time it took subjects, after they had completing the task, to intervene in a staged conversation between the experimenter and a supposedly different subject, who – if no interruption followed – kept on talking for ten minutes after the subject had finished her task. Of those primed with rudeness 65% interrupted, against 38% for the control group and 17% for those who had been primed with politeness.\(^67\) In both experiments subjects showed no awareness whatsoever on the relation between the content of the words in the scrambled sentence task and their behavior. Hence subjects who are primed outside of awareness are also unable to take the influence of the controller into account and must thus count as deficient in self-control.

*Reasoning Disruption*

Once a message is identified and scrutinized for its accuracy, a self-controlled agent must be able to reason about it.\(^68\) Such reasoning sets a self-controlled agent apart from a creature that has only reflexes. However, reasoning can be disrupted by a controller if she manages to put cognitive load on the controlee’s mental system and press for a quick decision. Putting
cognitive load on a system is achieved by dispersing cognitive resources by means of activating diversionary reasoning. If a decision is made under such conditions, an agent is likely to rely on simple heuristics, such as following authoritative persons or people that are likable.

Fortunately, we have a defense mechanism against such influence attempts in the form of a mechanism that is known as reactance. This phenomenon is conceived of as a system that warns us when we are pressured to make a quick decision (this at least corresponds to one cluster found in questionnaires that attempt to measure the phenomenon). The cognitive underpinnings of the system are unclear, but there is some evidence to suggest that two routes may be involved. In one the feeling that freedom is threatened leads to a motivation to reject a message directly, and in the other the feeling that freedom is threatened evokes the motivation to counter argue, so that the message may be rejected indirectly. In any case, there is significant individual variability with regard to the threshold for the emergence of reactance, and evidence is as of yet not solid enough to make a case for the level that can reasonably be required in standard types of situations. Moreover, feelings of reactance can be artificially repressed in an agent while control efforts are undertaken all the same. The latter process is also poorly understood however, so that it is too early days to indicate both a minimal threshold of reactance as well ethical prohibitions against manipulating it.

Weakness of Will Induction
A third way in which a controller can wield direct influence over a controlee is by depleting self-control resources. Thus, if a controller has the goal of preventing a controlee from engaging in a particular action that the controller knows is caused by a best judgment, she may intentionally deplete the self-control resource of the controlee. If the controlee were to have no awareness of the workings of ego depletion or the intentions of the controller, then her self-control (understood broadly) declines. Likewise, if a controller would know how to inhibit the release of serotonin and so trigger weakness of will the self-control of the controlee would decline.

2.2 Self-controlled Mental States
With regard to self-controlled mental states we need to consider two intuitions. One tells us that autonomy is about epistemic competence and the other sees autonomy as independence from other agents. I can deal quickly with the epistemic competence intuition. On some accounts – most notably on the objectivity account developed by Bernard Berofsky – the level of autonomy is contingent on the extent to which beliefs on which an agent acts are accurate representations of the world. Such accuracy allows an agent to interact effectively and flexibly with the world around her, and this would be what autonomy centrally consists in. For my purposes I can sidestep controversies about this intuition, because there is a consensus with regard to liberal autonomy to the effect that the state should in principle not intervene in the lives of citizens because it believes that citizens do not see the truth. Surely, it is agreed that self-controlled agents must act on the basis of true beliefs to the extent that they are able to act with some minimum of effectiveness, but this demand is understood negatively as the absence of severe clinical conditions of delusion.

In the remainder of this section I will therefore be concerned with the intuition that autonomy is about independence. Specifically, self-controlled mental state formation is impaired when it results from processing in which the intention of a controller is not taken into account (and when the mental state at a later time causes the controlee to behave in a manner intended by the controller). Knowing about the intentions of a controller preserves
self-control if this information serves as input into the practical reasoning system – provided it is not disturbed – and enables the agent to reach a relevant judgment (which may be either to accept the influence or to resist it).\(^74\) Thus there are two stages at which self-control of mental states is secured in the face of influence attempts. The self-controlled agent must first recognize the intention behind a message of a controller and then be able to properly process it. In order to bring out intuitions on what is disturbing about the loss of self-control, philosophers have so far discussed thought experiments that involve grotesque cases of brainwashing. A relevant criticism of this literature is that concerns about manipulation are only practically relevant insofar as it is possible (and the effects are significant).\(^75\) I therefore want to give empirical evidence for thinking that controllers can realistically devise messages in such a way that they cannot be properly recognized (2.2.1) and processed (2.2.2) by controlees.

### 2.2.1 Recognizing Intentions

The simplest way to take control via mental state formation is by changing beliefs. Giving true information to another agent violates self-control if the agent does not take the intention behind the communication into account (and if the processing of the information has a behavioral effect that the controller intended). Such a loss of self-control is more likely if the information is false, because in that case the controlee is less likely to correctly interpret the intention behind the communication (at least if the controlee expected to be told the truth; when confronted with default liars, the reverse would be true).\(^76\) However, if the liberal state is to establish whether an intention to manipulate exists on the part of a controller and whether the controlee could have been expected to recognize this intention, it must have criteria that allow it to determine ethically whether these conditions are met. As I will argue in section 4, establishing whether citizens are misleading each other involves invasions of privacy that are unacceptable from the point of view of liberal autonomy itself. In this regard, properties of messages that induce a type of stimulus processing that evades the recognition of an intention will count as more appropriate. Below I will do three things. I want to give evidence for the existence of such message properties and processing channels; I want to show that stimuli processed in this way can lead to behavior; and I want to demonstrate that such influence cannot easily be corrected.

### Unconscious Influence

We have already seen that it is possible to activate a goal directly by means of priming outside of an agent's awareness. Whether mental states that are stored in long term memory can also be changed outside of awareness is a separate question. For a start, the brainwashing experiments that figure so prominently in philosophical discussions of manipulation are probably not possible. That is, although psychiatrists have in the 1950s reportedly succeeded in wiping out the memories of patients, attempts to instill new attitudes in them by headphones have failed miserably (so that the unfortunate subject had to acquire even their most basic skills all over again).\(^77\) However, there is evidence for less spectacular unconscious attitude change by means of evaluative conditioning (which has been documented when stimuli are presented both supraliminally and subliminally) and non-evaluative covariation learning.\(^78\) Because proving the existence of such processing is crucial to my overall argument, I must ask readers to endure a rather dry and technical excursion into psychology.

Evaluative conditioning (EC) has long been thought of as a process. According to this view, if we like a conditioned stimulus (CS) more because it was paired with a likeable unconditioned stimulus (US), then our liking of the CS was the result of evaluative condi-
tioning. EC is thus seen as the process by which the evaluation of the US transfers to the CS. However, subsequent research has demonstrated that there are many different types of transfer, so that it is now widely agreed that evaluative conditioning is more usefully referred to as an effect rather than a process. Any one of these processes violates self-controlled attitude formation if transfer of affect occurs outside of awareness. That is, not only must subjects be aware that the stimuli are presented together, but they must also be aware of the transfer from unconditioned stimuli (USi) to conditioned stimuli (CSi) and the intention on the part of a controller to present the stimuli in this way. Below I will run through four mechanisms that produce evaluative conditioning effects and assess what is known about types of awareness for these mechanisms, so that I can subsequently draw out the implications for self-controlled attitude formation.

An easy way to test whether subjects are aware of EC effects would be to rely on a measurement technique that is known to measure inaccessible attitudes. Some have argued that implicit attitudes count as such attitudes. Measures of implicit attitudes are distinct from ones that measure explicit attitudes (which are obtained by means of questionnaires in which subjects report their liking towards an attitude object, typically on a scale, while time pressure and cognitive load are usually absent) and of more recent date. They come in several variants, but all of them measure small differences in response time. In the Implicit Association Test (IAT), subjects are asked to categorize an attitude object (e.g. a face) as belonging to one of two categories (e.g. black faces or white faces) which are represented on opposite sides of a computer screen. Key to the setup is that subjects have on a previous task already categorized another attitude object as belonging either to the category positive or negative, which are located on the same position as the categories in the second task. If for example a black face on the second task is categorized faster to the category of ‘black faces’ when it is located on the side of the screen that was previously occupied by the category ‘negative’, then it is assumed that there is a stronger implicit link between black people and negative valence. The strength of this link is represented in terms of reaction time. In the Associative Priming Task (APT) subjects are asked to evaluate an attitude object as good or bad. However, the attitude object is briefly preceded by a prime, which is supposed to evoke an automatic evaluative reaction. If the valence of the prime and the attitude object are congruent, then the attitude object should be associated with the relevant valence faster (and the association should occur slower if the two are incongruent). Again, the differences in reaction time are taken to indicate implicit attitude strength (between the primed word and the evaluation). I will return to what these measurements may refer to later. Here I only want to note that on the basis of a review of relevant studies, Betram Gawronski and colleagues concluded that there is no evidence that people would lack awareness of an attitude because it is implicit. Hence, evidence that EC effects appear only on implicit attitude measurements will not count as evidence for unconscious effects of EC so that we will have to assess direct measurements of awareness for the relevant conditioning mechanisms.

The first EC mechanism to be discussed is implicit misattribution, which is the only mechanism that consistently leads to EC effects outside of awareness. According to the Implicit Misattribution Model (IMM), this EC effect occurs when the affect from the US is mistakenly transferred to the CS if they are presented simultaneously. The experimental set up in which such effects have been found is designed to ensure the absence of awareness (which is supposed to facilitate misattribution). That is, a cover story is presented in which participants are told that they are to play a surveillance guard and are to press a bar as quickly as possible when a designated stimulus appears on the screen (amidst distracting stimuli), while the CS is paired with different USi (with a similar valence). A unique prediction of the
IMM is that the increased salience of the CS but not the US should enhance EC effects, because more salient stimuli are more likely to be the object of attribution. Indeed, when pictures of CSi were made bigger (and hence more salient) the EC effect is found to be larger, but not when pictures of USi were enlarged. Another unique prediction of the IMM is that the affective response to a US will be more likely to be misattributed to a CS if the US is not very evocative (where evocativeness is defined as causing arousal, being highly accessible and having an extreme valence). For if the US is highly evocative one may assume that attribution mistakes are less likely. Interestingly, it was found that EC effects for mildly evocative USi only occurred for participants who were unaware of the US-CS co-occurrence while EC effects of highly evocative USi only occurred for participants with awareness.

A second mechanism that leads to EC effects is stimulus-stimulus association, in which the CS is associated with a valence via a link with the US. This effect is achieved when the same US is paired with a CS (either sequentially or simultaneously). The best evidence for this effect comes from studies in which the US is revaluated. That is, if the valence association of the CS is mediated by its link with the US, then the CS should change along with the US when the US is independently revaluated. And this is exactly what is found. Whether contingency awareness is necessary for stimulus-stimulus EC is contested. Because stimulus-stimulus conditioning effects have been observed for pairs that occurred subliminally there can be little doubt that it can occur outside of awareness. On the other hand, it has also been consistently found that contingency awareness enhances EC effects for CS-US pairings. According to a recent suggestion, the way to resolve this issue may be that only CSi that are already embedded in rich associations require contingency awareness in stimulus-stimulus pairing.

The final two mechanisms occur only when contingency awareness is present and are therefore designated as propositional mechanisms. Pavlovian conditioning leads to EC effects in sequential trials in which for example a picture (CS) is followed by a shock (US). Because the picture predicts the shock, which has a negatively valence, the picture itself is valenced negatively. This effect is, as all Pavlovian conditioning, more vulnerable to extinction effects than other EC effects. Finally, stimulus-attribute association is similar to misattribution but differs in that it only occurs when there is contingency awareness. The boundary conditions for its occurrence have not been studied yet however. But just as Pavlovian conditioning depends on the truth of the prediction that the US will occur, so the belief that the CS has a characteristic of the US should probably be seen as a proposition that can be invalidated by subsequent information.

Of these mechanisms only misattribution is fully independent of (and possibly even be impaired by) contingency awareness, so that its application would be a clear case of self-control violation (provided the intended behavioral consequences occur). For Pavlovian, stimulus-attribute and possibly stimulus-stimulus conditioning it may be a necessary condition that one knows that the US and CS occurred together. Whether this contingency awareness will lead to demand awareness (i.e. knowing the intention behind the co-occurrence) will depend on factors of the context which have so far not been studied. But one would assume that the evocativeness of the CS and the centrality of its relation to the US in the overall message will determine whether the intention to influence attitudes by means of stimulus pairing is picked up.

Evidence for implicit non-evaluative covariation learning has recently emerged as well. When Michael Olson and colleagues in their surveillance experiment replaced evaluative USi such as positive words (e.g. joy) by non-evaluative words (e.g. size) no conditioning effects were observed – which was in line with previous findings. However, when they subliminally primed subjects beforehand with words related to size and ran the same experi-
ment, conditioning effects were found such that the CSi that had been paired with words related to size in the surveillance task were significantly more strongly associated with size after the trials. And as in the case of evaluative misattribution learning, subjects have no awareness of an association between CSi and the US. This suggests that implicit non-evaluative covariation learning does exist, but that it is contingent on the CSi being accessible.88

**Behavioral Effects of Unconscious Influence**

Evidence for mental state formation outside of awareness is not the same as evidence for loss of self-control. For if mental states just sit there (in the mind) and do not move anything, then the receiving agent cedes no self-control whatsoever. It must therefore be shown that mental state formation induced by relevant conditioning procedures and subliminal influence actually leads to behavior (or subsequent conscious thought processes) and under which conditions this occurs. If there is such evidence, then at least the scientifically informed controllers can be expected to know about the behavioral effects of their manipulations.89 A controller may not be able to predict when the controlee encounters the relevant conditions under which mental states influence behavior, but that seems not necessary for taking control.90 Admittedly, there is not a wealth of experiments that demonstrate attitude effects of conditioning mechanisms and subliminal influence which in turn cause behavior. But if we know how conditioning mechanisms have effects on different types of attitudes and how these types of attitudes in turn affect behavior, then we can infer that (and how) conditioning mechanisms can be conduits of control.

Thus, I will first deal with the effect of conditioning mechanisms on implicit and explicit attitudes. Different models make different predictions about this effect. On the strict dual systems account, there are two distinct processing mechanisms with their own output, each of which is thought to have an exclusive effect on either implicit or explicit attitudes. Conditioning mechanisms (which are a type of so called associative processing) are believed to shape only implicit attitudes whereas deliberation about information (which is also known as propositional processing) is believed to shape only explicit attitudes.91 However, later studies have found clear evidence of implicit attitude change after deliberation about an attitude object, as well as explicit attitude change as a result of EC trials.92 That is, associative and propositional processing each have effects on both implicit and explicit attitudes. The real question then is how these effects are mediated.

The Associative Propositional Evaluations (APE) model makes the most precise predictions in this regard.93 It also posits two processing mechanisms but maintains that they are distinguished on the basis of whether the mechanism is regulated by truth values. That is, on the APE account associative processing (which it understands as EC mechanisms and subliminal processing) does not ascribe truth values whereas propositional processing (which it understands as effortful deliberation) does. As in the case of the strict dual systems model, associative processing has an effect primarily on implicit attitudes and propositional processing has an effect primarily on explicit attitudes. However, the model also assumes that implicit and explicit attitudes are established online, and that they tap for content into the same ‘associative store’. Hence, EC and subliminal influence trials can have an indirect effect on explicit attitudes that is mediated by their effects on the associative store (that is tapped into when explicit attitudes are generated). Likewise, propositional processing about a strong argument works with representations from the associative store, which are thus activated and reinforced (and so may be reflected in subsequently generated implicit attitudes). The prediction that can be derived from this model is that propositional processing
should typically have a greater effect on explicit than on implicit attitudes, and that associative processing should typically have a greater effect on implicit than on explicit ones.

Now, a problem with the original formulation of this model was that EC mechanisms were all treated as associative in nature. We have seen that this is not the case. In fact, a meta-analysis of effects of mainly propositional EC trials (using a sequential presentation order and the same US) shows that these have a greater effect on explicit than on implicit attitudes. However, when the misattribution mechanism is activated in conditioning trials, it is consistently found that the effect is greatest on implicit attitudes (while the effects of stimulus-stimulus conditioning are more ambiguous). This suggests that a revision of the original APE model is in order to the effect that associative processing refers only to misattribution conditioning (and possibly some types of stimulus-stimulus EC). This suggestion has been corroborated in a recent test of the APE model. In a series of studies misattribution was induced (by using a variant of the surveillance task and presenting different USi simultaneously with the CS) and pitted against a trial in which information was presented in the form of propositions. Not only were the effects on implicit attitudes greater in the case of associative processing than in the case of propositional processing (and the reverse). It also turned out, as one would predict on the basis of the APE model, that effects of associative processing on explicit attitudes were mediated by implicit attitudes (which completely predicted the occurrence of explicit ones while the reverse was not true) and that the effects of propositional processing on implicit attitudes were mediated by explicit attitudes.

Thus, whereas propositional EC mechanisms have an effect primarily on explicit attitudes, associative EC mechanisms have an effect primarily on implicit attitudes. Interestingly, and this is the second step in the present argument, these two attitudes affect behavior under different conditions. On this point models on attitudes and behavior are broadly in agreement. The MODE (Motivation and Opportunity as DEterminants) model – which is the most influential one – maintains that attitudes can affect behavior in two ways. They can be generated after reflection about the consequences of possible actions or they can affect behavior spontaneously. Which one of these ways is activated depends on the subject’s motivation to process deliberatively and the opportunity to do so (as constrained by time and task demands). When people have little motivation to engage in deliberative processing and little opportunity to do so, implicit attitudes are found to be more predictive of behavior. Explicit attitudes, by contrast, are more predictive when motivation and opportunity are high (although effects of implicit attitudes on deliberative behavior are not absent).

We said that for control to be ceded to a controller via attitude change, the controller must know which type of attitude will be affected by the unconscious influence procedure that she uses and under which conditions the attitude that is predicted to have changed will influence behavior. We can now say what this means in psychological terms. If a misattribution EC trial is used to change implicit attitudes, the behavior that the controller seeks to elicit must occur under conditions of high time pressure or task demand, or with individuals low in need for accuracy.

**Correcting for Unconscious Influence**

A final objection to the present account is that even if attitudes can be changed by a controller in a way that predictably evokes behavior, we should not be bothered too much by this because a controlee can quickly regain self-control. This is the position of Daniel Dennett. He thinks that even if it were possible to brainwash people in the way that is common practice in philosopher’s thought experiments there is not necessarily a reason to be all too concerned about this. For after “a number of years” in which the brainwash victim deploys her
critical reasoning skills she will reach a point at which we would call her decisions autonomous once again. Supposedly, during these years of decision-making and information processing, the brainwashing victim will have engaged in a lot of reflection in which a great number of her brainwashed beliefs and desires have been scrutinized. Maybe she has also gained new desires and beliefs in a critical way. If we add up all the scrutinized and newly formed mental states we would find that after ‘a number of years’ the person is just as autonomous as before. Loss of self-control with regard to mental states, the argument continues, would be a temporary phenomenon and can therefore for all practical intents and purposes be ignored.

In order to properly assess this argument we must move away from imaginary thought experiments and consider the real world effects of EC mechanisms and subliminal stimulus presentation. Because both mechanisms primarily affect implicit attitudes, for an agent to maintain self-control it should be possible – if Dennett’s objection is true – to change or ‘shed’ such attitudes. This, however, is ridden with practical difficulties. First of all, deliberative corrections that negate an attitude may change only explicit attitudes. In experiments in which stereotypes are consciously negated, implicit association strength between the attitude object and the stereotype may in fact increase rather than decrease. In one study, Gawronksi and colleagues presented subjects with faces of black people and white people and asked participants either to press the NO key whenever a black face appeared together with a negative (congruent) stereotype or a YES key when a positive (incongruent) stereotype appeared together with a black face. When implicit attitudes were measured, those in the second condition had more positive attitudes towards blacks while those in the first condition had now actually more negative implicit attitudes towards blacks! Thus, propositionally negating an implicit association appears to activate both representations in the associative store and may so reinforce its strength irrespective of the nature of the relation.

In this section I have argued that mental state formation by means of unconscious influence is possible, that the changed attitudes predictably affect behavior and that the controlee cannot feasibly correct for most of this influence. Although the techniques by which unconscious influence is wielded are not easy to use, they pose a serious prima facie threat to liberal autonomy. I will discuss to what extent unconscious control of attitudes is permissible together with similar difficult cases in section 4.

2.2.2 Processing Alien Intentions

Even if the intention of a controller to instill an evaluative judgment or to present a goal is clearly recognized, self-controlled mental state formation may still be undermined. This occurs when either upon encoding or upon retrieval the receiver of the message falsely attributes the alien goal or evaluation to herself. This would impair self-control because the controlee can in such instances no longer properly scrutinize the goal or attitude.

The psychologists Julius Kuhl and Miguel Kazén have presented evidence for precisely such a phenomenon in an interesting experiment. It simulated a work day in an office for the purpose of which subjects had to choose from a list of 27 tasks that they were to perform later, while 9 were assigned by the experimenter and the remaining 9 were left over. The hypothesis that informed this first in a series of similar experiments was that subjects with a specific personality type would be more likely to falsely recall office tasks that were assigned to them by the experimenter as self-selected. The personality type in question is state orientation, which refers (among other aspects that are not relevant here) to an impaired ability to cope with negative affect after failure on a task and a heightened tendency to rumi-
nate about such failures. By contrast, so called action oriented individuals more quickly cut their losses and focus their attention on the next action at hand. The distinction is captured by a survey that asks subjects to report their common reaction to situations. A typical question on this survey asks whether, upon repeatedly failing to contact a person by telephone, one’s reaction would be “I can’t stop thinking about it, even while I’m doing something else” (state orientation) or rather “I easily forget about it until I can see the person again” (action orientation).\footnote{104} Now, the assumption of the office experiment was that the rumination of state oriented individuals impairs an emotional preference checking system, which is used to attribute a message to a source. (The nature of this system will concern us only later on in this chapter). Under these conditions, it was furthermore assumed, state oriented persons should falsely self-ascribe (i.e. introject) office tasks that are given to them by the experimenter. A series of experiments has established that this is indeed the case.\footnote{105}

In a number of variations on the above experiment it was investigated which conditions facilitate introjections for state oriented individuals. In this regard Kuhl and colleagues reasoned that if rumination about negative emotions (i.e. state orientation) inhibits accurate self ascription of tasks, then inducing negative affect should facilitate introjection. This was indeed what they found when they induced (and then measured) negative affect by giving subjects meaningless tasks, by showing sadness arousing movie clips and by having the experimenter exert pressure on subjects.\footnote{106} One may object that this body of research is not in fact evidence for loss of self-control because the receiver might simply have liked the goal or evaluation so much that she mistook it for a mental state of herself. On this interpretation, the agent may not have been aware of the origin of a goal, but may all the same have appropriated it in some unconscious self-control bestowing way. If this were true, then one would expect that introjections would occur mainly for items that are most attractive to subjects. In order to rule out this possibility, Baumann and colleagues had subjects rate their preference for the listed office tasks in advance of the experiment. What they found was that introjections overwhelmingly occur for items that are rated low in attractiveness.\footnote{107}

Thus, there is evidence for loss of self-control in that the source of a message (including, supposedly, the intention of that source) may be confused in the course of processing (by state oriented persons), especially if they experience negative affect and perceive goals to be unattractive.

### 2.3 The Shortcut from Self-control to Autonomy

According to some philosophers, most notably Daniel Dennett and Bernard Berofsky, something like the account of self-control that I have just outlined is all there is to autonomy. But this is in fact a minority view. Most philosophers assume that something must be added to the capacities for self-control, and that this something is somehow related to the self. What Dennett and Berofsky deny is not the existence of this intuition, but the existence of psychological processes to which it could refer. The intuition, they argue, results from folk-psychology, and that is not a reliable source. For Dennett the intuition is misleading because it supposes that there would be a Central Headquarter in the brain that is both conscious and in control of higher mental functions. This is a misguided assumption, he thinks, because the feeling of being a self is the result of a fiction that functions to ascribe responsibility. It is, in his words, a center of narrative gravity that helps us understand our actions but that does not actually move anything (or can for that matter own any mental states).\footnote{108} Berofsky similarly maintains that reflective reasoning – which is supposed to speak for the self by many theorists of autonomy – may bear little relation to what is actually driving behavior.\footnote{109}
Thus, what Dennett and Berofsky deny in terms of my taxonomy is that the intuition of autonomy as a point of view is a reliable basis for a theory of autonomy. In this section I assess whether a conservative reading of the evidence can support the idea of autonomy as a point of view and claim that this is indeed not the case. However, I also argue that this assessment need not have the implications for a role of the self in a theory of autonomy that Dennett and Berofsky think it has.

For the purpose of my assessment I distinguish with Peter Carruthers between two theories about consciousness (or a point of view) and decision-making. One holds that effects of conscious thoughts on action are direct, while another approach takes them to be indirect. The first position is referred to as the self-monitoring model. It assumes that there are two, independent, systems for reading intentions: one for reading the intentions of others and one for assessing our own intentions. The self-monitoring module can on this account not only access propositional attitudes but also directly influence them, so that conscious events in the forms of intentions can directly activate a motor instruction. The other view is known as the mind-reading model and maintains that knowledge of propositional attitudes is acquired solely via the system that evolved to interpret the minds of others. On this account, the mind-reading module generates an interpretation of why we are doing what we are doing only on the basis of introceptive and extraceptive stimuli. Because it has no direct access to the practical reasoning system, conscious thought can also not directly settle the choice of a motor instruction.

With Carruthers I also take there to be three groups of relevant evidence for deciding which model is correct. First, the self-monitoring model poses two mechanisms, one for reading the minds of others and one for reading our own, so that one would expect to find cases of patients with local brain damage in which the two systems come apart. Advocates of the self-monitoring model argue for autism as a case where the mechanism for reading the minds of others is deficient while the one for reading our own intentions is intact. Closer examination of the empirical evidence does not support this position however, since a number of experiments have documented greater relative deficits in the introspection of propositional attitudes in autists. Passivity symptoms in cases of schizophrenia are claimed to represent the opposite case. Persons who have this condition are perfectly able to read the minds of others but have the feeling that the actions they perform themselves are not their own. However, the best explanation of such alienation symptoms is related to lower-order systems. When a motor instruction is given, an efferent copy is broadcast and compared with both the initial intention and perceptual data about the behavior. A mismatch results in a feeling of alienation. The best explanation for schizophrenia is that the comparator system (that carries out this procedure) is deficient. But there is no reason to assume that the comparator system involves meta-representations of propositional attitudes or even introspective experience of them. If this is correct, then cases of schizophrenia cannot count as evidence against the mind-reading model.

Second, there is evidence from research on split-brain patients and on normal subjects which shows that metacognitions about an action and the actual cause of an action can come apart. This is what one would expect to find if the mindreading model were correct. Split-brain patients are shown to confabulate stories about their actions (mostly in the left hemisphere) that are unrelated to the motivation for behavior that issued in from the right hemisphere. Thus, if the right hemisphere of a split-brain patient is exposed to a picture of snow and the left hemisphere to a picture of a chicken leg, and the patient is asked to select items, then the left hand (which is guided by the right hemisphere) is likely to pick a shovel. The obvious explanation for this choice would be that the right hemisphere associated the
shovel with snow. But that is not what split-brain patients tell the investigator. Since the interpreting is going on in the left hemisphere, which has no access to the snow picture, subjects just make up a story about the relation between the shovel and chickens. The fact that they do not simply say that they do not know why they selected the shovel suggests that it is the default mode to make sense of our actions without direct access to propositional attitudes. Similar effects can be triggered in normal subjects, as a number of experiments in social psychology shows that people falsely attribute actions to themselves if they hear a word that reasonably explains their action no more than five seconds before the action (while no other thoughts may interfere). This is suggestive evidence for the mindreading model as well.

Third, if metacognitive processes have access to practical reasoning processes and are able to intervene in them, then we would expect an innate ability to correct reasoning errors. Yet evidence from the psychology of reasoning suggests that naive subjects are particularly bad at spotting mistakes in reasoning. An alternative (although not exclusive) hypothesis would be that conscious reasoning skills evolved rather in order to engage in persuasion contests. This hypothesis has much better support, since it neatly explains phenomena such as the innate confirmation bias.

Because the aim of formulating a minimal theory of autonomy places the burden of proof on those who want to argue for more demanding notions of autonomy, I take it that the evidence for the self-monitoring model is at least not strong enough. Hence, we are to assume from here on that conscious events do not directly cause motor instructions. By corollary, the point of view from which we think we act cannot be said to have the special autonomy-conferring status that proponents of autonomy as authenticity, such as Gary Watson, Harry Frankfurt, David Velleman and Michael Bratman ascribe to it. It is important to stress though that arguing against the authority of the point of view of the agent is not the same as arguing against a role for consciousness either in action or in a theory of autonomy. That is, although the point of view of the agent as such does not confer authority on actions, this does not rule out the possibility of indirect influence of conscious actions, which may still be said to be important in a theory of autonomy. More important still, whereas the evidence for the direct involvement of conscious processes in action is weak, there is very strong and uncontested evidence for the importance of conscious processes in the formation of mental states.

To sum up, I agree with Dennett and Berofsky that theorists who think that the autonomy of actions is bestowed by the agent’s point of view are on shaky ground. However, it is one thing to claim that the point of view of the agent confers no relevant authority on actions and quite another thing to claim that there is no role for the self in a theory of autonomy. For the latter claim to be true it would have to be the case not only that the point of view of the agent is immaterial in the formation of mental states, but also that autonomy as coherence (in relation to both actions and mental states) lacks a reliable empirical correlate. In the next section I will claim that this premise is wrong.

3 Authenticity

What an argument for autonomy as coherence should look like seems clear enough. If one were able to distinguish a subset of mental states that can be said to represent the self, then
actions caused by such states can be called authentic. However, a problem with this approach is that there is a lot of disagreement among psychologists and philosophers alike about the referent of mental states that represent the self. Some psychologists think of the self as a central node in a memory network, others see it as a disparate set of modules, and yet others conceive of the self as a connectionist network. Likewise, philosophers who work on autonomy have widely diverging views on the self and see it either as a mode of information processing, a set of narratives or a set of policies. I cannot hope to resolve these fundamental disagreements and establish which one of these concepts captures the ‘real self’. I also cannot hope to develop normative accounts of authenticity for all of the central theories of the self separately (that would be too laborious). I will therefore remain neutral as to what the self refers to in the mind. The only assumption that I make is that when new states are acquired, there will be special places in the mind where states are stored on which the organism relies when it acts, or that new states are tagged as belonging to the self so that they can be retrieved when the organism acts. If there were no such mechanisms, then nothing would prevent one from acting on beliefs and desires that one ascribed to others. And such a life would lose its coherence.

On the basis of this assumption, minimal authenticity can be defined with reference to two criteria. The first is appropriate acquisition. If we could define plausible criteria for appropriate acquisition, then we could say that actions that issue in from states that were not acquired appropriately are not authentic actions. Defining such criteria will be the main task of section 3.2. The second criterion is access to relevant self-representations. If memory systems are stuffed with appropriately acquired self-related states, but reasoning systems receive hardly any input from them, then actions still do not speak for the self. From an evolutionary standpoint such differences are to be expected. When seeing eye to eye with a saber tooth tiger it would be seriously non-adaptive if the brain would take the time to extensively access memory systems in order to evaluate needs and desires. Instead, the system that regulates such scanning should block all access save the most elementary environmental input and skills that are relevant to fight-or-flight reactions. In section 3.1, I want to demonstrate that such differences in self-access are not only to be expected, but that psychology offers evidence for their occurrence as well as the conditions under which they occur.

3.1 Authentic Action
An account of authenticity as coherence with appropriately self-related states must from the outset face a variation on the objection against autonomy as a point of view. This objection holds that just as the point of view of the agent cannot be trusted as a source of authenticity, so self-related states from memory as they are acquired in psychological surveys cannot be trusted to cause action. It is possible, after all, that in reality unconscious memory states do the causal work. In response to this objection I will outline relevant retrieval mechanisms that operate both consciously (3.1.1) as well as unconsciously (3.1.2).

3.1.1 Explicit Authenticity and Action
The first important case of authentic action consists in relating a possible course of action to relevant explicit self-related states. I distinguish two ways in which such states are accessed. One concerns the mere interpretation of what we are doing, and the other involves higher-order reasoning.
A Minimal Account of Authenticity as Interpretation

In a highly original account, David Velleman maintains that autonomy consists essentially in interpreting what we are doing. Central to his theory of autonomy is the assumption that intentions are the output of theoretical reasoning that is informed by several intellectual desires. The most important of these are the desire to know what we are doing and the desire to know why we are doing what we do. According to Velleman it is only because of the motivational force of such intellectual desires that we are not limited to acting on innate drives but can also act on predictions about what we are going to do. For if such a prediction is linked to the desire to know what we are doing, it can attain motivational force and hence outcompete other desires and issue in actions. A reason for action can now be defined as a thing that enhances self-knowledge upon being intended. That is, if we act on a prediction, then what we will be doing is already known to us – it was what we expected to do after all – so that our self-knowledge increases. On the basis of this theory of action, Velleman defines autonomy as acting on predictions that are most coherent with explicit self-representations.

To this end he introduces, in a discussion of Dennett's views on the narrative self, a “narrative-composing module”, which “will … declare a winner [among potential courses of action] simply by telling the more coherent continuation of the story.” And the most coherent continuation of the story is the one that is best in line with our beliefs and desires.

There are a number of problems with this account. First of all, there are serious difficulties with Velleman’s concepts of intention and a reason for action, which has been pointed out by Michael Bratman. When we act, we not only expect a goal to occur but also the side-effects of that goal. Yet if expectations are intentions by virtue of increasing our self-knowledge when they are performed, then side-effects that we clearly do not intend (say, waking up a person by switching the light) would have to be described as intentions as well. Similarly, Bratman argues, the expectation that I will stumble in front of an audience would increase self-knowledge and hence have to count as a reason for action. Both consequences are unacceptable. The only way to resolve this problem would be to restrict the expectations that can settle a decision to those that are based on an agent’s relevant preferences. But such relevant preferences would have to be practical desires for success in action, so that the intellectual desires for self-knowledge and understanding are no longer the drivers of behavior.

Secondly, Velleman posits a desire to know what one is doing that would lead one to do what one is knowing. There can be little doubt in any conscious person that we have a desire to know what we are doing. The issue, however, is whether this desire causes us to act upon what we are expecting to do. Velleman thinks that evidence from social psychology is at least coherent with his view, and, supposedly, that this would strengthen his case. The psychologists Wegner and Vallacher have indeed argued that we select a ‘proponent act description’ before we act, and that this selection is guided by a concern for acting consistent with beliefs and desires that make sense of the action. And decades of research on cognitive dissonance have indeed shown that people go to great lengths to maintain consistency. But none of this research has any bearing on the question of whether we select such act descriptions – or expectations – because we want to know what we are doing. Instead, all the evidence that Velleman has mustered rather points in the direction of a desire to act in way that is consistent with act descriptions. For the sake of parsimony, we should therefore pose – for the regulation of action at least – a consistency desire rather than an intellectual desire to know what we are doing (or the desire to know why we are doing what we do).
On the basis of these criticisms I want to propose a more minimal version of a theory of autonomy as interpretation that retains, I think, the spirit of Velleman’s account. It takes as its starting point the assumption that the mindreading module is primarily involved in interpreting the actions that are in fact selected by the system 1 practical reasoning modules. The results of these interpretations can in turn indirectly influence the practical reasoning process if the selection of future action schemas is regulated, at least in part, by the desire to act consistently with self-interpretations. Moreover, this construction allows me to say that acting upon these interpretations results in autonomy to the degree that these interpretations (i.e. act descriptions) are formed on the basis of (appropriately acquired) self-related states. That is, when more states that are made proponent in action identification enter into the consistency assessment process, then actions that issue in due to a desire for consistency (that is properly related to the consistency assessment process) are more autonomous. Take the case of pressing a doorbell. This single action can be interpreted as, respectively, stretching a finger to reach for the button, ringing the bell, or looking if a friend is at home. It is clear that the highest level interpretation touches on far more concerns that the lower level ones. Looking if a friend is at home raises many more questions with regard to social norms (Is this an appropriate time?), means-end efficiency (Should I not rather give him a call?) and ethics (Is it ethical to visit a convicted pedophile?). Hence, many more self-related states would be activated than the simple interpretation under which we stretch a finger to reach for a button. Thus, the higher the level of identification, the more likely it is that self-related memories will enter into the consistency assessment process.

It seems to follow then that a minimal theory of autonomy can simply set a certain level of act identification given classes of situations. Unfortunately, personality differences make such a move problematic. To see why, we must go into some more details of Vallacher and Wegner’s action identification theory. It states on the one hand that if higher and lower level act identities are available, there is a tendency for the higher level act to become proponent. Evidence for this principle is found in the readiness by which people accept a higher order act identity that is presented to them when they are identifying at a low level. On the other hand, when act identities cannot be maintained at a certain level (because the act cannot be carried out effectively) the tendency is to move down the hierarchy, where actions are broken down into their component parts. On the basis of this principle Vallacher and Wegner predict that people who are less skilled at carrying out a range of tasks (so called “chronic klutzes”) should in general maintain lower act identities. When the researchers compared responses on a questionnaire that asked subjects to self-report success at a range of tasks (such as typing, playing an instrument, planning a party and building furniture) with responses to questions that asked subjects to indicate the level at which they would identify an action (e.g. whether they would describe ‘making a list’ as ‘getting organized’ or as ‘writing things down’), they found a significant correlation in support of this prediction. That is, people who report more difficulties in carrying out actions effectively are more likely to display the tendency to identify actions at a lower level. But that would mean that personality differences influence people’s propensity to be autonomous, and that makes formulating a minimal threshold a tricky affair. Dealing with this practical problem in the context of liberal autonomy will be the task of section 4.

A Minimal Account of Authenticity as Reflection

Distinct from interpreting what we are doing (which may be the work of a specialized mind-reading module) is the activity of reflecting about what we are to do. All hierarchical theorists of autonomy assume that a special relation with self-related states is established in
reflection, but John Christman is the only one who has developed this view in significant
detail. He discards the idea of self-governance as identification, since it means either that we
must “identify with all aspects of ourselves that we acknowledge as such”, which would force
us to acknowledge addictions, compulsions and the like. Or it means that we must identify
with “those principles, values, and ideals that we hold in most esteem” so that we are led by
our “best selves” – but then we would have to view everyone who falls short of her ‘ideal self’
as deficient in autonomy, which would include all of us.131

In an attempt to strike a balance between these extremes, Christman requires only that
upon reflection on a ‘basic organizing value’ – but we could also insert here ‘course of
action’ – we do not experience alienation (i.e. we must both belief and feel that we should
not repudiate the value or course of action).132 Obviously, we need some guarantee for thinking
that this reflective judgment is properly related to the self. To this end, Christman
demands that the value (or course of action) be processed by “the characteristic mode of
thinking, embodiment and feeling that the agent’s self-schema manifests”. This ‘characteris-
tic mode’ refers in turn to the schemas, heuristics and scripts that are activated “over a variety
of circumstances”. Moreover, from among these stable processing modes only the ones that
are activated when we reflect (where this reflection is unimpaired by manipulation, pathol-
gy and the like) speak for what we could call the autonomous self.133

There is more to Christman’s theory of autonomy, but what concerns me here is the
extent to which reflection can count as authentic on his account. In this regard I am skepti-
cal, for three reasons. First, it is not clear in how many circumstances reflection is to occur,
or even by which criteria the appropriate number and type of circumstances could be
selected. If this is correct, then it seems more promising to identify psychological mechan-
isms by which self-consistency is assessed and to demand that these are not impaired.
Second, self-schemas are unique individual modes of processing information, but if they are
to guarantee autonomy then they must take the form of reflection on Christman’s account.
Hence, if there are individual modes of processing that are not reflective, then there can be
a conflict between a state of alienation triggered by characteristic processing and a state of
non-alienation triggered by reflection. If that is true, then the requirement of reflection
seems to be biased against certain ‘characteristic’ processing modes. This bias is problematic
for at least a liberal theory of autonomy because, as Berofsky notes, it is one thing to claim
that autonomy involves some reflection and quite another to claim that only reflective judg-
ments speak for the self.134 While the former commitment is reasonable from the perspective
of a minimal liberal theory of autonomy (see section 2.1), the latter is not.135 Third, a liberal
theory of autonomy should not decide which types of feelings or beliefs (such as non-alien-
ation) should function as the standard of evaluation in reflecting on self-related states. For
such a theory it suffices that self-related states are appropriately acquired and that reasoning
systems have appropriate access to them. In the latter respect, a minimal theory must ask
instead how access of self-related states to system 2 processing is regulated.

On this question we must, absent relevant work in psychology, rely on some tentative
findings from neuroscience. The short summary of studies on this issue is that reflection that
involves a judgment on preferences of the self (as opposed to reflection on an abstract prob-
lem) appears to go together with the activation of a specific area in the brain that is known
as the Cortical Midline Structures (CMS).136 Activation of these structures during self-re-
fection is moreover shown to reduce conflict in cases where there are multiple correct pos-
sibilities, such as in occupational choice.137 That is, system 2 conscious reasoning about
self-relevant dilemmas does seem to resolve personal dilemmas, and it may well do so by
relying on self-related states (given activation patterns). What is more, the impairment of
these self-reflective processes due to cognitively demanding tasks is associated with less activity in CMS and more activity in lateral prefrontal areas of the brain (while the reverse also holds true). Thus, the above evidence suggests that self-related processing by system 2 may be blocked by cognitive load, and is in that sense similar to the interpretation system (where cognitive load decreases task proficiency and downshifts act identification).

3.1.2 Implicit Authenticity and Action

The case for implicit authentication of action in a theory of autonomy can be inferred from the work of Harry Frankfurt. In section 2 I discussed Frankfurt's theory of autonomy as acting on second-order desires that are decisively endorsed. Several critics have noted that although it makes intuitive sense to maintain that autonomy is dependent on some form of reflection, it is not intuitively clear why reflection by itself would speak for the agent. The later Frankfurt has responded to this question by claiming that reflection does so only when it is in line with our ‘essential volitional essences’ that consist of mental states of deep caring. Frankfurt hereby abandons the active nature of the decisive endorsement of a higher-order desire and stresses instead that becoming ‘wholehearted’ – as it is now called – is not “under … immediate voluntary control”. It is not a matter of merely “telling stories about our life”, since “we can be only what nature and life make us”. Instead, having a free will is to “be wholehearted in [the will]”, that is, to have no volitional conflict. Wholeheartedness does not mean that there can be no “inner opposition to [one’s] will”, but it does mean that one “must be resolutely on the side of one of the forces struggling within” and “not on the side of any other”. Being satisfied is defined as the “absence of restlessness or resistance”, or “having no ambition for improvement” of the current situation. Thus, satisfaction consists not in the adoption of any “cognitive, attitudinal, affective, or intentional stance”, but is simply a “state of the entire psychic system”.

Although there is much to criticize in this account, I also want to maintain that the core intuition of autonomy as acting in harmony with mental states that are not under volitional control is defensible. With regard to the critique, Frankfurt does not explain what the phrase ‘volitional essence’ refers to. He also fails to make clear what entity could be on a ‘side’ in the case of volitional conflict, while it is likewise not clear why the amount of conflict would determine the amount of autonomy. There is at least a strong intuition against autonomy as harmony. But these critical points do not invalidate the impetus of Frankfurt’s approach, because they are for the most part related to Frankfurt’s conception of the self as a perspective. If we think of the self instead as a set of representations in the mind, we have no need to introduce ‘volitional essences’, much less an entity that takes sides between them. Instead, we need only require of an action that it issues in from a practical reasoning process that has appropriate access to implicit self-representations. This alternative reading of the implicit coherence intuition is contingent on the existence of a relevant retrieval mechanism. To this end I will outline in broad lines the theory of volition developed by Julius Kuhl, known as Personality Systems Interaction Theory (PSI), which poses precisely such a mechanism and presents evidence for its existence.

A Psychological Account of Implicit Self-consistency Processing

In PSI volition is seen as the output of a complex system that consists of two main processes, self-control and self-maintenance, which are regulated on the basis of positive and negative affect. Self-control is concerned with inhibiting impulsive actions and maintaining a
single focus on goals. It is conscious and effortful and consists of two component systems that are called intention memory and object-recognition. Intention memory is concerned with maintaining abstract symbolic representations of intentions in working memory and inhibits intuitive behavior in order to prevent premature behavior. The object-recognition system is a lower-order system concerned with the detection of discrepancies and warns others systems in the brain in case of their occurrence. The other major system, called self-maintenance, is concerned not with the suppression of alternatives but rather with the integration of a wide array of desires and implicit self-representations. Its two component systems are extension memory and intuitive behavior control. Extension memory integrates cognitive and affective information from various subsystems and processes it much faster than analytical thinking due to parallel processing. Intuitive behavior control translates intentions into routine actions and also regulates spatial awareness. Volitional behavior is supposed to result from carefully calibrated switching between these two modes and to be regulated by negative and positive affect. An increase of negative affect is a sign of danger and activates the object recognition system, while a decrease in negative affect indicates safety and activates extension memory. An increase of positive affect is believed to signal the absence of problems so that intuitive behavior is activated, while a decrease in positive affect supposedly demonstrates that needs are not met so that intention memory is activated.

If there were in fact such a self-maintenance system, then this would be the type of implicit consistency checking mechanism that can support a place for authenticity in a minimal theory of autonomy. It is therefore most interesting that Kuhl and colleagues have deduced specific predictions from PSI for the experimental setting of the office work day simulation described in section 2.2 and put them to the test. Note that if there is a system that automatically assesses the coherence of a course of action with states of the self then it may do so at the encoding stage (when office tasks are allocated) or at the retrieval stage (when the memory task is completed). Only the assessment of coherence at the retrieval stage might be evidence of the type of mechanism that we are looking for.

In line with this possibility, Kuhl and colleagues hypothesized that because individuals who are worse at automatically downregulating negative affect (i.e. state oriented individuals) are more likely to introject tasks given to them by others, the retrieval of task selection may involve an affective self-consistency checking mechanism. Specifically, they argue that there may be a check of (emotional) preferences about the goal or attitude object (which is generated on line) which may result in a commitment marker being tagged to that goal. If this commitment marker is scanned when a goal is retrieved, then subjects who are best at retrieving task selection (i.e. action oriented individuals) should show the longest response latencies (since scanning takes time). The assumption was tested in a number of studies which found that upon retrieving task origin there is indeed a longer time lag for action oriented individuals (compared to state oriented individuals), especially under conditions in which negative affect is induced (i.e. meaningless tasks, pressure and sad mood induction). As we noted earlier, negative affect induction before task retrieval increases the introjections rate (for state oriented individuals), so that there is good reason to assume that at least part of the introjection indeed occurs at the retrieval stage and that the time lag is related to this effect.

Let me stress that I leave open against what exactly consistency is checked by this automatic process. These entities may correspond to what is found on implicit attitude measures, because the process operates so fast. But it cannot be ruled out that the attitude that corresponds to what is found on explicit reports are also accessed. Moreover, how these mental states are instantiated in the brain is also not clear. This structure may consist of subsym-
bolical connectionist networks (which is what PSI assumes) or symbolical modules, but this issue is highly contested. What is to be retained for now is that there is a process of automatic and unconscious preference scanning for goals and attitude objects that aids the recall of selected actions and may reasonably be supposed to be activated in action contexts. By virtue of having a role in selecting an action we say that the entities that are accessed by this process can be called mental states of the self. The greater the access to such states, the more authentic a person becomes.

Once it is agreed that both implicit and explicit retrieval is relevant for authenticity it may be wondered which one speaks for the ‘real self’. A liberal theory of autonomy, I argue, should not take sides on this issue. Intuition of course tells us that we should side with explicit retrieval, but if Kuhl and colleagues are correct, explicit retrieval may activate a smaller number of self-relevant states than implicit retrieval. Alternatively, one could argue that harmony of the two retrieval systems leads to more autonomy. There is currently no research that speaks to this. But if explicit retrieval were to rely more on explicit attitudes and implicit retrieval more on implicit attitudes, then there would be relevant evidence that indicates that a state of self-awareness can reduce the gap between explicit and implicit states. However, it is not realistic to require citizens to engage continuously in meditation practices and the like that foster such harmony.

Hence, a minimal liberal theory of autonomy should only demand that access to the conscious self is unimpaired by cognitive load and it should demand that implicit processes are unimpaired by negative affect. Because of this liberal stance, the theory is able to ascribe autonomy to the entire spectrum ranging from over-reflective and inhibited personalities to impulsive hedonists. And this is precisely what liberal autonomy should look like.

3.2 Authentic Mental States

For mental states to count as authentic it will not do if beliefs and desires are simply stacked on to each other. If we think of the memory system upon which implicit or explicit retrieval rely as a box, then if that box simply contains all the beliefs and desires the system happened to encounter in its environment the set of mental states of a person will not count as autonomous. And we would say so not merely because the resulting incoherence would disturb efficient action execution (and thus diminish self-control), but also because newly acquired beliefs and desires would not be related to existing mental states in a way that preserves the authenticity of the self. The task of this section is to give the most convincing minimalist account of criteria for such mental state acquisition without specifying what such a self should look like.

The most straightforward proposal for a principle of authentic mental state formation is to require that all mental states be processed by bringing to bear on them the mechanisms for interpretation and reflection that I outlined in section 3.1. Although such a proposal might satisfy the demands of the intuition of autonomy as coherence or a point of view, it does not satisfy the demand for realism, because such processing (at least in the case of reflection) would be way too resource intensive. This concern has prompted John Christman to propose a modification. If we cannot realistically be required to reflectively endorse new mental states, then maybe we can require that we endorse new mental states were we to reflect on them in light of how we acquired them. However, the problem with defining autonomy on the basis of hypothetical reflection is, as I noted earlier, that there is no practi-
cal way in which others can establish whether a person would have reflectively endorsed a mental state.  

A second solution to the problem that reflection on new mental states is too resource intensive has been developed by Alfred Mele in a discussion of the authenticity of values. He specifies the conditions under which reflection is to occur, where the reflection bestows autonomy by means of being endorsed by the point of view of the agent. Specifically, Mele identifies two principles according to which this process is to be structured. Initiatory principles stipulate when assessment should issue in and procedural principles stipulate how assessments should proceed. Mele mentions a number of such principles for the purpose of illustration, but maintains that they should not be fixed by a theory of autonomy. Instead, autonomous agents should design and implement these principles themselves. Although this radical subjectivism about standards of mental state acquisition may seem attractive from a liberal point of view, it cannot be the basis for a minimal liberal theory of autonomy. For one thing, it is not realistic to suppose that mechanisms that initiate reflection are under conscious control. Such a mechanism would have to filter huge amounts of data in a way that is carefully calibrated with other unconscious heuristic principles. Such complicated design work appears to be beyond the reach of conscious decision-making. For another thing, because there would on Mele’s account be no straightforward way for the state to establish which initiatory or procedural principles an agent upholds, this construction would – as in the case of hypothetical reflection – be practically meaningless. That is, the liberal state cannot feasibly keep track of individually defined initiatory and procedural principles and thus could not protect autonomous mental state formation.

In the remainder of this section I want to present an alternative account which modifies Mele’s framework by replacing the intuition that autonomy is a point of view with the intuition that autonomy involves coherence. This move allows me to find objective criteria of authenticity violation. That is, in order for initiatory mechanisms to support a theory of autonomy as a point of view, they must be endorsed by the agent in reflection – hence Mele’s untenable subjectivism about initiatory mechanisms. But once we take coherence as the guiding intuition for designing a theory of autonomy, we have an externally verifiable criterion for identifying initiatory mechanisms. I will here outline two such coherence-checking initiatory mechanisms, while taking a liberal (and hence minimal) approach to the types of assessment that guarantee authenticity. The first type of consistency-checking mechanism is domain specific. In this regard research has so far focused on a mechanism that is concerned with maintaining a balance with regard to the valence of persons. Although this mechanism issues in automatic domain specific processing I want to argue that this does not undermine authenticity, since it remains restricted to lower-order processes. The second mechanism is general and concerned with bringing a host of self-relevant mental states to bear on new mental states. If it is plausible to claim that these mechanisms form the main portion of the initiatory defense structure, then we can say that authentic mental states are the ones that have passed through these initiatory processing channels.

The operation of the domain specific mechanism is studied in balance theory. This theory holds that relations between people tend towards a balance in which we like the friends of our friends, dislike the enemies of our friends, and so on. In triad relations – i.e. relations between three people – there is balance if there are either no negative relations or an even number of negative relations. If such balance is established at encoding, then this would be a case of a mechanism that detects inconsistency and initiates an automatic corrective response. In a recent experiment it has been shown that this is indeed the case. If subjects were given a negative valence of a source first (e.g. Peter is not nice) and afterwards
his negative evaluation of a target (e.g. Peter dislikes Paul), then the subject’s attitudes towards the target balanced out and was positive (e.g. I like Paul – the enemy of my enemy). However, if the valence of the source (i.e. Peter) was given only after his negative evaluation of Paul, no balancing took place (i.e. the negations did not cancel each other out). Likewise, in another experiment it was shown that if the valence of the source is changed afterwards (e.g. Peter is in fact not nice at all) the balance that ought to occur in relation to his negative evaluation of the target (i.e. the two negations should cancel each other out) fails to materialize. If balance were established upon retrieval, then the order of presentation should not have an influence. As these experiments show that the order of presentation does have an influence, they tell us that there is a mechanism that automatically establishes a specific type of consistency in our evaluations of persons upon encoding. However, because balance processing aims to maintain consistency, which is dependent on the ascription of truth value, associative processing should not be subject to it, if the APE model is correct.

The second initiatory mechanism that can be argued to meet minimal standards of authentic mental state acquisition is the unconscious consistency checking process that is studied in PSI research. In section 3.1.2, I offered evidence for unconscious consistency checking at the retrieval stage, but there is also good evidence to suggest that it occurs upon encoding. According to PSI what should happen at the encoding stage is that emotional preferences towards a goal or attitude object are scanned, after which it may be tagged with a commitment marker. On the basis of earlier research, Kuhl and colleagues assumed that many of the presumed aspects of the emotional preference scanning process are located in the right hemisphere. For example, whereas the recognition of faces of other people correlates with more activity in the left hemisphere, the recognition of one’s own face correlates with more activity in the right hemisphere. The right hemisphere is also known to be more active when integrative processing is observed and when increased activity of bodily states is registered. In order to manipulate the activation of the right and left hemispheres, Kuhl and colleagues relied on research that indicates that muscle contractions on one side of the body activate the opposite brain hemisphere and its functions (the hemispheres are connected to opposite sides of the body for reasons that are not well understood). Thus, squeezing a ball with your left hand should activate your right hemisphere, and if that hemisphere is indeed more involved in self-related, integrative and emotional processing, then the number of introjections for state oriented people should decrease.

Interestingly, the ball squeezing manipulation allows one to accurately isolate encoding effects from retrieval effects. As such, experiments using the office set-up that was described earlier demonstrate that one minute of left-hand ball squeezing before encoding (i.e. selecting and being assigned office tasks) by state oriented individuals totally annihilated the surplus of false self ascriptions. Moreover, both state and action oriented individuals showed a greater tendency to introject after right hand ball squeezing (and concomitant left hemisphere activation). The result has been replicated for attitude formation (where the attitude objects were tests for measuring everyday intelligence, which subjects rated for their effectiveness). This consistency-checking mechanism can be said to be initiatory to the extent that there is suggestive evidence for thinking that it sends a signal to reflective processing. This evidence builds on an intermediate stage in the argument within PSI theory on introjection, which is that performance on tasks in which subjects are to judge the coherence of triplets (e.g. ocean, sand and bucket) is impaired in state oriented individuals after negative affect induction. Interestingly, the psychologists Sascha Topolinski and Fritz Strack suggest how
this effect may initiate relevant reasoning. Specifically, they propose that the processing of coherent triplets occurs more fluently and that this fluency arouses positive affect. This positive affect may in turn be read by a monitoring system and automatically influence or even become the object of system 2 reasoning. In a series of studies, Topolinski and Strack have made a cumulative case for this proposal by showing that coherent triplets are processed faster, that they function as primes that facilitate positive affect and that they are liked more (compared to incoherent ones). Thus, if integrative coherence judgments do indeed account for the relevant consistency checking, then positive affect may bias reasoning about self-consistency without giving any insight into the nuts and bolts of the process. On this account, we are conscious only of a positive feeling when a goal is found to be coherent with our mainly emotional preferences. When a negative feeling occurs, by contrast, this could conceivably initiate central processing. But research is needed to corroborate these speculations about the initiatory function of this mechanism.\(^{164}\)

In any case, it is unlikely – as it is unlikely for automatic balance regulation – that associatively processed stimuli will be subjected to this coherence-checking mechanism. That is, the mechanism has been shown to operate in the case of goals or attitude objects that were deliberatively chosen, but it is unlikely that it would kick into action when two stimuli are implicitly associated with each other, let alone when affect is misattributed from a US to a CS. Admittedly, there is research that might be interpreted to suggest that conflict detection leads to appropriate changes of implicit attitudes indirectly. That is, a number of experiments have shown that when subjects have discrepant explicit and implicit attitudes about the same attitude object, they are more willing to engage in deliberative processing about information that is relevant to the attitude object (and that might resolve the discrepancy).\(^{165}\) In order for this effect to occur there must presumably be a separate system that receives input from the propositional and associative processing systems about an attitude object and checks whether they are consistent. This is of course not the same as evidence of a consistency checking mechanism for implicit-implicit conflict. But it could be argued that when discrepancy between implicit and explicit measures is reduced, and explicit measures are subject to consistency detection, implicit attitudes are by virtue of being aligned with explicit measures indirectly responsive to conflict detection. Such an argument is unpersuasive though, because processing information that resolves the implicit-explicit attitude discrepancy by changing implicit attitudes often widens the discrepancy, because – as we saw in section 2.2 – negations often do not carry over to implicit attitudes.\(^{166}\)

If this is correct, then associatively processed stimuli bypass both types of initiatory mechanisms. If there is no plausible way in which initiatory standards can be lowered, then the possibilities of finding natural defense mechanisms against authenticity undermining influence have been exhausted. Because these assumptions seem to be correct, we are forced – as in the case of self-controlled action and mental state formation – to look for thresholds beyond natural defenses.

### 4 Practical Thresholds of Liberal Autonomy

In detailing the constituent components of minimal liberal autonomy I have come across a number of unsolved puzzles that concern mainly covert influence on both action and mental states. Specifically, with regard to self-control it is to be determined which extent of
manipulation is acceptable and with regard to authenticity which minimal levels of cognitive load, negative affect and associative processing are acceptable. In this final section I propose practical standards that can bring my minimal theory of liberal autonomy into equilibrium with intuitions about appropriate state action. I have of course already discussed some practical considerations, but I have done so only insofar as they were germane to the plausibility of mechanism that are constitutive of autonomy. Here I consider solutions to problems that persist after the possibilities for modifying such mechanisms have been exhausted. And I will do so for both self-control (4.1) and authenticity (4.2).

4.1 Thresholds of Self-Control

Whether liberal state action is imperative when self-control is undermined is – I propose – contingent on two factors. First of all, state action must be proportional to the violation of self-control (just as it must be proportional to a violation of freedom). In other words, the violation of self-control must be sufficiently central to the controlee before state action is imperative. Second, state efforts to protect autonomy should not bring with them violations of autonomy in another respect (or rather, the latter violations of autonomy should not outweigh the gains won by protecting autonomy). Specifically, efforts to protect autonomy can impair autonomy as a side-effect in the form of both violations of privacy and violations of the controlee’s consent.

Lack of Centrality

To begin with, not all instances of manipulation seem serious enough to warrant state action. Consider the following cases. Suppose I know about the process of mirroring (either from theory or from experience) and just to amuse myself I shake my foot with the goal of making you shake your foot. I hereby tap into an unconscious mechanism by which you regulate your body movements in interaction, and so bypass your ability to know about my intention. Hence, my action would result in a clear loss of self-control on your part. Or suppose that I direct a thriller movie in which I want to create an experience of suspense whenever a certain evil character appears. In order to do so, I play a piece of music that I know arouses an eerie feeling every time the evil character enters the scene. From experience (or from seminars at the film academy) I know that the viewer will make an associative link between eeriness and the evil character (even if I know nothing about the details of evaluative conditioning). If the editing is subtle and the story an enthralling one, I will achieve my goal even with an experienced public.

Now, both in the foot shaking and the movie example autonomy conditions are violated intentionally and yet it seems deeply wrong to say that liberal autonomy is violated in any meaningful way. In these cases the obvious source of this intuition pump seems to be the relative futility of the behavior and attitudes. In other words, it seems so utterly immaterial to a person’s life whether she shakes her foot or thinks that a particular movie character is eerie that it is unreasonable to suggest that the liberal state should take action to protect her. By analogy, although breaking your neck in the course of a fistfight will land me in jail, pushing you off the sidewalk will not. Although the latter act is wrong in some sense, it is just too marginal an offence to require state action. And just as there is no clear cut-off point at which violence becomes an offence, so there is unlikely to be a point where intentional autonomy violation becomes serious enough to become a concern for the liberal state. Still, if this analogy were to be useful, there would have to be a standard of centrality of self-control loss equivalent to bodily harm. I propose that such centrality of self-control loss pertains to
both mental states and behavior of the controlee, where the centrality of mental states is measured against the attitude’s position in the controlee’s goal hierarchy and the centrality of behavior is dependent on its social and private consequences.

I have no intention of delving into debates in psychology on the details of goal hierarchies. Instead, I will simply assume that something like the goal hierarchy of three stages developed by the psychologists Charles Carver and Michael Scheier will be true. At the most basic level of this model there are sequences (or motor control goals), one level up are programs which are described as ‘do goals’ (e.g. going to the movies), and still higher up are principles and traits (such as ‘be thoughtful’) which are connected in memory to the self-concept.168 With regard to mental states this simple model allows us to say that – all things being equal – changing mental states that involve ‘do goals’ is of greater consequence than changing mental states that involve motor instructions. Hence, whether we think of a movie character as eerie or not can be said to be of little significance because its position relative to the goals that structure our behavior will in the typical case be peripheral.

Manipulations of behavior are more central if they have a greater effect on future actions by the agent. This effect can be mediated either by the effect of the behavior on the social environment or by the effect on the psychological organization of the agent herself. In the first case the agent makes a commitment which changes the perception of the agent in the mind of others. For example, if the manipulated act were a marriage, then the belief held by others that the manipulated agent is married changes their future conduct towards her, and hence has a substantive effect on her future actions (depending on the relationship). Such an event might also involve, and this is the second case, a change in self-perception that will have a major impact on the content of the agent's goal hierarchy and so influence future actions (where again the level in the hierarchy determines the centrality). Hence, all things being equal, the manipulation of actions that are public will be more central than the manipulation of private actions and actions that influence goals higher in the goal hierarchy level are more central than actions that influence goals at the bottom.

Irretrievability
The present proposal still allows for many cases that do not even look like threats to liberal autonomy as we intuitively understand it. Suppose that after a series of painful losses I finally want to defeat you in a game of trivial pursuit. Knowing my share of psychology, upon opening the box I start a discussion about soccer hooliganism and in the course of the ensuing discussion ask you – by way of a rhetorical gesture – to imagine being a hooligan in order to understand their side of the story. By activating in you schemes associated with low intelligence (which results from empathizing with soccer hooligans) I cause you to underperform during the game and so I achieve my premeditated goal of defeating you.169 Now, on the centrality condition I meet the criteria for relevant liberal autonomy violation if we add that the consequences of your loss will be severe. Suppose that in light of your previous string of victories against me, you overconfidently bet your house on winning the contest. Although the effects of the covert influence are now central to the controlee, it seems wrong to suggest that the liberal state therefore has an obligation to step in and prevent me from engaging in this petty trap of mine.

This intuition, I maintain, is fueled by the principle that state action should not excessively trample upon privacy. Monitoring or reconstructing actions such as the trivial pursuit case would surely constitute a massive invasion of privacy. And if it is true that we value privacy because we value having control over personal information, then the injunction against privacy violation is in fact a concern for autonomy itself.170 Does this mean that the
liberal state can never act against manipulation? Not if one can show that there are realistic ways in which manipulative intentions and knowledge can be retrieved without violating privacy. I want to argue that this condition is met when the intentions of potential manipulators are fixed by an organizational setting.

Specifically, if agents are members of organizations, then there are two ways – one direct and the other indirect – in which intentions and knowledge can be retrieved appropriately. Direct retrieval is possible because overall organizational strategies and the roles of its members in relation to that strategy are often accessible to observers. Members of the organization may have given interviews (to the press or researchers), strategy documents may be available and it is even possible that surveys have been conducted. Indirect retrieval is possible by deducing intentions and knowledge from statements or patterns of behavior. If no literal statements of intent or knowledge are available, then it may be possible to at least deduce from documents and interviews whether agents must have had specific intentions or had sufficient knowledge about the manipulative nature of their actions. Consistent patterns of behavior may also be good evidence of intentional behavior as well as, more importantly, specific types of knowledge that tasks require for their effective performance.

What is more, a concern for privacy also brings with it that the main objective of state measures against manipulation should not be to act against individuals (since this involves privacy violating monitoring), but to design policies that regulate or ban organizational practices that violate autonomy. That is, it must be shown that we can reasonably expect individuals in organizations to have manipulative intentions and knowledge in order to determine whether state policies against their organizations are in order.

Informed consent

Another case which suggests that our current conception of self-control is too demanding is the therapeutic use of hypnosis. Once under hypnosis, a client in such a session cedes self-control because she can no longer properly take either context-relevant or longer-term goals into account. And yet intuition unambiguously informs us that such influence is acceptable from the point of view of the liberal state. One explanation for this intuition refers to the well-being of the addressee. On this reading, the hypnosis is legitimate because it is in the interest of the client. But this explanation is insufficient, because therapeutic hypnosis cannot legitimately be applied at any one time (except for cases in which a subject has lost all contact with reality). Given this conditionality of our intuition, what seems required for hypnotic therapy (for mild clinical cases at least) to be legitimate from the point of view of the liberal state is proper consent of the patient. And because proper consent is an exercise of autonomy, protecting a citizen’s autonomy against a practice to which she has consented constitutes yet another case in which we are faced with a trade-off between (respect for) different aspects of autonomy.171

It is contested which properties consent must have in order for it to condone violations of the autonomy of the consentor. Surely, from the perspective of the liberal state there must be credible guarantees that a person who has consented to thorough manipulation has arrived at that decision without undue pressure or coercion. As a first step towards specifying these properties we could make a rough distinction between cases in which tacit consent is acceptable and ones in which explicit consent is required. We would also want to distinguish between the types of contexts in which consent to autonomy violation might be required. That is, consent can be given at the level of a message, at the level of the (institutional) context in which types of messages can be expected and society at large (with its legal framework).
I posit that there is an interaction between the centrality of covert influence and the appropriateness of either tacit or explicit consent. As I conceive of it, instances of consent to autonomy violating practices cluster in three groups according to their centrality. Firstly, futile cases of covert influence do not violate liberal autonomy by virtue of their lack of centrality. This includes cases such as unconsciously priming a person for games of trivial pursuit and the like, where other agents are completely in the dark about the covert influence but for which we still do not think that the liberal state has any role to play. Such cases are found predominantly at the level of individual messages. Secondly, there is a group of cases of covert influence that because of being somewhat more central requires tacit consent. Such cases would include movies, soaps and stage hypnosis. We can assume that literate viewers know that their emotions are aroused outside of awareness in such cases and hence that by entering a cinema or switching to a channel they have tacitly consented to such influence. In such cases the agent typically does not know the content of the future influence, its centrality nor the method by which it is wielded. In fact, such knowledge might annul much of the joy derived from the experience (e.g. of watching a movie). Besides these cases in which tacit consent is given at the level of (institutional) contexts (in which specific message types can be expected), one can also imagine tacit consent being given at the level of society.

One may argue that tacit consent to covert influence suffices in a liberal society that treats its citizens as responsible agents. But I think this argument is wrong, because of the danger of runaway manipulation. That is, tacit consent is a shady concept, and once the risk of abuse becomes too large it becomes implausible to say that tacit consent is a sufficient safeguard. Aggressive cults for example can always argue that even the most extreme long-term priming and hypnosis efforts could have been anticipated at some level by participants. In others words, if tacit consent would be the only type of consent required, then the risk of abuse becomes too serious. If this is correct, then we can distinguish a third group of cases that are highly central and therefore require explicit consent. Although I cannot draw a hard line in this regard, a central case would be religious rituals that use methods of implicit belief induction (e.g. bodily synchrony, music and evaluative conditioning). With regard to explicit consent as well, criteria can be more or less stringent depending on what the controlee is supposed to know. That is, a range of positions can be occupied ranging from thinking that highly central covert influence requires that the full swat of conditions is consented to (source, content, method, strength of effect), to thinking that carte blanche consent suffices (i.e. merely knowing that someone will somehow covertly change attitudes in a given context).

Let me illustrate the ambiguity of intuition in this regard on the basis of an episode of the Channel 4 program ‘Hero at 30.000 feet’ by the psychological illusionist Derren Brown. In this program, Brown had hundreds of volunteers screened for one of his new mental stunts. As in previous programs, these volunteers presumably all signed a contract (in previous shows this happened on screen, where Brown would tell the participant that ‘in the contract it says that we can do anything we like with you’). One of the candidates – Matt – was then followed for a month with hidden cameras without him knowing that he had been selected (nor did he know that his entire family was in cahoots with Brown). The objective of Brown was to select a candidate with a boring life that lacked meaning, and to inject purpose into it by means of unconscious influence (especially hypnosis). Thus, Matt was hypnotized in his sleep by Brown at night and instructed to take risks in his life, after which he was planted into dangerous situations as a form of (involuntary) behavioral fear-reduction therapy.
Now, here we have a case in which a person gives carte blanche consent to serious intrusions of his mental independence (at the end of the show we see Matt opting to buy a new house and looking for a new job, presumably because of him having gone through a number of terrifying experiences in hypnosis which included being tied to a railway when a train was approaching, taming a crocodile and landing a plane with dozens of passengers). I think that it is fair to say that it is not clear whether the liberal state should allow for such interventions, since there are intuitions on both sides. In order to account for the ambiguity in this and similar cases I propose to distinguish between what I will call the present-autonomy account and the balanced-autonomy account.

On the present-autonomy interpretation, carte blanche consent to central autonomy violation in branding will do for the liberal state. It requires no guarantees from consenters as a matter of liberal principle (except for absence of coercion and pressure) because the ultimate decision-making authority resides with the current instantiation of the agent. On the balanced-autonomy interpretation, by contrast, carte blanche consent to central autonomy violation will not do. From this perspective the default position of the liberal state is to protect all life slices of autonomy. The agent herself can choose to give up future slices of autonomy by allowing other agents to associatively change her values. But on the balanced-autonomy reading more stringent demands are made on the consent procedure, because future autonomy is allocated more weight. For this reason, the agent must not just know to whom she consents, but also to what she consents. That is, she must know the content of the influence, the method by which the influence is communicated and the extent of the influence. If one of these requirements is reasonable from a balanced-autonomy perspective, then carte blanche consent is insufficient. Which one of these two approaches to consent is the correct one is unexplored territory; answering it would require extensive normative analysis that I cannot hope to provide here. I therefore remain undecided as to which of these approaches to autonomy is the correct one.

Finally, with regard to the level at which explicit consent applies, we must note that the level of society at large is problematic because there is no right of entry in international law. If there is no possibility to enter another country, then there is also no possibility to leave your own country if you refuse to give consent. And consent without the option of rejection is meaningless. Furthermore, explicit consent at the level of individual messages is possible – from a practical perspective – only if these messages are limited in number. Thus, explicit consent to covert influence will be restricted in practice mainly to contexts in which message types can be expected.

4.2 Thresholds of Authenticity
The most pressing problem in finding a minimal interpretation of the notion of autonomy as authenticity is that associative processing bypasses initiatory mechanisms. I will here discuss three solutions to this problem and make a case for accepting the final one.

The first solution can be extrapolated from the work of Robert Noggle. He makes a distinction between desires that are formed on the basis of beliefs and those that are formed on the basis of quasi-beliefs. A quasi-belief is defined as a belief that lacks one or more of the four properties of a full-blooded belief: integration, formation by cognitive-epistemic means, possibility of introspection and the willingness to mentally assert the mental state in question. The kind of states that Noggle has in mind when speaking of quasi-beliefs are the ones brought about by conditioning. Rats, for example, can be conditioned to run if they are hungry, provided they are rewarded in the right manner. In a similar way, behavior therapists
can change behavior in a way that bypasses the patient’s awareness.\textsuperscript{178} Noggle next acknowledges that we cannot simply require that quasi-beliefs be rooted out of our cognitive system, because they are a constitutive part of it and fulfill important functions. Hence for him the task of a theory of autonomy is to stake out which quasi-beliefs are acceptable and which are not, and Noggle’s proposal is to say that an alien desire is one “that is generated by a quasi-belief whose content conflicts with that of a straightforward (possibly only tacit) belief.”\textsuperscript{179}

The advantage of this approach is that it defines an objective standard for authentic attitude formation: concordance with existing mental states. On the downside minimally two points must be mentioned. First, the standard of concordance between beliefs and quasi-beliefs may be unrealistic, because we noted that divergences between implicit attitudes (which are primarily affected by misattribution EC) and explicit attitudes towards the same object are endemic.\textsuperscript{180} Second, observers who want to determine whether conditioning violates the autonomy of an agent would need to know the full range of beliefs in order to establish whether the conditioned quasi-beliefs conflict with them. It is not clear how such knowledge could be acquired practically. For this reason, the liberal state cannot know when conditioning must be prevented.\textsuperscript{181}

A second way of dealing with associative processing in the context of authentic mental state formation is to say that the centrality of changed mental states determines the degree of authenticity loss. It is even conceivable that a reasonable threshold can be found in this regard. However, a problem for this approach is that there are cases of central mental state and behavior change that – despite their centrality – cannot plausibly be said to impair minimal authenticity because they are part of the human condition. In particular, unconscious processing is a structural part of both dyadic interactions and group life. A constitutive component of dyadic interactions is so called mirroring behavior, which generates rapport that is required to sustain comfortable trust-generating interactions.\textsuperscript{182} In this way attitudes are changed (about the valence of the interaction partner) while progressively intense forms of mirroring can even result in the mimicking of behaviors. All this occurs outside of awareness. A similar phenomenon occurs in the context of group behavior, where so called automatic goal contagion is facilitated by group membership. In one experiment, students who saw competitive behavior acted more competitively themselves in a follow-up task only if they were told (or naturally perceived) that the people who initially displayed competitive behavior were part of the same group (in this case the same university).\textsuperscript{183} Again, subjects had no awareness of this bias, while such effects can reasonably be assumed to lead to lasting attitude change. Surely, it would be absurd to suppose that the liberal state should act against interactions and associations on the ground that their effects are central to participants.

This brings me to the third solution, which is to say that minimal authenticity is only violated if the associative processing channel was activated by another person. This solution is able to explain why dyadic interactions and group life do not in themselves violate autonomy (as is the case on the centrality account) while it still captures much of the intuition that the effects of associative processing are problematic because they are not properly related to existing mental states. It may be objected that on the current definition authentic mental state formation is indistinguishable from self-controlled attitude formation. In response, two points must be noted. First of all, it is not clear why this would be an objection. As long as the conditions for proper mental state formation (for self-control and authenticity) are fueled by different intuitions it seems immaterial if they would turn out to have identical practical consequences. But secondly, authentic and self-controlled mental state formation are not identical. Violating self-control involves three things. The intention to bring about a certain state in a controlee, accurate predictions about the state that is actually brought about, and the
intention to use a message that prevents the controlee from knowing about the control. Violating authenticity requires only the intention to use a message that in fact bypasses initiatory mechanisms and some awareness that the message type in question is processed in an impaired way. Hence, there are cases of intentional attitude change in which one form of autonomy violation occurs but not the other. That is, a controller can take control of another agent in a way that does not bypass initiatory mechanisms (so that authenticity is not violated). And an authenticity reducer can bypass initiatory mechanisms without having accurate knowledge about the effect of the relevant message on the receiver’s behavior (so that self-control is not violated). Moreover, because there are fewer constraints on what counts as a valid intention to subvert authenticity, the retrievability and consent conditions are more modest than in the case of self-control. That is, what is to be retrievable and consented to in the case of authenticity is, respectively, the intention to use a message that bypasses initiatory mechanisms (not mindreading capacities) and some awareness that the message has effects that are associated with relevant processing channels (not the intention to bring about a certain mental state or behavior in another agent).

Once the authenticity requirement of autonomy is said to depend on the intentions of other agents, there are also better prospects for defining thresholds for authentic action. With regard to the retrieval of explicit self-related states we could formulate the negative requirement that other agents may not deliberatively increase the cognitive load of another agent. (Because in the case of interpretation cognitive load would bring down the level of action identification, and in case of reflection it could reduce access to explicit self-related states). With regard to the retrieval of implicit self-related states, a feasible negative requirement would be that other agents should not induce negative affect. The success of this modification in formulating practical and retrievable thresholds is mixed however. Making cognitive load contingent on it being induced does not make much of a difference, since it is not clear in which situations such inductions should be allowed (and how much). The prospects for the negative affect induction requirement look much better though, because one can in principle test whether it is correlated with false self ascriptions (FSAs).

This completes my analysis of the minimal components of autonomy, which can now be pieced together. An autonomous agent is self-controlled and authentic with regard to both action and mental state formation. An agent acts in a self-controlled way if her system 2 reasoning system is not impaired, if her goals are not unconsciously activated and if she exhibits appropriate impulse control. Self-controlled mental state formation is guaranteed if the agent is not deluded and if she is aware of the intention on the part of a controller to evoke a specific behavior (and takes this intention into account in processing a message). An agent acts authentically when she has appropriate access to her implicit and explicit self-related states, which is secured by the absence of negative affect and cognitive load induction. Finally, authentic mental state formation is guaranteed if initiatory mechanisms related to conflict detection are not bypassed intentionally by another agent. For influence attempts to violate autonomy with regard to both self-control and authenticity it must be true that the effects of influence are central, that the relevant intention (which is not the same for self-control and authenticity) of the influencer is retrievable and that relevant consent (which is also not the same for self-control and authenticity) has not been given.

It is my contention that this framework is maximally integrative of intuitions about liberal autonomy under constraints of consistency and realism. In order to falsify this claim it need be possible only to add either an interpretation of one of the seven core intuitions or a mechanism that tells us in what these intuitions consist, without this addition being unreal-
istic or causing inconsistency. Throughout this chapter I have argued that at least conative hierarchy, evaluative reflection, the authoritative point of view of the acting agent, substantive theories about the self and excessively intellectual theories of action fail to do so. If these are the most likely candidates for falsifying the claim to maximal inclusiveness, and if my identification of the seven core intuitions of liberal autonomy is acceptable, then it is reasonable to assume that at least something like my minimal theory of autonomy meets my design specifications. If it is also accepted that my design specifications are an appropriate safeguard against perfectionist worries, then this theory supports anti-perfectionist autonomy-minded liberalism. Finally, with regard to the subsidiary aim of this chapter I think that it is fair to say that my theory is sufficiently rich in empirical detail to allow for the concrete assessment of autonomy violation in society. But whether this is indeed the case can be established only in the course of such an assessment, which I am about to carry out in the upcoming chapters.

Let me close with a conditional statement on the value of liberal autonomy. Even if autonomy violations – say, in the form of independence violations – are sufficiently central to the receiver, if they can be retrieved in the right way and if they have not been consented to, it need not follow that the liberal state is to take measures against them. For that argument makes the problematic assumption that autonomy is the supreme value that trumps all other values. Instead, appropriately context sensitive political philosophy must carefully weigh moral, ethical, prudential and realistic concerns for every policy or policy area. Hence, instances of autonomy violation are only pro tanto problematic.
Now that we know the empirical conditions for minimal liberal autonomy it seems that there is an easy way to determine whether branding undermines any one of them. We could take a representative sample of for example advertisements, show them to subjects in a laboratory and determine whether the thresholds for autonomy have been transgressed. Because no ready standard for cognitive load was available, we would restrict ourselves to the thresholds for negative affect induction, evaluative conditioning and subliminal influence. Thus, we would see whether the acquisition of the brand-related mental states co-occurred with negative affect (for state oriented persons), whether the mental states were formed as the result of covariation learning, and whether subjects were aware of the stimuli that caused their brand-related mental states to change. In order to establish the centrality of the attitude change, we would have to repeat the experiments over an extended time period. And for the purpose of determining whether producers of branding knew or should have known that the relevant aspects of the message they produced violated autonomy, we would also want to track down and interview a representative sample of the people involved in the production of the relevant advertisements.

Unfortunately, such research has not been carried out yet. Admittedly, as I will show in chapter 4, academic consumer research does show that there are cases in which advertisements transgress the specified thresholds for autonomous mental state formation. But since it is hardly feasible for the state to test every single brand message for autonomy violation, the sample that has been studied in consumer research is far too small to inform regulatory state action. As a way around this problem, my strategy is to give an account of structural features of branding that generate pressure in the direction of autonomy violation. Thus, if any one of such features were shown to drive autonomy violation in branding, then the state could realistically develop regulatory policies. I define structural features of branding as stable conditions, where conditions are sufficiently stable if they are unlikely to change in the foreseeable future. These features include universal mechanism of human psychology (genetically determined or shaped in the same way by universal environmental conditions), institutions supportive of market economies and cross-cultural social norms. If we have a list of such conditions, then we can check whether they push branding in the direction of transgressing autonomy thresholds (on their own or in interaction). And if we can locate the position of organizations within this structure, then we can also hope to determine what agents with specific roles can be reasonably expected to intend.

For this strategy to work, we need first of all a theory about how psychological, economic and cultural factors relate to each other. This is a contested issue in the social sciences, and not one that I can hope to resolve here. Instead I propose to opt – in line with the minimalist design of the overall argument – for the most conservative theory. Because results of the hard sciences are less contested than those of the humanities, a conservative approach would
favor the theory of which the greatest share of assumptions are grounded in the hard sciences. A further requirement would of course have to be that the theory makes non-trivial falsifiable predictions, for otherwise the gain in reliability does not help to explain anything. On the basis of these criteria I posit that the epidemiological approach to culture developed by Dan Sperber is to be preferred.\(^2\)

On this account, culture is made up of causal chains of representations that are mediated by processing devices (i.e. human minds). There are mental representations and public representations, and these are interchangeably linked to each other. Mental representations are located inside processing devices and public representations outside of it. When we zoom in on one such a causal chain, we might see something like this. The vibration of a set of vocal cords in a throat causes a phoneme (public representation), which is then transduced by the auditory system in another person's brain. After some complicated inferential processing the interpreted thought is stored in the memory of a speaker, where it may one day in some form – after some motivational processing – cause (mental representation) the motor area of the brain to trigger a vibration of the vocal cord (public representation), and so the process continues (until a thought is lost in all relevant memories). The highly complicated array of such chains in a society is a culture, and branded messages are a subset of a culture (or several cultures in the case of global brands).

In order to explain why some chains of representations stabilize and others do not, Sperber points to psychological and ecological factors. The most important ecological factors are the availability of external memory devices (e.g. writing), institutions engaged in the transmission of the representation (e.g. churches), and the recurrence of the situation in which the representation gives rise to an action which causes it to be transmitted (e.g. danger). Psychological factors include the relevance of a representation to the background knowledge of a person, the ease with which a representation is memorized, and the motivation to communicate (or otherwise act on) the representation. Furthermore, as in epidemiology proper, Sperber does not offer an overarching theory for the spread of representations, but merely a set of tools with which specific chains can be explained on an ad hoc basis.\(^3\)

Sperber's work focuses mainly on psychological factors, for the purpose of which he has developed a ‘massively modular’ account of the mind.\(^4\) On this view there are perceptual modules that process external stimuli and conceptual modules that receive their input from a perceptual module or from another conceptual module (or both). Because processing is a cognitively costly affair, Sperber assumes that processing in higher-order conceptual modules only occurs for representations that are in a buffer which he calls attention (provided they meet an input condition of the higher-order conceptual module). Competition among representations moreover is fierce and regulated by a number of relevance principles.\(^5\) With regard to memory, Sperber assumes that beliefs about states of affairs (which he calls intuitive beliefs) are formed on the basis of innate templates for which parameters are set by the culture in which a person grows up. As such, ideas about objects follow specifications from a ‘physics module’, those about nature the ‘folkbiology module’ and those about other people the ‘mindreading module’. His assumption is that representations that violate such templates are inherently interesting and more memorable, which according to him might for example explain the spread of myths and religious beliefs. The motivation to communicate a representation, finally, has not yet received much attention in Sperber’s theorizing.\(^6\)

On the basis of this theory we can begin to assess the structure of the conditions of branding. Whereas psychological and cultural conditions will prove to be complex topics (which are dealt with later in this chapter), economic conditions of branding are relatively straightforward. It will prove useful to lay them out briefly here, because they constrain the framing
of subsequent relevant conditions. Simply put, the main economic condition derives from
the assumption that brands that cause more sales have greater survival value. By corollary,
brand managers that happen to produce brand features that cause more sales will have
greater survival chances within companies. To be sure, the causal chain of brand production
is a lot more complicated. The brand message itself is produced by advertising creatives who
may simply follow fashion trends among colleagues, craft messages on the basis of prejudices
or follow the advice of some marketing guru. But their fate is decided by corporate execu-
tives who are in turn under pressure from shareholders to increase profits. Surely, there is
also a wide array of considerations on the minds of executives – such as adherence to com-
pany traditions and personal biases – and there is certainly also a process of negotiation with
creatives, but at the end of the day a company must make a profit. And we can assume that
the companies which produce branding that causes more sales will outcompete companies
that do not. Although tracking measures are imperfect, firms do have some notion of the
relation between their advertising budget and sales over time. Hence, companies with agents
who happen to prefer branded message features that more efficiently increase sales (or prefer
agencies that produce messages which increase sales) will be selected. For this reason,
autonomy violation that results from branding that increases sales will count as structural.
Other forms of branding may also thrive, but since they will be the result of company tradi-
tions, confused wisdoms of the advertising trade or personal whims they are inherently insta-
ble. In terms of Sperber’s framework we could summarize the above by saying that branding
organizations are institutions which transmit public representations because of a feedback
loop with sales.7

On this basis we can already say that at least unconscious goal activation for branding
that has a direct impact on behavior (i.e. in-store branding) is likely to be used structurally.
For if a goal is activated unconsciously, the consumer is not able to apply persuasion knowl-
edge to the message and will hence – all things being equal – be more likely to act on the
goal. This assumption is elegantly illustrated in an experiment in which the Wal-Mart logo
– which is branded for thrift – reduced spending behavior, whereas its slogan “Save money.
Live better” increased spending.8 This interesting effect is best explained with reference to
the persuasion knowledge that consumers bring to bear on explicit slogans in shops. They
know that store managers are out to persuade them and hence automatically correct – and
even overcorrect – for such messages. Thus, in shopping contexts unconscious goal activa-
tion will overall be more effective in increasing sales and can hence be expected to be
applied for structural reasons. Whether introjection that is due to negative affect induction
is equally effective under conditions in which persuasion knowledge is applied is a question
that has not been tested yet and therefore remain an open question.

With regard to branding that changes mental states, matters are more complicated
because the strength of mental states that result from associative processing is far weaker
than the strength of mental states that result from deliberative processing. That is, we saw
that deliberative processing generates stronger implicit attitude effects (as well as explicit
attitude effects) than associative processing.9 Because the persuasion threat is much more
direct in the shopping context than in the context in which advertisements are processed it
is not evident that associative processing is ever more effective in advertising. With regard to
mental state formation in branding we therefore need a more elaborate argumentative strat-
agy in order to establish whether or not it involves structural autonomy violation. Such a
strategy might look something like this. If we could isolate the set of combinations of proper-
ties of branded messages that are most efficient in increasing sales (under specified psycho-
logical and cultural conditions) then we could determine whether any subset of them struc-
turally transgresses autonomy thresholds. Properties of branded messages refer to content (themes, values or emotions) as well as form (rhetorical mode or persuasion technique). Although it is theoretically possible that there is a crude correlation between a single one of these properties and sales effects, this is unlikely. It is more realistic to suppose that there are correlations between combinations of these properties and sales. One can imagine for example that certain emotions enhance the effect of a persuasion technique, or that certain themes are most effectively communicated when metaphors are used.

There are two relevant routes by which the relation between such sales-enhancing properties of branding and autonomy violation can be established. The difference consists in the order in which relations are established. The direct route looks for correlations between properties and sales first, and relates the resulting subset of properties to psychological and cultural conditions in a second step in order to determine their structural nature. It can then be established whether this still narrower subset of properties violates autonomy. The indirect route relates sales to psychological and cultural conditions, which results in a set of constraints for properties being effective in increasing sales in the long run. Hence, for properties to structurally cause sales it is necessary for them to overcome these constraints. If we can determine how they must do so, then we can establish whether this occurs in a way that violates autonomy.

I pursue the direct route in section 1 on the basis of consumer research, but because it has failed to find correlations between sales and relevant properties of branded messages (i.e. those that can in principle be implied in autonomy violation), this route is blocked to me. This forces me in the remaining sections to push the indirect argument. Here I focus on general psychological conditions, because they are most structural. Among psychological conditions I select motivation on the ground that if there are general motivations, then all consumption behavior (and hence sales) must be caused by them. In section 2 a case is made for the existence of such general motivations. Section 3 shows that a large chunk of products is bought for public consumption and offers empirical evidence for thinking that such consumption is in large part motivated by three higher-order social motivations. Pre-

---

**Figure 3.1** A simplified model of the causal chain of branding as a positive feedback loop with sales. Boxes stand for mental representations and ovals stand for public representations.
dictions are derived from motivational psychology about the conditions under which higher-order social motivations should cause specific types of product choices, and these predictions are supported by evidence from consumer psychology. In section 4 I extend this line of reasoning by formulating a set of hypotheses that allow one to test whether the relations found in laboratories carry over to the level of society. In the final section, relevant managerial implications are explored (that is, implications relevant to the question of autonomy violation). Given the complexity of human psychology, we may assume that there are not only cues that trigger general motivations, but also psychological obstacles that obstruct this triggering. If either the triggering itself or the circumvention of obstacles is more effective by means that violate autonomy, then such autonomy violation is of a structural nature. I will argue in section 5 that there are such cases, and therefore predict that autonomy violation in branding is structural.

This argumentative design is not only more manageable than the empirical approach sketched out in the beginning of this chapter, but it is also more relevant from a normative point of view. For if we know the structural causes of a problem, we can respond with more measured responses than if we just know the size of a problem.

1 When Branding Leads to Sales

The direct route to knowing whether branding structurally violates autonomy begins with isolating the properties of branding that are most effective in enhancing sales. Fortunately, a substantial academic community attempts to answer this question. In particular, those working on the phenomenon brand equity attempt to establish which consumer attitudes towards brands generate profits (1.1) and those working on advertising effectiveness attempt to find which techniques most effectively enhance sales (1.2).

1.1 Brand Equity and Consumer Attitudes

Attention in marketing theory shifted towards long-term effects of promotion from at least the early 1980s. Theorists such as David Aaker and Kevin Keller argued that the difference between the price of a branded product and a similar unbranded product must be due not to individual advertisements or a seasonal campaign, but to promotion of a brand over the course of decades (although there is of course some role for inter-consumer communication as well). Ever since, several attempts have been made to put a number on this value, which I will discuss under the heading of financial approaches to brand equity. Putting a number on brand equity does not explain how it came about. To this end, so called consumer based approaches to brand equity attempt to identify the mental representations in the consumer’s mind that lead her to pay a premium for a brand.

The financial value of brand equity is determined by measuring the difference between the price of a branded product and the price of an equivalent product that is not branded. This is established by analyzing sales figures of the brand against sales figures of an equivalent generic product. The difference cannot be due to any attribute of the product or service, and so – if distribution differences are controlled for – must be due to the brand image. So called firm level analyses extend this product level analysis by assessing the revenue of a firm that can be attributed to the brand. This requires more complicated subsequent calculations.
that are carried out in various ways. For present purposes it suffices to sketch one prominent approach taken by the agency Interbrand. Their calculation of brands as assets proceeds in three steps. First, the annual earnings of the branded products of a firm are established. Because the brand is treated as an asset, one needs to add to these earnings the projected future revenues of the same products. From this sum taxes, operating costs and capital costs are subtracted. In a second step one must subtract from the resulting economic earnings the value that is generated by the product attributes themselves (so that only the value due to the brand image remains). Third, a more fine-grained assessment is made of the riskiness of future earnings. To this end, projections of demand for the brand are made according to criteria such as market position and diversification.\footnote{10}

Other research projects attempt to explain the sources of the financial brand equity in the minds of consumers. Just as in the case of financial brand equity, consumer brand equity researchers fail to agree on a methodology or on basic concepts. The first problem is that much of the research does not measure financial brand equity but the intention to purchase or laboratory product choice. And these measures are of course imperfect proxies. To give one example, although brand awareness is found to correlate significantly with purchase intention in the laboratory, there is no such clear-cut correlation with sales data.\footnote{11}

Second, the survey measures of the psychological ‘sources’ of brand equity are not very informative for our purposes. A very popular measure is ‘perceived value’.\footnote{12} But what exactly does the survey response that a brand is perceived as valuable tell us? Sure enough a correlation with purchase intention is found (who would want to buy a worthless object?), but saying that someone buys a product because she values it is like saying she has a preference for it. It merely leads us to reformulate the question ‘What is the source of the brand?’ into ‘Why do people value a brand?’ More useful are studies that attempt to distinguish brand equity that is ascribed on the basis of attributes of the branded product or not. This research shows that the ratings of attributes of a product (say the durability of a phone and its signal strength) are less important than ratings of nonattribute perceptions for predicting brand choice.\footnote{13} This is a valuable insight in itself, but for our purposes we are particularly interested in which aspects of these nonattribute perceptions predict brand choice. We want to know which themes, values, persuasion techniques and rhetorical modes cause the sale of brands. And to this level of analysis consumer based brand equity has simply not yet progressed.

1.2 Advertising Effectiveness Research

For such a more detailed analysis we would have to select a form of branding that is the easiest to manipulate and monitor. For this purpose T.V. advertising is a good candidate, because cable connections can be controlled on the household level. Furthermore, we would want to test the effect of these ads (and their features) not in the form of intentions to purchase the advertised product penned down on a survey form, but in the form of actual purchasing behavior in the real world. Such research does in fact exist, and is known as split cable research.\footnote{14} For such research a small community is selected in which two subgroups are selected that are balanced for demographics, previous purchases and shops visited. Because of the ability to manipulate T.V. content at the individual household level, it is possible to test the effectiveness of a particular ad (shown in group 1 households) against variations on the same ad, a rival ad for the same brand or no ad at all (shown in group 2 households). It is moreover possible to correlate evaluations of advertisings obtained in laboratory settings or by telephone interviews (such as recall) with sales figures for that same ad in the
real world. Because all the participants use ID customer cards for their shopping, which are tracked by the experimenters, purchasing effects can also be measured at the individual level.

What a meta-analysis of these enormous studies demonstrates is surprisingly unremarkable. Although the measure is still in use in many firms, it turns out that there is no (or at best a tiny) correlation between subject’s recall of an advertisement (obtained in experiments or via telephone) and purchasing behavior. A better idea for marketers is to concentrate their T.V. ads in time (say at the beginning or end of their campaign) rather than to distribute them evenly, because this distribution results in higher sales as measured over the course of an entire campaign. The authors offer no explanation for this effect on the psychological level though. It also turns out that, on average, ad copy that is changed leads to more sales than copy that is repeated over and over during a campaign. The most likely explanation for this effect would be to suppose a tedium effect of repetition that keeps back the effect of unchanged advertisements. Because the effect of changing advertisements is greatest for new brands, a preference for novelty may drive the effect. More interestingly, one split cable study tested several laboratory evaluations of advertisements (such as liking), which appears to have no or almost no predictive power. True, a moderate correlation was shown for the so-called ‘ARS persuasion measure’. But that measure tests product choice (from a set of products of the same category) in the laboratory after seeing an advertisement for the relevant product. And a correlation between product choice in the laboratory and the real world after seeing the same ad is not all that revealing from a psychological point of view. In sum, neither brand equity research nor the most sophisticated advertising effectiveness research has yet established correlation between sales and message aspects that are relevant from the perspective of autonomy.

2 Human Motivations

Because no relevant correlations between branding properties and sales have been established in consumer research we will have to infer these properties from first principles. This is what I called the indirect route. I plan to develop this more elaborate approach in four steps. If there are universal human motivations (step 1), then branding that enhances sales must appeal to these motivations (step 2). If we could isolate the subset of motivations that are most relevant to consumer behavior (step 3), then we could determine whether the most effective appeal to these motivations involves autonomy violation (step 4). This argumentative strategy assumes – and this is the first step, to be dealt with in this section – that we can find universal higher-order human motivations, and that is not an uncontested claim. For although psychologists universally agree that there is a set of basic innate motivations, they do not agree on the complexity of these innate motivations. I want to argue though that there are at least no decisive arguments against assuming that a limited set of evolved motivation modules produce all the varieties of human desires, and some theoretical arguments in favor of it (2.1). I then go on to offer suggestive empirical evidence for thinking that there are stable higher-order motivations (2.2).
2.1 Theoretical Arguments for Evolved Higher-order Human Motivations

The most comprehensive case against the idea that the mind consists of a set of complicated innate modules – including motivational modules – has been made by the philosopher David Buller. His first argument is that there would simply be insufficient genes to code for such detailed higher-order systems in the brain. It is known that about half of the human genome codes for the brain, and that a large bulk of those genes is involved in coding for perception functions. Hence, Buller argues, there are simply not enough genes left to do the coding of more complex higher-order functions. This argument ignores that the expression of genes relies on complex interaction with proteins, so that much detailed mapping is off-loaded, and genes merely control the general outline. If this is true, then a small number of genes could in principle code for the complicated innate algorithms that supposedly constitute higher-order modules.

Second, Buller maintains that developmental neurobiology would show that whereas ancient structures of the brain are under relatively tight genetic control, more recent cortical structures are not. That is, when the massive amount of neurons that are produced in early life are pruned, only subcortical pruning would be endogenous, while cortical neurons would be pruned only on the basis of experience. Because higher-order modules would presumably have to be located in the cortex, this would mean that they cannot be genetically programmed and hence that there are no such modules. This argument is not plausible either. For one thing, many cortical structures develop normally in animals that are artificially deprived of any relevant experience, which can only be due to genetic control over the initial locations of the neurons. Furthermore, if mental modules are formed exclusively on the basis of experience then they could still count as universal and adaptive provided the environmental pressures under which were formed are universal.

Finally, Buller argues that higher-order modules cannot exist in principle, because they require for their functioning information from other modules. He gives the example of the status module which would according to him have to rely on feedback input from other modules that measure status ranks in domains such as religion, politics or sports. If it accesses such modular databases (with regard to religious status rankings for example), Buller continues, then it is no longer a module. However, if it does not access such databases at all, then it cannot assess status behavior. If this is correct, then there cannot be complicated modules such as the status module. There are two serious problems with this argument. First, it is not necessary to assume that a status module must rely on information from (unrealistically) specialized modules for religion or sports. If there are universal signals of deference, then a status module could simply be equipped to read such signals of deference and infer that the deferred-to person must have high status, without having to know anything about the trait or activity that led the deferred-to person to become the object of deference. But a more fundamental flaw is, second, the assumption that modules must be encapsulated by definition. In this regard Peter Carruthers has proposed to make a distinction between encapsulation and inaccessibility. A module is encapsulated if it cannot draw on any outside information besides its input, where input is what activates the system. This would be the conception that Buller has in mind. It would for example mean that the mindreading module is turned on by intentional behavior and receives no information beyond that. However, an alternative conception of modularity is possible. Modules are inaccessible when they can not only receive input from one another (i.e. information that activates their processing) but can also search each other’s database for subsequent information. Crucially, such entities are still
modules because they cannot access each other’s processing. If the inaccessibility account of modularity is coherent, then Buller’s third argument against massive modularity also fails.

Let me now very briefly mention some theoretical reasons for thinking that there are higher-order modules. First of all, biological systems in general tend to be built in a modular fashion, because such a structure is less prone to developmental errors, because local damage is more easily contained and because a structure made up of modular parts can be more flexibly (and independently) adapted in response to environmental problems. And since the mind is an organ, it is likely to be modular as well (while abundant evidence of people with brain lesions who suffer a defect in only one functional domain supports this assumption). Second, an organism as complex as Homo sapiens must solve many learning challenges, and it is unlikely that these are all solved by a general learning mechanism. Instead, one would expect such processing to occur in specialized modules that are activated by inputs that are specific to a certain problem or opportunity domain. The number of modules would then depend on the number of stable problems that occurred in the Pleistocene, to which the mind is adapted. This account is coherent with thinking that modules will sometimes overlap. For because brain activity is so costly to the organism, there are good reasons to assume that modules will share processing systems with each other if they carry out the same computations. With that said, I conclude that there is a suggestive theoretical case for thinking that higher-order motivation modules exist.

2.2 Evidence for Higher-Order Human Motivations

If the mind is indeed modular, then there will be a fixed number of desire modules that have evolved to respond to stable problems. What we want to know then is how many motivational modules exist, or at least what the most important such modules are. There is certainly no lack of such lists, but these are of little use if we do not know what counts as evidence for a motivational module. Let us first discuss some proposals that clearly cannot count as such evidence.

The psychologist Steven Reiss had thousands of subjects pick motivations that they thought were driving their behavior from a list of hundreds that he had come across in the psychological literature and then extended himself. After some factor analysis he managed to categorize motivations in 15 groups (to which he later added another category), which he afterwards framed as evolved solution to problems from the Pleistocene. In a similar study, the psychologist Larry Bernard and colleagues reversed this order. That is, they first considered stable problems faced by our Pleistocene ancestors, which they grouped according to the social system in which these must have emerged. Specifically, at the individual level they posit a self-protection challenge, at the level of the dyad ones for mating and affective awareness, at the small group level the challenge of relationship maintenance and parental care, at the large group level the challenge of coalition formation and at the highest level not so much a challenge as the danger of motivational infection by memes. Bernard and colleagues then conducted a survey that was to establish whether these categories represented distinctive categories. Whatever the outcome of the studies of either Reiss or Bernard, and whatever the order of definition of motivations, there is no compelling reason for thinking that they will tell us anything about evolved motivations. The reason for this is that using surveys assume that conscious goal states are a window into motivational modules. And that assumption is true only if evolved motivations would cause actions exclusively via the conscious system 2 reasoning process, which is highly unlikely.
So what does count as evidence for a higher-order social motivation module? One type of evidence would be a delineation of the range of inputs that activate a motivation module. Obviously, absent fine-grained neuroimaging (which is not yet possible), in order to know whether a stimulus activates a module we must look for patterns of behavior that consistently correlate with exposure to a relevant stimulus (and that must be predicted in advance on the basis of assumptions about stable problems of humans living in the Pleistocene). Furthermore, as in basic motivational systems that regulate energy reserves, fluid balance, carbon dioxide levels and body temperature, there should be special feedback mechanisms that inform social motivation modules about the state of relationships to other people.30 Just as blood sugar levels tell us to take another bite and insulin tells us to stop eating, so social feedback mechanisms should regulate the strength of a social motivation. In this regard, so called sociometer theory maintains that one such mechanism is global self-esteem. In light of very strong correlations between social exclusion and global self-esteem, it proposes that self-esteem is an evolved meter that signals to the organism whether it is recognized by members of a group.31 Because of the potentially disastrous effects of ostracism in Pleistocene life, a meter that alarms the organism that it is about to be kicked out of a group would be a great advantage. In following through on the evolutionary logic of the idea of a sociometer, Lee Kirkpatrick and Bruce Ellis propose that there should be many such meters because there were multiple stable social problems in the Pleistocene. Besides group affiliation, humans have an interest in also gauging the state of at least relations of dominance, prestige and love.32 On the basis of this framework, I will argue below that there is suggestive evidence for higher-order (and partly overlapping) social modules for group affiliation, mating and status rank.

Affiliation Motivation

There may have been three stable challenges for individuals in relation to groups in the Pleistocene Environment of Evolutionary Adaptiveness (EEA). First, due to the prolificacy of in-group conflict, an individual must have adapted its behavior in order to avoid ostracism. Second, an individual would have wanted to know whether group members could be trusted to do their share of cooperative work. And third, in times of war an individual would have wanted reliable protection from a group. In this section I briefly discuss three lines of research on group behavior, which taken together suggest that evolution may have designed motivations that help individuals cope with these challenges and specify which inputs trigger these motivations.

Whatever the precise social organization of human life in the Pleistocene may have been, survival chances inside groups must have been significantly higher than outside of them. Not only are reproductive chances slim on one’s own, but survival probability is also diminished, given the lack of food sharing and protection against predators or hostile bands. For this reason, one may expect evolution to have designed a mechanism that alarms the organism when it is in danger of being ostracized from a group.33 Sociometer research suggests that to this end self-esteem may function as a meter that tracks recognition from group members. Self-esteem may function in this way because self-reported self-esteem correlates highly with the level of inclusion in real as well as imagined situations. Moreover, if exclusion is experimentally induced, levels of self-esteem plummet. In one experiment, subjects rated and rank-ordered three other participants in terms of how eager they were to participate with them in a cooperative follow-up task for which three persons were required. After the questionnaires were collected, the subject was told in one condition that she was excluded from the follow-up task because of a random selection procedure, and in another condition
that her exclusion was based on the ratings of the other three participants. Self-esteem decreased only in the latter condition. What is more, exclusion appears to activate the motivation to affiliate, since a series of experiments has established that it is correlated with greater agreement to incorrect majority perceptual judgments, greater willingness to join regular (e.g., a job training group for students) as well as dubious groups (e.g., a mind-control group for spoon bending), better memory for social information, and greater unconscious mimicking of an interaction partner (especially if the interaction partner belongs to the same group).

More evidence for the motivation to engage in affiliation behavior comes from social identity theory. Experiments within the so-called minimal group paradigm show that cues for group membership such as wearing the same t-shirt influences the subsequent allocation of resources in favor of group members, while subjects are not aware of this bias. Apparently, there is a database in which membership characteristics are stored and whenever one or more of these are perceived a desire to benefit the person is generated. Although the theory was not designed with an eye to validating predictions about evolved motivations, a case can be made for thinking that the need to cooperate may have driven the evolution of these mechanisms. For experiments show that social identity is a better guarantor of keeping members of a group together in the face of attractive alternative options than are perceptions of investments or norms. If there are moreover honest signals that indicate that a person is motivated by the social identity mechanism and if this mechanism operates automatically, then others can rely on these signals in deciding on whether to cooperate or not. Evidence for the existence of such signals is offered in chapter 5 (when we turn to evolutionary explanations of religion).

A third type of input to the affiliation module is fear of death and related coalitional challenges. Evidence for the relation between death threats and positive in-group behavior has been collected by decades of Terror Management Theory (TMT) research. This theory holds that fear of death is so paralyzing that we are strongly motivated to escape into worldviews that assuage it. In particular, TMT proponents have speculated that those individuals who are better at avoiding thinking about death—due to a 'survival instinct'—must have been more efficient in whatever tasks they were carrying out, and hence must have had better survival chances. This reasoning has been criticized by several evolutionary psychologists though, on the ground that evolution does not build general survival instincts (instead it builds mechanisms that avoid specific threats) and that the anthropological record demonstrates that worldviews are as much fear-inducing as fear-assuaging. Whereas TMT explains the tendency of individuals to affiliate with groups as an element of the escape into a worldview, an alternative evolutionary account sees the need to join a group as primary and worldview preference as derivative. On the latter account, the need to affiliate is explained as an adaptive response to coalitional threats. Given the prolificacy of warfare among Pleistocene communities, an urge to cling to a group in times of extreme danger is likely to have enhanced survival chances. Evidence for this account comes from experiments in which subjects upon perceiving a fear of death cue affiliate with (in this case: sit closer to) group members even if these individuals previously attacked their worldview.

**Mating Motivation**

With regard to mating, motivational structures may have been a response to two stable problems: mate acquisition and mate retention. Evolutionary psychologists claim that both are structured by complicated assessment and behavior mechanisms, but I will focus here on the motivation to acquire mates because it has been the object of research in consumer
psychology (and thus allows me extend my argument in that regard later on). The motivation for mate acquisition is claimed to be structured by complicated design architecture in two respects: there may be sophisticated innate criteria for preferring one potential mate over another and there may be pre-programmed mating strategies. I will discuss these components of the motivation for mate acquisition in this order.

Some criteria for mate assessment are shared by men and women while others are adaptations that have arisen in response to sex-specific problems. The first type of preferences has emerged because both men and women have an interest in a partner with so called ‘good genes’ (i.e. genes that benefit potential offspring) as well as compatible and diverse genes.41 A reliable way of measuring the quality of genes is fluctuating asymmetry (i.e. deviation from perfect symmetry of a trait that is on average symmetrical in the population) of a set of bodily features. Such asymmetry is correlated not only with lower fitness, but also with lower mate preference of both men and women in industrialized as well as developing countries.42 It is not clear however whether humans assess asymmetry as such or conditions that are associated with it, such as (for men) peer status, masculine faces, masculinility, voice quality, dance movements and scent.43

A criterion that men may use to assess potential mates, beyond symmetry, is signals of residual reproductive value. Crucial elements that determine a women’s residual reproductive value – such as the quality of her ova or the storage of a special type of body fat – is modulated by the production of estrogen. Hence, bodily features that are shaped as a result of the release of estrogen may signal reproductive value, provided such features are not easily produced by other means. That is, there should be a serious cost involved in producing such features for them to function as honest signals. Although there are serious costs associated with the release of estrogen – such as the diminishment of the functioning of the immune system – it might be possible to develop relevant features by different and non-costly means. With regard to body fat such concerns are immaterial though, because the relevant type of fat is directly useful for the development of the fetus.44 That is, so called gynoid fat is different from android fat in that it is less easily accessed for survival functions and contains nutrients that are essential to the fetuses’ brain development. It is stored in different depots – breasts, hips, buttocks and thighs – and such storage turns out to correlate with attractiveness ratings by men.45 There is also evidence that estrogen release influences the development of faces (e.g. facial bone growth is capped, so that the lower face remains small), smoother skin and a higher pitched voice.46 But in this regard it is not clear why women with low reproductive value have not developed such features independent of estrogen release and so faked having a high reproductive value.47 That is, future research must rule out the existence of feasible alternative developmental systems.

For their part, women appear to be equipped with innate criteria for assessing men in terms of their ability and willingness to provide material benefits (i.e. shelter, food, protection and teaching children). There are two relevant types of evidence for this claim. The first is the existence of what Randy Thornhill and Steven Gangestad call ‘extended female sexuality’: the fact that human females are sexually active and initiate sex when they are not fertile (where the fertile period is limited to a few days before ovulation). This behavior is most plausibly interpreted as a strategy by which women compete for the acquisition of those men most able and willing to provide material benefits. This is plausible because in other species extended sexuality is associated with males giving material benefits to partners, with reproductive benefits of females and with females assessing men on the basis of different criteria when fertile (i.e. criteria other than the ones that indicate the provision of material benefits).48 Second, with regard to food, the ethnographic record shows that of the 95 foraging societies...
about which data are available, 64% of calories are provided by men, while male contributions to a family unit correlate with lower child mortality and higher reproductive rates.49 However, for women to be able to assess men for their capacity and willingness to contribute material benefits, there must be relevant criteria. With regard to capacity, the most relevant criterion appears to be status. The mechanism for the assessment of status is the subject of the next section. Here I note merely that there are prestige orders based on for example hunting skills in foraging societies, which are relevant in terms of men’s ability to provide calories, while social dominance would be relevant in terms of protection (of the women herself as well as her offspring). Both prestige and dominance are qualities in a mate that women value.50 The willingness of men to provide these benefits (and not to run away with another woman after conception) may be assessed by kindness, generosity and emotional openness, which women across cultures value highly on self-report measures of mate preference.51

Interestingly, as we noted, across species females with extended sexuality have different preferences when they are fertile. This ‘ovulatory shift’, in which mating preferences shift from concerns relating to material benefits to concerns for good genes, has also been consistently observed with human females in a number of respects. That is, it is consistently found that women at peak fertility (i.e. 2 to 3 days before ovulation) prefer the scent of men who are rated as more symmetrical on 10 bodily traits and men who self-report to be socially dominant. They also prefer male faces that are rated as more masculine (provided they rate for short-term sexual interactions), more muscular male bodies, male voices that are lower pitched, and men who display more dominant behavior based on ratings of video-taped material.52 Because this shift of mating preferences in the direction of good genes disappears when women take contraceptives that inhibit hormone release associated with ovulation, it seems plausible to assume that the ovulatory shift may be caused by estrogen.53

The second way in which mating motivation may be structured by complicated mechanisms is in the form of mating strategies. In the Sexual Strategies Theory (SST) developed by David Buss and Steven Schmitt, a distinction is made between short-term mating strategies and long-term mating strategies (which involves extended courtship, parental investment, the emotion of love and resource contributions). If it were shown that the choice for any one mix of these strategies is regulated by automatic mechanisms, then this would suggest that there is a complex innate mating motivation. Likewise, if there would be automatic feedback mechanisms that inform the organism about success in mate acquisition, then this would also count as suggestive evidence for innate motivational mating mechanisms.

To begin with, on the basis of parental investment theory – i.e. women invest by default much more in children than men, in the form of nine months of pregnancy – SST predicts and finds that men more actively seek short-term relationships across cultures and whatever the male-female ratio.54 More interesting yet, the choice of mating strategy may be regulated by hormonal shifts in both men and women. When it comes to women, there is evidence that during the fertile phase – possibly because of the influence of estrogen – their strategy shifts towards extra-pair copulations with symmetric men (provided their partner is not himself symmetrical). That is, in a number of studies it has been found that self-reported desire for extra-pair copulations strongly increases during the fertile phase, while other studies show that this effect occurs mainly when a woman’s partner scores high on measures of fluctuating asymmetry of body traits.55 Women also report a lower desire for intimate sex when they are in their fertile phase, which is the type of sex they supposedly typically have with their partner.56 Finally, women in their fertile phase are consistently shown to walk and move more, while they also report a greater interest in attending social events where they
might meet men.\textsuperscript{57} With regard to men, higher levels of testosterone are associated with a more short-term mating strategy in a number of studies. It is not clear however whether this effect is due to temporary shifts in testosterone, to testosterone exposure in utero (leading to a more masculine brain) or both.\textsuperscript{58}

With regard to feedback mechanism there is evidence for effects of mate value on global self-esteem and effects of self-esteem on mating strategy. Effects of mate-value on self-esteem have been tested experimentally. In one study it was shown that the effect of negative feedback on a participants’ mate value on global self-esteem is greater than feedback on a person’s qualities as a friend. In a follow-up study to this experiment it was moreover found that when participants received feedback on their mate value, the negative effect on global self-esteem was greater for men if the feedback was about their status and greater for women when it was about their attractiveness.\textsuperscript{59} A cross-cultural study obtaining self-report measures from participants in dozens of modern nations suggests the function of such shifts in self-esteem. For it was found that global self-esteem is consistently associated with a higher frequency of reported short-term mating strategies of men, while there is either no (or a reversed) such association for women.\textsuperscript{60} This suggests that there may be a specialized device (or sociometer) for measuring success and failure in mate acquisition that scales up or down self-perceived mate value, which in turn may determine which mix of long- and short-term mating strategies is pursued.\textsuperscript{61}

\textit{Status Rank Motivation}

The function of dominance in primate groups is relatively clear. A dominance hierarchy is a ranking based on the superiority of one member over the other in a physical confrontation (or the threat thereof). Signs of dominance are universal across the primate world and include large size, physical strength, vigor, good health, being adult, being male and having a high-ranked mother.\textsuperscript{62} For females, a higher place in the dominance hierarchy results in greater access to food, grooming and access to higher quality mates.\textsuperscript{63} It is possible that primates come to learn that ending up high in the hierarchy brings benefits, but such learning is error-prone, and it is therefore more plausible to assume that primates have an innate desire for climbing in the status hierarchy. Moreover, humans show universal signs of deference and authority, which suggests that such behavior schemata are programmed (it would be an extreme coincidence if all human groups came up with exactly the same signs). Universal deference signs include submissive cowering, stooped posture, eye aversion, retreating from someone, nervousness and grinning.\textsuperscript{64} Universal dominance signs for humans include specifically (besides the one mentioned in the beginning of this section) facial authority signals. That is, cross-cultural studies find that holistic authority judgments of faces are consistent between cultures, although the precise constellation of features that accounts for these judgments has not yet been uncovered (e.g. smiling is universally seen as a sign of deference whereas eye brow position is interpreted differently across cultures).\textsuperscript{65}

That such features play a role in the formation of actual status hierarchies has been corroborated in a great number of studies on spontaneous rank orderings among humans and voice and facial dominance features. In one study Allan Mazur and colleagues even significantly predicted the number of promotions of officers on the basis of ratings of their faces for dominance.\textsuperscript{66} However, what we want to know is how these status signs are used in human behavior and whether this use can count as evidence of evolved modules. On this question there are two relevant models. One sees the use of status signs as regulated by testosterone while the other points to specialized sociometers.
The sociologist Allan Mazur has developed a model of dominance contests as games of stress induction. According to him, rank allocation in humans and nonhuman primates alike occurs not so much by means of fighting as by staring competitions. Alter and Ego both display universal and culturally diverse dominance signs (which I mentioned earlier) which induce stress in the other, measured in terms of depressions of thumb blood value and cortisol value in the blood. Both negatively affect health. The contest ends when Ego decides that the stress is not worth the effort, and accepts a subservient rank position to Alter. Winning contests boosts levels of testosterone in the blood (which has been measured in a great number of physical and non-physical contests\(^67\)), which may stimulate more confident expressions of dominance signs in the future, while losing contests has the opposite effect.\(^68\)

A second line of research extends the initial sociometer model by measuring self-esteem in terms of feelings of superiority (according to question such as ‘Are you easily pushed around?’). This domain specific meter is hypothesized to be specialized in the regulation of dominance displays. In an experiment conducted by Kirkpatrick and colleagues this hypothesis was put to the test. Subjects were asked to write an essay after which they received negative feedback by an evaluator in order to induce feelings of rejection. In a follow up task they were asked to prepare sandwiches for the evaluator in which they could choose to select hot sauce. It was argued that the decision on the amount of hot sauce offered subjects the opportunity to aggress against the evaluator. Regression analysis showed that measures of self-esteem in terms of superiority were a positive predictor of the level of aggression (in terms of the quantity of hot sauce), whereas self-esteem in terms of social inclusion was a negative predictor.\(^69\) I noted earlier that status rank may be conferred not only on the basis of dominance but also on the basis of prestige. If so, then there should be a distinct mechanism for regulating status rank behavior on the basis of prestige-relevant criteria. To this end, Kirkpatrick and colleagues devised another questionnaire that measures prestige-related self-esteem by asking questions like ‘Are your opinions valued by others?’ and ‘Do you make significant contributions to your social group?’ Because dominance related self-esteem informs the organism that it should engage in more contests, it should correlate with higher levels of testosterone (which is correlated with more aggression). However, people who are high in prestige rely on the presence of others in their vicinity for their position and therefore should want to avoid aggression. A high prestige self-esteem score should therefore decrease the desire to engage in aggression and not or negatively correlate with reported aggressive behavior. These predictions were subsequently partially validated.\(^70\)

Many uncertainties pertain to both models. As to the first model, the role of testosterone in causing dominance displays is still not settled, whereas sociometer theory of dominance and prestige relies still entirely on explicit measures of self-esteem.\(^71\) Moreover, both models only predict when humans are motivated to engage in status contests and do not predict when humans will be motivated to acquire status signs. It is furthermore unclear which adaptive strategies losers of status contests might employ to further their prospects, since a position at the bottom of the rank-order elicits aversive reactions. In this regard it is interesting to note that research on Vervet monkeys has demonstrated that serotonin levels in the blood of dominated individuals are lower, while experiments on human subjects show that experimental reductions of levels of serotonin result in irritability, sleep disorders and aggressiveness.\(^72\)

Taken together, research on evolved higher-order motivations suggests that much of social life may be driven by three modules that may include one or more sub-modules (e.g. the status rank module includes both dominance and prestige meters) and that may share cogni-
tive architecture with each other (e.g. the status rank and mating modules both motivate status signaling and may hence share systems that evaluate status rank in a group). The fact that subjects are not aware of the relations between inputs, motivation and behavior suggests that the mechanisms function automatically, which is exactly what one would expect of innate modules. Some evolutionary psychologists have taken this argument a step further and proposed that there is a pyramid of evolved needs in which parenting, mate acquisition and mate retention are at the top, because reproduction would be the primary evolutionary requirement. Such arguments should be advanced with great care however, because motivations are highly context dependent. It is one thing to assume that a given set of motivations will be of primary importance given the stability and importance of Pleistocene problems they evolved in response to, and quite another to place these problems in a hierarchy. For my purposes there is no need to take a stance on this issue.

3 Motivations and Consumption

From among the motivational modules that a massive modularity account implies we would want to limit ourselves to a subset that is most relevant to consumption. One subset of motivations for consumption has recently been proposed by Dan Ariely and Michael Norton, who want to distinguish between conceptual consumption and unmediated consumption. Consumption that is not mediated by concepts is motivated by immediate experiences of aspects of the product. Eating a cookie is presented as the paradigm case. However, the consumer research that Ariely and Norton discuss reveals that even experiences of elementary consumption, like eating and drinking, are themselves mediated by expectations. That is, an influential study showed that knowing that a beer contains vinegar affects taste only when this information is known before drinking, which is evidence for thinking that concepts influence the actual taste experience. It is not clear then whether there is all that much to which unmediated consumption could refer.

A more useful distinction is the one between consumption because of private desires or because of public desires. A private desire is not private in the sense that other people had no role in shaping it. Instead, private consumption means that the consideration of what others might think about the product plays no role in the decision to buy it. There is disagreement about cases in which consumers are not aware of other-regarding motives. However, given the predominant role of unconscious processes in decision making (chapter 2.2) it seems only reasonable to also count unaware motivations as public behavior. Scholars have traditionally struggled to define such unconscious motives to please or impress others in acts of consumption. But on the basis of my discussion of evolved human motivations an easy solution presents itself. We call all acts of consumption public that are caused by motivational modules that evolved in reaction to a social problem. This includes the modules for mating, affiliation and status, but also the motivations for consistency and accuracy that may have arisen in response to persuasion contexts.

Given that social relations are of primary importance in the EEA as well as today, we would expect that public consumption plays a major role. This expectation is easy to test. If social signaling drives consumption, then products that are more visible should have a lower price/quality relation. Angelo Chao and Juliet Schor tested this hypothesis for lipstick, mascara, eye shadow and facial cleanser, which subjects ranked as progressively less public.
They then had professional artists and students test the products, and give quality evaluations. According to expectations, highly visible lipstick had no positive price/quality relation, whereas the privately used facial cleanser did.\textsuperscript{78} If the theory of evolved social motivations developed by evolutionary psychologists is correct, we should also find evidence for acts of consumption resulting from cues that have been identified as inputs to higher-order motivation modules. Such evidence is given in the remainder of this section for affiliation (3.1), mating (3.2) and status rank (3.3) motivations.\textsuperscript{79}

3.1 Consuming Community

There are two relevant types of motivation to affiliate that can be activated in consumption. Consumption can occur because of affiliation with the people who use the brand (i.e. the brand community) or with a brand-external group.

There is evidence for the involvement of brand communities in the motivation to consume from both psychology and sociology. Psychological research suggests that coalitional threat cues of danger can activate the motivation to affiliate. Specifically, in an experiment carried out by Aric Rindfleisch and colleagues, one group of subjects was first asked to describe the emotions that thinking about their own death aroused. In a follow up task both this existential threat group and a control group were asked to fill in a survey on their communal-brand connections (involving questions such as ‘I really identify with people who use this brand’) to brands of mp3 players and sunglasses. Connections with the brand was higher for participants in the existential threat condition.\textsuperscript{80} Rindfleisch and colleagues also carried out a national survey in which it was found that people who report social insecurity show strong communal brand connections.\textsuperscript{81} Admittedly, the strengthening of brand community feelings when feeling insecure (or after perceiving an existential threat) is not evidence for an effect on consumption behavior. After all, the reported preference for affiliation brands for insecure and fearful subjects might not translate in actual purchases.

But if products are indeed bought to express group identity, then we should find that brands actually function in this way in social networks. That is, we should find evidence from sociology for consumers feeling part of a group that uses a product. Sociological evidence for the existence of so called brand communities comes from field studies by Albert Muniz and Thomas O’Guinn. They took notice of one brand community member and visited him in his neighborhood (called Fairlawn) where the researchers tracked down the network of fellow brand community members as well as members of other brand communities. In the end they interviewed members of the community of users of Saab, Mac and Michelin about what Muniz and O’Guinn understand as the three defining criteria of a community. They found evidence for common awareness (brand users felt part of a brand community), rituals and traditions (e.g. Saab drivers apparently flash their lights when passing by) and feelings of moral responsibility (brand community members help each other out when encountering product related problems).\textsuperscript{82}

More direct evidence of a relation between activation of the affiliation motivation and purchasing behavior exists for both pre-existing groups and entire generations. In a series of experiments carried out by Baumeister and colleagues, affiliative consumption was defined as the choice for a product that was related to a group the subject was a part of (instead of a brand community) while the relevant cue was social exclusion (instead of existential fear). The relevant result is that subjects who felt rejected after an experimental manipulation were more likely after the experiment to buy with their own money from a store in the laboratory items that expressed their belongingness to a group over items that did not. That is,
compared to a control group, they were no more likely to buy more practical items (such as a notepad, a magazine, bath gel and cookies), but they did buy more items that showed off their group affiliation (e.g. a wrist band of their university).\textsuperscript{83}

Brand managers can also activate the affiliation module by tapping into nostalgic behavioral programs that satisfy the consumer’s need to belong. That is, if a brand can be associated with an item in the set of nostalgic phenomena of a consumer, then whenever a consumer’s shopping behavior is motivated by the affiliation module, the brand will be considered more favorably. Katherine Loveland and colleagues carried out a number of studies that demonstrate that a need to belong is indeed correlated with a preference for nostalgic products.\textsuperscript{84} In one study they had participants fill in two questionnaires on their need to belong and an implicit measure of belongingness in which they had to recognize a subliminally flashed nonword letter string that actually had the shape of words related to belonging. A high score on these measures correlated with a preference for Dutch brands that had been rated as nostalgic (e.g. LU Prince and Peijnenburg). Moreover, when the need to belong was induced artificially by priming an interdependent self or by inducing social exclusion, subjects respectively picked more nostalgic movies or ate more nostalgic cookies.\textsuperscript{85}

### 3.2 Good Mate Value for Money

It was earlier pointed out that the mating module shares much of its architecture with the status module because of the sexual benefits that a high rank position brings with it. In the realm of consumer products, high status goods are an obvious way to show one’s position in the prestige order. Thus, if the mating module drives consumption, then men who are primed with attractive women should prefer buying status goods over saving money, compared to a control group (of men who are not primed with attractive women). This inference was corroborated by Vladas Griskevicius and colleagues, who found that men who were primed with attractive women were much more likely to buy a new watch, a vacation or a new car rather than save money.\textsuperscript{86} In another experiment it was shown that mating-primed men also prefer conspicuous goods (vacations, watches and cars) over everyday goods (kitchen staples, toiletries, household cleaning products) compared to a control group. Women showed no such behavior change.\textsuperscript{87}

In a large follow-up study, Griskevicius and a different set of colleagues pinned down two conditions in which primes of attractive women induce status-seeking behavior in men.\textsuperscript{88} One condition is the mating strategy to which men are inclined. On the basis of a questionnaire, in which a range of questions such as ‘Sex without love is okay’ were answered, men were given a rating on the so called sociosexual attitudes scale. On this basis the categories low-investment men (i.e. those who are more interested in short-term mating opportunities) and high-investment men (i.e. those who are more interested in long-term relationships) were formed (where an individual fell in either category by virtue of having a score at least one standard deviation above or below the mean). It then turned out that only low-investment men who were primed with pictures of eight attractive women were more likely to choose products that were rated by other students as conspicuous consumption goods (e.g. designer sunglasses, stereo sets for cars) over nonconspicuous goods (e.g. low-cost jeans, toaster oven). That is, high-investment men (as well as women in either category) showed no significant priming effect. This result was corroborated by having men from both categories rate a set of products. Besides some filler products there was the option to buy one of two wallets that were of supposedly comparable quality, appearance and price. However, one wallet had a low-budget retail store logo on it while the other was a replica that carried the
name of the prestigious brand ‘Coach’ on it. Again, low-investment men (who were this time primed with a romantic story) showed a greater preference for the high status item while the prime had no effect on high-investment men. As such, this study rules out the possibility that low-investment men chose high status items just to feel good about themselves (as buying a counterfeit is unlikely to do so).

A second condition (besides a person’s sex – and hence level of parental investment – and mating strategy) for the facilitation of status-seeking behavior is the type of prime that is used. To test the effects of this variable, Griskevicius and colleagues devised romantic stories that differed in the expected duration of the sexual relationship that was unfolding. In a vacation affair story, a man and a woman had a romantic dinner on the day before they departed, while in a campus affair story the romantic dinner was between students who expected to stay on the campus for an extended period. When the male participants were asked to indicate how much they were willing to spend on eight types of products that had previously been rated by other students for representing conspicuous consumption, only low-investment men showed an effect of the prime. More interestingly, this effect only existed for the short-term affair story. Surely, all this presumed status signaling by men only makes sense if women actually find men who engage in conspicuous consumption more attractive. Griskevicius therefore tested in yet another study how interested women were in a relationship with a man on the basis of a short biography. Between two groups of women, the manipulation was the name of a car. In one condition the description of a corporate executive included the fact that he owned a Honda Civic while in another condition he owned a Porsche. Women indicated a greater willingness to engage in an affair with the Porsche owning man, but only when the affair was short-term.

The mating motivation of women has been shown to be activated by presenting pictures of attractive men, by certain hormones and by pictures of attractive women (i.e. rivals). The main output behavior of mating motivated women is not the acquisition of high status goods, but rather behavior that enhances appearance. In an experiment, Kristina Durante and colleagues took the purchase of sexy clothing as such behavior. In a setting in which a fashion web shop was simulated, women who were primed with pictures of beautiful men indeed chose significantly more clothing products that had been rated as more sexy in a pre-test.

Interestingly, Durante and her colleagues also tested the effects of hormonal shifts during the menstrual cycle by measuring the level of the luteinizing hormone (LH), which is released about 48 hours before ovulation. Participants took part in the webshop test described above both when they were ovulating and when they were not (telephone interviews in advance allowed the researchers to estimate participants’ period of ovulation by asking about it obliquely). As a cover story, participants were told that the experiment was about fashion, health and relationships. A first finding of Durante and colleagues was that women in the ovulating condition chose 10% more clothing pieces that has been evaluated as sexy (compared to women in the non-ovulating condition). In follow up studies, the researchers attempted to isolate the variables that moderated this effect. Specifically, they wanted to know whether appearance enhancing behavior functions as the peacock’s tail (i.e. to impress opposite sex members), as a deer’s antlers (i.e. to impress same sex rivals) or as a lion’s manes (i.e. to impress both).

Interestingly, it turned out that the surplus effect of ovulation on sexy product choice disappeared when women were primed with attractive men or with unattractive women, which must count as evidence against the thesis that hormones may enhance efforts to impress potential mates. Even more interestingly, ovulating women chose 25% more sexy products when they had previously thought about attractive local women (as compared to
non-ovulating women). Because there was no effect at all of either seeing pictures of attractive women or thinking about attractive local women on non-ovulating participants, these results suggest that LH (or the accompanying estrogen) may activate a switch in the mating module, so that perceptions of potential rivals in the mating market stimulate efforts to impress potential mates with behavioral programs such as dressing sexy. And brand managers can of course tap into this motivation by presenting images of attractive women when branding products for women.

3.3 Selling Power to the Powerless

There is ample anecdotal evidence for thinking that branded products are used as status symbols, but what we would really like to see is more reliable real world evidence for this claim. Such evidence is given in a study by J. Lycett and R. Dunbar on the display of mobile phones – products that are heavily advertised for their status appeal – by men in an English pub.91 What they found was that men showed their mobile phone more not when more women entered the pub in absolute terms (or when there were more women relative to men), but when more men entered (and when there were more men relative to women). This suggests that at least mobile phones might be used in male-male status contests. Furthermore, if products function in status competition, then we would want to see evidence of them eliciting effects that can be predicted on the basis of models that were discussed in section 2.2. Specifically, on the basis of the model of Mazur one would expect to find that people who use reliable high status products should win more stress contests and hence have higher levels of testosterone (in either saliva or blood samples). In this regard Gad Saad and John Vongas found that when men drove an expensive Porsche sports car they produced more testosterone than when they drove a family Sedan.92 Because participants committed themselves to strictly sticking to the speed limit, this effect can reasonably be argued to be due to perceived higher status.

We also noted that the two models of status-seeking behavior in humans did not make predictions on the conditions in which people seek to acquire status signs (which is different from using status symbols in status contests), which is most relevant from the perspective of motivation and consumption. In this regard the psychologists Derek Rucker and Adam Galinky have built upon the finding that low levels of serotonin due to low power in monkeys and humans alike are deeply aversive and motivate the organism to alleviate the state of powerlessness. If consumption goods can signal rank position, they argue, then low status individuals may be particularly motivated to acquire high status goods to re-enter the status contest ring. In a series of experiments Rucker and Galinsky offered evidence for such effects in the realm of consumption.

In a first experiment, subjects were asked to recall a situation in which they either were in a position of power over another person or were the underlying party. Those in the low power condition showed a greater willingness to pay for products that were pre-tested as signaling high status (executive pen, briefcase, fur coat, and a silk tie) against products that were not (ballpoint pen, sofa, dryer, washer, and a minivan). A second experiment controlled for the possibility that the products themselves rather than the status that was associated with them accounts for the effect. This time the same product (a picture frame) was described either as conferring high status (e.g. ‘scarce’ and ‘unique’) or as conferring low status (e.g. ‘mass-produced’ and ‘available to anyone’). Again, low status subjects were more willing to pay for the high status option.93
In a third experiment, subjects in both the high and low power condition were asked how much they were willing to pay for an executive pen with the label of a prestigious private university and to indicate how much future power they expected to gain from the product. Low power subjects were not only willing to pay more, but also expected greater future power from the purchase. Fourthly, when subjects had to generate advertising slogans for a luxury product, the ones in the subordinate (low power) role made more appeals to status, while subjects in superior roles (manager) stressed performance. In order to rule out the possibility that the powerful emphasize performance only to hide their aims (of gaining more power), in another experiment Rucker and Galinsky presented high power subjects with a product that was described as high in performance but not at all stylish. If high power subjects really prefer performance, they should still prefer this product description over one in which the frame was described as stylish but of low quality. And so they did. In a sixth experiment, it was shown that high power people had less interest in visible (or conspicuous) consumption than low power people. That is, high power individuals had less preference for high-end clothing with a logo (compared to the same clothing without logo) than low power subjects.

If low power people are indeed more likely to consume high status goods, then we can make the subsequent prediction that people who belong to low status groups should on average spend more of their budget on status goods. If one is willing to assume that ethnic minorities in the United States have on average a lower position in status hierarchies, then research by the economist Kerwin Charles and colleague validates this prediction. On the basis of US data on consumption patterns, they showed that ethnic minorities spend a greater share of their income on visible consumption (clothing, jewelry and cars) than non-minority citizens. Because the share of visible goods decreases when income levels go up within a minority group, there is good reason to assume that individuals low in power indeed attempt to enhance their status by means of conspicuous consumption.

It is not immediately clear how these results relate to the models of status-seeking behavior of Mazur or Ellis and Kirkpatrick. But a reasonable speculation would be to suppose that whereas winning status contests (and the accompanying experience of high power) may induce subjects to engage in more contests with the same repertoire of status signs, losing contests may facilitate behavioral programs that seek to add status signs to the repertoire.

### 4 Social Motivations, Consumption and Branding

So far we established that there is suggestive evidence for assuming the existence of motivation modules for affiliation, mating and status rank, and that there is evidence from psychology and sociology to suggest that these modules cause consumption behavior. This is all that my argument requires, as it follows that the kind of branding (for visible products) that manages to plausibly present a product as satisfying desires that are connected to these modules will be most effective in causing sales. However, there are quite a number of assumptions involved in taking a case that is built mainly in laboratories to the level of actual branding effects in society. It would therefore be appropriate to at least outline how this extension can be falsified. To this end I have devised a number of hypotheses that use natural experiments,
set up by history, to test whether attempts to associate brands with higher-order motivations indeed correlate with higher sales.

4.1 Affiliation, Advertising and Existential Threats
With regard to group identification conditions, we must begin with the observation that we have evolved to affiliate with those who share traits with us. This first criterion is difficult to test however, since it is practically impossible to retrieve physical and psychological traits of actual consumers and then relate them to traits in real branded messages. At best, we could argue that group identification is easier with ads that feature people with average faces, clothes and values. But then, group affiliation emerges precisely then when people differ from the standard (e.g. by wearing a red t-shirt), so that it is uncertain whether this feature would work. On the other hand, when features in advertisements are too distinctive, most people will reject the features, which will hamper sales. Thus, the gain in affiliation caused by similarity of unique features will be offset by the loss in affiliation due to greater risk of conflicts of the specific values with those of the consumers. The second criterion is that we tend to identify with those who are much around us. Although it will probably be the case that advertisements featuring groups of people that are very often repeated are more likely to lead to group identification, the threshold for significant effects will in all likelihood be much higher than the level of repetition that any advertising campaign has yet achieved. But we may suppose that, all things being equal, group identification will be higher for advertisements in which groups of people are associated with a brand than for those in which a single person is associated with a brand.

If this is true, then we would expect increased sales for such brands under historical conditions that are known to lead to greater group identification. But how to isolate such historical conditions? We noted in section 2.3 that coalitional threats activate the affiliation module. Hence, if a society-wide event were identified that can reasonably be assumed to have induced existential fears across the population, then those brands that are advertised for group identification may be expected to be bought more often. As such society-wide events, I propose to select two particularly dangerous moments in the Cold War: the Berlin Crisis from June 4 to November 9 1961 and the Cuban Missile Crisis of October 1962.

**Hypothesis 1**: Sales of the brand that was advertised more consistently for group affiliation than the equivalent product will be higher during the Berlin and Cuban Missile crises than during the preceding, intermediate and following years.

4.2 Mating, Advertising and the Sex Ratio in Society
In order to test the effect of mating related branding on sales we can fill in more details of the research design. We would want to have two historical records of advertisements for brands of comparable quality (as established in blind company tests) in the same market. In this way we would control for most fluctuations in demand external to the brand image. In this regard the car manufacturers BMW and Mercedes would be a good pick because they come close in terms of quality. Next, we would need to find a condition similar to the sexual priming exercise of Griskevicius and colleagues on the level of society at large. For this purpose I propose to turn to the sex ratio in society – i.e. the amount of men in relation to reproductive women – because it influences the selection of mating strategies. That is, if there are more females than males in the mating market, power shifts towards men so that
their innate preference for short-term relations comes to shape sexual relations. Skirts, for example, turn out to be shorter if there is a shortage of men on the mating market (at least in the period 1885-1976) and a cross-cultural study that compared 185 countries found that a shortage of men consistently correlates with a higher birth rate.\textsuperscript{97} Particularly striking is the recent finding that there was a 100\% increase in extramarital births in postwar Bavaria when the sex ratio dropped to 0.6 due to male war casualties.\textsuperscript{98} As such, the motivational basis of advertising induced consumption could be tested by a simple prediction.

**Hypothesis 2**: In the 1950s, German sales of the car that turns out to have been associated in advertising with sex were higher than sales for the car that has not been advertised equivalently in the same period, compared to sales for both cars in the 1970s (when the sex ratio was more equal).

Testing the motivational basis of brands that are associated with status is trickier, because recent research suggests that female mate preferences may be – paradoxically – more demanding in the context of a low sex ratio as a defense mechanism in a short-term mating environment.\textsuperscript{99} More (cross-cultural) research is needed on this mechanism as well as male responses to it before meaningful predictions can be made.

### 4.3 Status Rank, Advertising and Foreign Occupations

A natural experiment on the role of status in the real-world effectiveness of branding would have to involve a manipulation in which a cue of low power (or high power) is experienced nationwide. Such a condition might be very rare. So much so, in fact, that the only plausible candidate condition may well be foreign occupation. We can certainly assume that this situation evokes feelings of low power in the population. The realization of defeat is likely to cause feelings of powerlessness by itself, while dominance displays by conquering soldiers will reinforce this state. A good case would be the German occupation of the Netherlands, which was swift (the conquest took only four days) so that there should have been a clear break in terms of feelings of power. (Better still might be the German occupation of France, because the humiliation was more intense in that case. But the realization of defeat may have been a more gradual affair in that case.) If the psychological work by Rucker and Galinsky is correct, this event should have activated the status rank module so that consumption for status products went up. If we could find two equivalent products where one was consistently branded for high status whereas the other was not, then we could predict that the high status brand should have higher sales (all things being equal).

**Hypothesis 3**: The product that was advertised more consistently for high status than the equivalent product will sell more during the first week of German occupation of the Netherlands in 1940 than during the same week in preceding years (relative to equivalent products that had not been advertised for status).

In order to avoid confusion I should stress that the aim of research along these lines would not be to demonstrate that consumption is driven by the male/female ratio, diplomatic crises and occupation armies. The aim would also not be to confirm the idea that different types of branding are particularly successful under these three conditions. Instead, the aim of the studies sketched above would be to confirm that the respective motivational mechanisms
indeed drive spending behavior, and only to this end extreme historical conditions are selected as natural manipulations of history.

5 Managerial Implications of the Motivational Structure of Branding

So far in this chapter I have observed that much of the brand premium is explained by visible consumption and that such behavior is driven to an important extent by higher-order social motivations. If this is true, then brand managers have an interest in crafting messages in such a way that they are most appealing as signals of mate value, status or affiliation. In this section I explore under which conditions such signals are most appealing and establish the factors that constrain this appeal. The most important such constraint will turn out to be the cognitive motivation for consistency. If the motivation for consistency were indeed a constraint on effective brand promotion, and brand managers would hence have a motive for bypassing it, then there are two reasons for thinking that the normative implications are serious. Firstly, checking for consistency is the core property of the initiatory mechanisms that secure authentic mental state formation, and bypassing these mechanisms constitutes a case of authenticity violation. Since the associative processing channels by which initiatory mechanisms are bypassed do not allow intentions to be identified, such influence would also undermine self-control. Secondly, because the motive for bypassing the consistency architecture is to more effectively appeal to an innate motivation module, the resulting violation of autonomy would count as structural.

5.1 Widening the Brand Community

A first managerial implication of the affiliation motivation is that (brand) community images that appeal to more specific values will be more effective in constructing credible signals of belongingness and create greater loyalty. Let me discuss credibility first. Although social identity research shows that we easily identify with a group, such identification is contingent on similarities (e.g. wearing the same t-shirt). This means that once a branded product spreads and is used publicly, it will get an automatic affiliation boost. That is, people who are motivated to express their belongingness will now consider the brand a more credible signal of belongingness. This means that this effect is unrelated to properties of branded messages. Such a message can of course in principle show other people displaying the brand and so simulate a virtual brand community, and this strategy is likely to have an effect. But a more effective way to turn a brand into an effective signal of belongingness to a group is to make the brand express values that a consumer holds dear. This strategy is likely to be more effective because our ancestors are more likely to have preferred signals of belongingness that indicate that the individual ascribes to the values and norms of the community. Such signals are more revealing of true commitment than signals that express similarity of dress and the like (see chapter 5).

This brings me to loyalty. If consumers of brand A signal belongingness via similarity of surface properties and those of brand B via similarity of values, then all things being equal consumers of brand B will be less likely to switch to a different community. The reason being that commitment to values is located far higher up in the goal hierarchy, so that aban-
donment would force the individual to reject more goal commitments. If either the credibility or loyalty argument is correct, then brand managers have a structural motivation – all things being equal – to be specific about the values the brand community signals.

However, all things are not equal, and that brings me to the second incentive. It states that in order for a brand community to be more successful in generating sales it will have to include a maximum number of consumers. There may be a serious trade off involved for some product categories where the wealth of consumers is more important than their number. But even if we take these cases into account, in general the maxim will still be: the more brand community members, the better. In order to expand the consumer base, a brand would have to signal values (given that these are most effective in fostering belongingness) that appeal to a maximum number of consumers. However, even if there are certain values that are shared almost universally among consumers, appealing to these values is no guarantee of success. Universal values, after all, are limited in number – if the cross-cultural work of the psychologist Schwarz is correct there are about ten. Given competition from other community brands, there would be a vicious race to exploit these universal values, and the massive appeal to them would mean that a simple reference to the universal value will no longer be distinctive. From now on, the brand that is most convincing in associating itself with a value will attract most consumers. This forces us to answer the question what makes a community value more convincing to a consumer. We can assume that this convincingness will depend on the possibility for consumers to relate to situations that express the value. However, once situations are specified, values are interpreted in advertisements and other aspects of marketing. And interpretations massively expand the amount of conflicts that the consumer can perceive with what she deems important (higher order goals, preferences, moral codes and the like). This means that the second incentive to increase the number of consumers is hampered by the rise in conflict perception that reliable community commitments bring with them.

If this is true, then we would expect to find evidence for decreases in brand preference when consumers associate a brand with an out-group that they do not like. In a number of studies Katherine White and Darren Dahl have indeed found such effects. When they had subjects allocate real brands to the categories in-group, out-group and dissociative group (i.e. disliked group), it turned out that subjects rated dissociative group brands more negatively and had less self-brand connections with them. They also found that Canadians showed a lower product preference (and choice likelihood) for brands that were associated with the United States (i.e. a dissociative reference group), especially when they identified strongly with Canada (i.e. their in-group). Interestingly, this effect was moderately mediated by their self-reported ‘self-disidentification’ with a brand (i.e. responses to questions such as ‘I would not want to identify with this brand’), which suggests that conflict occurs at the level of explicit attitudes. These effects are found both in more and in less publicly consumed products (i.e. menu choice and pen use), but they are stronger in the former case.

This leads to a third incentive, which is to remove the factors that constrain the brand group inclusion process. If there are means to communicate reliable group commitment signals that bypass conflict perception mechanism, then a firm that uses those means will be more successful in broadening its consumer base – all things being equal – than a firm that does not. We saw in chapter 2.3 that there are such means, as both implicit covariation learning and unconscious stimulus presentation bypass conflict perception. And such bypassing constitutes a violation of autonomy. To recapitulate then, the affiliation motivation is most effectively activated when the community appeal is broadest, but the consistency motivation limits the success of doing so via explicit value displays. Because the incentive to use implicit
influence results from a conflict between innate motivations, the autonomy violation is of a structural nature.

5.2 Prudent Displays of Fitness

Given that explicit – attention grabbing – messages are more effective in changing long-term attitudes (as well as action), and given that the mating module causes much of our behavior, we might be led to believe that brand managers have an incentive to make explicit claims about the fitness signaling qualities of product. There is all the more reason to think so in light of the fact that members of the opposite sex must know about the meaning of the signal for it to have an effect. So the more explicitly a mating relevant quality is advertised, one would assume, the better signal receivers know about it, and the more successful the product should be in causing sales. The reality of advertising shows a very different picture however. We do not read that a shampoo makes you look younger and that youth is attractive to men. Nor do we learn that BMW is a reliable indicator of high status, and that this is scientifically proven to be associated with more one night stands. Moreover, while there are quality certificates for every imaginable aspect of a product (its ecological footprint, consumer union quality results, health approval stamps) there is no such seal for a proven effect in attracting sex partners.

This state of affairs should not surprise us once we take the pressures against explicitness of sexual quality displays into account. For one thing, in almost all games it is an advantage if you have control over what others know about you. And so it is in the mating game. If a man wears a t-shirt that advertises him as a womanizer, then women know immediately what he is up to. This could lead to defensiveness that is not in the man’s interest. Better to be able to decide on the spot how much of his intentions women should discern. Moreover, all things being equal, what is easily available is less attractive than what is hard to get (this principle if known in social psychology as the scarcity principle), which is another good reason not to flaunt desires to attract mates too explicitly. Second, many potential consumers have strong marriage commitments. Men who return home with sunglasses that were advertised consistently for their guaranteed success in attracting mates may expect to raise some marital eyebrows at the very least. And married people, needless to say, form a substantial market. Third, there are social norms against overt displays of sexuality in advertising, especially among women. Perhaps motivated by both sexual strategies and marital peace, such norms are in some form near universal and seem to make life a whole lot more relaxed. Hence, explicitly exhibiting sexual desires is likely to generate negative feelings because of the fact that a social norm is breached as well.

The evolutionary psychologist Geoffrey Miller has entertained another possible explanation which is that promises of mate acquisition are not made explicit because they are not credible. If they were put to the test, it may be that at best small effects are found that do not seem to justify the enormous costs of the brand premium. Miller’s reason for this expectation is that the most effective mating strategies must have evolved in the Pleistocene. If this is so, then we might expect witty conversations, dancing skills and athletic performances to be much more effective time investments than taking another job in order to pay for a BMW. This argument is hard to assess though because there is hardly any research on the comparative effectiveness of mate acquisition strategies.

In any case, if any one of these arguments is correct, then explicit claims about sex are to be avoided and brand managers who want to play the sex card should do so covertly. Or, as Miller puts it more emphatically, “the product’s nominal functions, features, specifications,
novelties, popularity, and branding must occupy the consumer's conscious attention, while the promise of signaling status and sex appeal must penetrate the unconscious as silently and unaccountably as a stealth bomber”. Indeed, unconscious stimuli would do the job, but so would evaluative conditioning (as long as the relevant conflicts are not perceived by both the consumer and the target audience that is exposed to future brand signaling). Hence, the structure of marriage, the rules of the mate acquisition game and social norms all press brand message into formats that violate autonomy. Although they are not as immutable as innate motivational modules, these normative structures are stable enough to think of the autonomy violation resulting from implicit sex appeals as structural.

5.3 Status Appeals That Fit
If brand managers want to present their brands as convincing signals of high status, then they must associate their product with signs that indicate a person’s position in one of society’s rank order systems. The design and branding of SUVs for example, is crafted in such a way as to signal that the driver is dominant. SUVs have darkened windows, are often black and advertised as larger than large. But most status appeals in branding are more subtle and involve associations with sign systems that bestow prestige on people, and these systems are messy: there is no single rank order in society, signs of prestige are not the same for every group, and especially cultural prestige is historically fluctuating.

Admittedly, status displays are credible signals of one’s position in a cultural status field in part because they are costly: it takes effort to acquire the information about the latest trends. In line with this expectation a number of experiments conducted by Jonah Berger and Morgan Ward have shown that fashion students have a preference for less visible brand signals even though these increase the risk of misidentification. And they have this preference especially if the product is visible (e.g. clothing) and if they are asked to think of walking with the product (e.g. a branded bag) through the street. However, the authors note that inconspicuousness is rewarding only for high-end products that cannot continuously be updated or renewed. Inconspicuous status signaling would in these cases be an efficient means to retain the perceived exclusivity of the brand. For most products though, the gradual loss of exclusivity will be the most efficient way to soak in as many consumers as possible. Thus, for most brands which appeal to status it will be more efficient to show advertisements in which high status people use the brand and to display the brand in contexts (e.g. villas and swimming pools) or objects (e.g. suits) that are associated with high status.

If this is correct, then we may assume that the process of convincing consumers that a brand signals status will cause a lot of conflict perception. Some consumers may feel aversion towards the high status person that promotes the brand, some may dislike luxurious contexts and some may dislike objects like suits. Displaying the brand as a status sign would also mean displaying potentially disliked traits. If this is true, then we should find that when subjects choose among brands that are rated as high in status, they should prefer the ones that are rated as more self-congruent. This is exactly what Aran O'Cass and Hmily Frost found in their study of attitudes towards the brands Calvin Klein, Target, Oakley and Polaroid. As in the case of affiliation, the interaction with the innate motivation for consistency presses appeals to the status motivation in the direction of conflict evasion. And as both motivations are equally stable, the resulting autonomy violation is equally structural.

Given the oblique character of my argument it is easy to lose sight of which claims have been supported and which remain to be proven. The strategy was to argue that if we know
the motivational bases for consumption, we can infer to which desires brand images are responsive. Brand value that depends on public signaling was argued to be motivated by the three partly overlapping modules for affiliation, mating and status rank, and evidence was offered for the assumption that consumption is caused by these motivations. It was then argued that brand managers who want to appeal to these motivations have an incentive to devise messages in such a way that they violate autonomy. Specifically, affiliation, mating and status appeals are bound by structural constraints, caused either by other innate motivations or by stable social norms, which lead to conflict perception. This gives brand managers the incentive to bypass conflict perception mechanisms and hence to violate autonomy. Thus, the entire argument has so far been about the structural forces that are expected to push branding in the direction of autonomy violation. In the next chapter the mechanics of the autonomy violation itself will be considered in order to see whether it actually occurs.
4 How Branding Structurally Violates Minimal Liberal Autonomy

What I have argued so far is that branded messages that violate autonomy are ones that take the form of evaluative conditioning or unconscious influence or are adopted after the induction of negative affect. In order to establish whether these techniques are used widely I took into account structural pressures for using them that emanate from social norms, economic conditions and psychological conditions. For consumer action (at the points of purchase and consumption) the pressure was straightforward. Unconscious goal activation by means of priming was found to be sufficiently effective in the short term, so that one should find ample evidence of its effectiveness in the realm of consumption (section 1).

For mental state formation, matters turned out to be more complicated because unconscious influence and evaluative conditioning mainly affect implicit attitudes, which are not as effective in influencing behavior as explicit attitudes (all things being equal). I argued however that when the use of brand images evokes conflict, implicit brand attitudes resulting from evaluative conditioning are more effective than explicit ones (in whatever way they are acquired), and I specified the conditions under which such conflicts were expected to emerge. If these arguments are correct, one would expect to find empirical evidence for attitude change brought about by branding in ways that violate autonomy both from laboratory experiments and from the historical record of branding. More specifically, one should find in laboratory experiments that for messages that activate the affiliation, status or mating module implicit attitude change is often more effective than explicit attitude change. And one should find in the historical record that advertisements that appeal to these same modules make more use of rhetorical techniques that are known to foster evaluative conditioning (or implicit attitude formation by means of associative processing more generally).

Unfortunately, no such research exists to date, so that I will restrict myself to more limited arguments. That is, I will give evidence merely of the effectiveness of mainly evaluative conditioning in changing implicit attitudes for brands and for the effectiveness of such attitudes in triggering purchasing behavior. I also identify a set of rhetorical techniques that facilitate associative processing and specify under which conditions their application in advertising messages is evidence of autonomy violation, so that my theoretical argument can be falsified (section 2). I move on to consider the conditions that moderate the need for state intervention. In this regard I argue that the effects of autonomy violating branding are sufficiently central to warrant state action (section 3), that the requisite intentions on the part of brand managers are retrievable (section 4) and that there are no adequate opportunities for consumers to consent to unconscious influence in branding (section 5).
1 Branding, Autonomy and Consumer Action

In chapter 2 I identified three main moderators of the autonomy of actions. The authenticity of actions is moderated by cognitive load and negative affect, while the amount of self-control of actions is affected also by the unconscious triggering of behavior. For these factors to influence consumers in a relevant way (i.e. a way that enhances sales), they must have an influence either at the point of purchase or at the point of consumption. With regard to cognitive load and negative affect little is known about their effect on consumers in any of these realms. Moreover, although both factors could theoretically enhance sales, it is not obvious that such tactics would work in practice. Negative affect is a dangerous tool to employ, because it is also associated with negative attitudes towards attitude objects. Cognitive load is likewise not without risk because the deployment of cognitive load by salespersons may trigger reactance in consumers. Hence, in the absence of ecologically valid studies, no conclusions can yet be drawn about the proliferation and workings of any one of these methods.

With regard to unconscious influence on consumer actions, by contrast, there is by now a significant body of research. Unconsciously triggered behavior undermines self-control, but this need not be the case for authenticity because self-relevance and congruence with the self may be regulated automatically. For this reason I will limit the current discussion to the effects of unconscious influence on self-control. There are three such routes of influence that are facilitated by different moderators and which are all problematic from the point of view of self-control (see figure 4.1). One route operates through a direct perception-behavior link (paths 1,2,7), another by means of the activation of trait concepts (paths 3,4,7) and a third route via the activation of goals (paths 5,6,7). The first route of mimicry is an ancient mechanism that activates behavioral schemes directly and facilitates liking for others.

Trait and goal activation are more complicated. In trait activation (or semantic activation), associative structures are activated if they are congruent with the self-concept and may activate behavior in the very short term. In goal activation, a goal of the so called ought self or ideal self is activated, and may be kept active as long as no feedback is received about its satisfaction. Moreover, a discrepancy between the ought or ideal self and the active self increases the motivation of the unconsciously primed goal (where the prime itself is conscious but its effect in causing behavior is not). Below I discuss studies concerning the effect of these three routes at the point of purchase (1.1) and the point of consumption (1.2) respectively.

1.1 Unconscious Point of Purchase Influence

The image haunting Dennett in his discussion of self-control is that of the Sphex (a type of wasp), whose rigid program lends itself to easy manipulations by pitiless biologists. The Sphex reproduces herself by building a burrow, dragging a paralyzed insect inside and laying her eggs alongside it. In doing so, the Sphex displays an interesting routine. She is programmed to throw down the insect at the entrance of the burrow and to walk around in circles in order to find out whether there is danger lurking in the dark. If all is clear, she picks up the insect and drags it inside. However, if the insect is pulled back a bit by a human observer while the Sphex is running around in circles – having almost finished inspecting
her burrow – and laid down outside of the hole, the routine repeats itself: the insect is pulled up to the threshold of the burrow and the Sphex starts a new inspection round. This process has reportedly been repeated up to forty times. Such cases leave one wondering to which extent human beings can be manipulated in a similar but of course more sophisticated way (either by aliens with a superior intelligence or by fellow human beings with superior knowledge). Surely, we will take note of repeating an action forty times. But if behavior can be unconsciously triggered – below the radar of our far more impressive cognitive defenses – then it can legitimately be asked to what extent human beings entering shops are like Sphexes entering their burrows. In this section I discuss the three routes by which such manipulation may occur. I first discuss mimicry of other consumers and of salespeople and then take goal and trait activation together (as it is not clear from all studies which one is activated) and discuss their application in terms of auditory, olfactory and visual cues.

Mimicry
There is by now an extensive literature documenting the effect of mimicry of body posture, speech patterns, syntax, accent, facial expression and hand movements on rapport. It turns out that the causality of this relation goes both ways. When we like somebody, we start to mimic that person, and we also start to like a person if she mimics us. The latter effect has been documented in real world environments. That is, it has been demonstrated that a wait-
ress who repeats orders by customers in fact earns more money from tips than if she paraphrases orders. And people whose body posture is being mimicked are also more likely to give to charity.\(^5\) We may expect two ways in which mimicry can influence consumer decisions in shops: consumers may mimic others, and they may be influenced by others mimicking them. That is, consumers may be influenced by product choices by others in the shop, and they could be more likely to purchase products that are promoted by salespeople who mimic them.

In a series of experiments, Robin Tanner and colleagues tested both of these possibilities.\(^6\) Evidence that consumers may mimic others in making decisions comes from an experiment on preferences for crackers. In it, participants were presumably to do a test on recall of advertisements. To this end, they were shown a video of a person who was said to be an earlier participant (but who was in fact a confederate) and read advertisements to them. She was told, in the video, by an experimenter that she could eat from two bowls that were placed in front of her. One bowl contained animal crackers and the other contained giant goldfish crackers, while the confederate – in two conditions – ate consistently from only one of the bowls. A second manipulation was to allow one group of participants who watched the video to eat from two similar bowls, while giving a second group no such option. Afterwards, participants were asked to rate – in a presumable unrelated task – product preferences for products which included both types of crackers. If mimicry mediates consumer preferences, Tanner and colleagues reasoned, then enhanced preference for one of the crackers among those who could eat from the bowls should be correlated with their mimicking the cracker consumption of the confederate in the video. And that was indeed the case.

In a second experiment, evidence was provided for thinking that mimicry by a salesperson can influence brand preference. Participants were divided into a mimic group and a no-mimic group and supposedly met a salesperson who promoted an imaginary new soft drink called Vigor. In both conditions the salesperson conducted a tightly scripted interview about soft drink use, so that as little room as possible was left for spontaneous interactions that could interfere with the design. In the mimic condition, the salesman mimicked body movements with a delay of 1-2 seconds (and for the opposite side of the body, hence the name mirroring). In the no-mimic conditions by contrast, he reversed the procedure, so that when a participant bowed over the table, he would sit straight up. In a survey following the interview it was found that both the liking for the brand as well as purchase and physical consumption intention were significantly higher in the mimic-condition.

Taken together, these studies show that in laboratory settings consumers can be induced to develop a preference for a brand by means of mimicry. Field studies will have to establish whether these techniques will work in practice. It must furthermore be noted that both cases of mimicry induction involved considerable amounts of time. With regard to consumers mimicking others, the application for brand managers would be to hire actors to ostensibly buy their products, which will typically be noticed only briefly by consumers. If mimicry is to boost sales then, it is to be proven that short observations also have such effects (as this is what happens in actual shop environments). With regard to consumers being mimicked, the present research could explain persuasion in longer salesperson-customer interaction such as in car sales.\(^7\) Such cases would also fall within the scope of my present argument, because care salesmen are often closely aligned with the brand manufacturer and can hence be said to be part of the organizational complex that manages a brand. For both kinds of mimicry the practical relevance is mediated by the risk of consumers finding out about the practice. Such effects have as of yet not been studied however.\(^8\)
Goal and Trait Activation

Studies investigating the more complex unconscious goal and trait activation can be organized according to the perceptual channel through which the influence stimuli enter the brain. As such, I will present experiments that demonstrate such effects for auditory, olfactory and visual primes. The perhaps most studied unconscious prime in retail environments is music. It is uncontested that music can have a beneficial impact on sales in general. Adrian North and David Hargreaves for example found that playing classical or pop music enhanced sales in a cafeteria more than did silence or playing pleasant music. More interestingly, in another study North and colleagues found that playing German music in a store increased sales of German wines, while playing French music increased sales of French wine. Afterwards, consumers displayed no awareness of the relation between hearing such music and their wine purchases. The temporal distance between the auditory prime and the purchase decision was not manipulated however (which could have been done by playing the music at the entrance of the store and measuring the average traveling time to the wine section), so that it is unclear whether the effect was due to unconscious goal pursuit or unconscious trait activation.

There is evidence for thinking that smell can activate behavior unconsciously as well. Rob Holland and colleagues mixed some citrus scent into a bucket of lukewarm water that was covertly placed in the laboratory. When subjects were asked to list activities they wanted to engage in, they were more likely to list cleaning activities if the bucket was in the room. In a follow up study it was found that subjects who smelled the citrus flavor were also more likely to clean up their crumbs after eating a cookie (which was established by counting the crumbs participants left over). In both studies participants were not aware of either the bucket or the citrus smell. A number of studies have furthermore demonstrated that smells that are rated as pleasant in general enhance sales. Taken together, this evidence suggests that brand stores that use scents to unconsciously activate goals and traits related to brand purchasing behavior may actually succeed in doing so.

The most sophisticated research in priming consumers has been conducted with regard to visual cues. Specifically, Tanya Chartrand and colleagues have made a strong case for thinking that such primes can unconsciously activate goals (rather than merely activate traits). In an experiment they presented a scrambled-sentence task in which subjects were to compose four-word sentences out of five scrambled words. For one group the spare fifth word was occupied by words related to status and in another group by words related to frugality. Subjects were then presented with a text in which a situation was sketched in which they had to choose between buying one pair of Nike socks (for 5.25 dollar) or two pairs of cheaper Hanes socks (for 6 dollar). As predicted, more subjects chose the Nike option in the status priming condition (48%) than in the frugality priming condition (19.2%) while they reported no awareness of a relation between the scrambled sentence task and the product choice task.

In follow up studies it was shown that the effect was due to goal activation and not trait activation because the effect grew stronger after a time delay (i.e. 3 minutes versus 8 minutes). More evidence for this interpretation emerged from a study in which one group primed with status made a hypothetical choice for socks whereas another group primed with status made a real choice for socks (i.e. they could actually take them home). Afterwards, subjects in both groups moved to another room in which they were told that a lottery was organized for participants in the experiment, and that they were to make a choice out of two options in case they would win. In one option they could choose a high status watch and 25 dollars in cash, and in the other option they could choose a lower status watch and about 75
dollar in cash. Goal satisfaction turned out to make a real difference. Whereas status primed subjects were 43% more likely to choose the high status option in the lottery task against thrift primed subjects, this effect disappeared completely for status primed subjects who had made a real choice for socks before (i.e. they could take them home). Because satisfaction of a goal has been shown to lead to its deactivation, this is strong evidence for interpreting the priming exercises as activating goals and not traits.

So far the studies are relevant insofar as they indicate the possibility for retailers to place status displays in their shops to activate unconscious status consumption goals. In a final study, Chartrand and colleagues proposed that brands themselves may serve the same purpose, since they have already been associated with status (as in the case of Tiffany, Neiman Marcus and Nordstrom) or thrift (for which were selected Wall-Mart, Kmart and Dollar Store). Store brand names of both groups were subliminally flashed repeatedly on a computer screen (for 60 milliseconds with a backward mask of 100 milliseconds) after which subjects were to choose between high and low status socks and microwaves. As predicted, subliminally primed store brands that are associated with status correlated more strongly with choices for high status options than store brands that are associated with thrift.

Looking back on the evidence of effects of mimicry as well as unconscious goal and trait activation on consumer choice we can return to the question of whether consumers in shops are controlled like Sphexes. There is some evidence that they are. Mirroring behavior by covert and overt store employees, smells, music and a range of visual cues can all be employed by store managers to trigger purchasing behavior outside of the awareness of consumers. However, human beings are not Sphexes. We have superior learning and reasoning capacities as well as the capacity to act upon such reasoning. Although consumers today may not know the nuts and bolts of unconscious in-store persuasion, it can be argued that they could potentially learn about these effects. Whether such an education project is realistic is a different matter however. For now, I conclude with the caveat that since the scope of my overall argument is limited to the effect of branding, these studies are relevant only with regard to effects of store brands and atmospheric effects in shops that sell a single brand exclusively (such as certain car retail stores or clothing stores).

1.2 Unconscious Point of Consumption Influence
The second realm in which branding may undermine the autonomy of consumer action is the point at which products are consumed. Because most products are consumed at home, the best shot at unconsciously priming people to consume more of a brand is via mass media. In one experiment, Jennifer Harris and colleagues tested the effect of television food advertising on eating behavior of both children and adults. In the first experiment children watched a 14 minutes episode of ‘Disney’s Recess’ which was interrupted by two breaks with a total of four 30 second real-world commercials. The manipulation was that one group saw commercials for unhealthy snack food, while another group saw children’s commercials for non-food products. As they watched, children were given goldfish crackers. As predicted, those who watched the food advertisements – which were unrelated to the goldfish crackers – ate more (8.8 grams to be precise). In a second experiment, adults watched a 16 minutes long episode of a comedy show with 4 minutes of commercials with different content in different conditions. In one condition four commercials for snack foods (with hedonic appeal) were added, in another condition four commercials for nutritious foods were inserted, and in a control group no food advertisements were included at all. Participants were told that the study investigated the
effect of TV watching on mood, and that commercials had not been cut out to make the TV watching as realistic as possible. In a presumably unrelated follow up study in another room, subjects then rated a range of healthy and not so healthy foods and were told that they could eat as much as they liked. After controlling for gender and restrained eaters (i.e. people who are dieting, as they are more vulnerable to food cues) it turned out that subjects in the snack commercial condition ate far more than those in the non-food condition and the healthy food condition. In both studies, subjects had no awareness of a relation between the food commercials and their food consumption. Further evidence for the explanation of the effect in terms of an automatic mechanism is the fact that none of the foods that were actually consumed was mentioned in the commercials. Snack food commercials associate eating with fun and enjoyment, which appears to activate eating behavior either by activating a goal or a trait. Which one of these mechanisms facilitated the effect was not investigated however.

In any case, because the effect of the hedonic snack advertisements appears to boost eating behavior in general, the technique is a very crude one from a brand managerial perspective. It is moreover questionable whether the steering of unconsciously activated behavior can be further refined in light of the finding that health food commercials did not increase the consumption of healthy foods over the other two conditions. That said, the unconscious effect of commercials on consumption is real and advantageous at least for market leaders in a food category.

2 How Branding Changes Mental States

With regard to mental state formation there is no need at this stage to analyze the consequences for self-control and authenticity separately. As we have seen, mental state acquisition induced by negative affect and evaluative conditioning is a potential violation of autonomy from both perspectives. Relevant differences will surface only when we consider retrievable intentions in section 4. What we would want to see here is evidence from the laboratory for thinking that the processing mechanisms that were said to bypass cognitive defense structures can be used successfully to change attitudes towards brands (2.1).

Moreover, for such instances of autonomy violation to be relevant on the present analysis, they have to stand in a structural relation to stable features of the branding environment. The most important such feature was said to be sales, so that we also need evidence for thinking that brand attitudes that are the result of processing that bypasses defense mechanisms are effective in triggering sales (2.2). It could be argued that these laboratory experiments tell us little about the messy real world in which branding occurs. In response to this concern I give suggestive evidence from the historical record of advertising for thinking that the normatively relevant type of evaluative conditioning has been deployed on a large scale, and specify how this claim could be tested more robustly (2.3).

2.1 The Mechanics of Autonomy Violating Attitude Change in Branding

Because there is no research on introjection of purchasing goals in branding, I will focus here on evaluative conditioning. In chapter 2 I established that there is a mechanism that can change attitudes out of the reach of both awareness and consistency checks (i.e. evalua-
tive and non-evaluative covariation learning). In this section my aim is to give evidence for the effectiveness of these mechanisms in changing attitudes towards both new and well known brands. Although there is no obvious reason for thinking that brands are attitude objects of a type that is resistant to conditioning effects, it is better to rule out this possibility. One would moreover want to see evidence of conditioning effects of brands that are paired with USi that are commonly used in advertising.

With regard to attitudes towards new brands, Steven Sweldens and colleagues compared the effects of misattribution and stimulus-stimulus evaluative conditioning (EC).23 In misattribution trials, the conditioned stimulus (CS) was simultaneously paired with several unconditioned stimuli (USi) of the same valence, so that the affect is more likely to be attached to the CS directly. In stimulus-stimulus trials the CS was sequentially paired with the same US, so that the affect was transferred to the CS indirectly (via the US). Pairing the CS simultaneously with the same positive US (e.g. pictures of people having fun) was reasoned to fall somewhere in between these procedures. In two experiments using only explicit attitude measures it was shown that American students thought more positively about Belgian beers after both types of EC procedures, but that the revaluation of the relevant US wiped out the affect transfer only for sequential-same trials. That is, when the pictures of happy people were associated with felonies after the conditioning trials, this affected the evaluation of the beer brand only if it had been associated with the same happy face sequentially.

Moreover, awareness of CS and US appearing together only affected students in the sequential-same condition. In another study it was shown that when students were asked after the conditioning trials to learn about which breweries produce which Belgian beer, the effects on sequential-same EC were wiped out, but not those of simultaneous-different EC. Again, awareness did not affect responses in the simultaneous–different condition in this experiment. Finally, and more interestingly, when students were told after the conditioning trials that in rating the beers they should not be affected by the pictures with which the beers had just been associated, only participants in the sequential-same condition were able to correct their explicit attitudes. But even they did still show a conditioning effect, which the authors take to suggest that the mechanisms underlying such conditioning may not be entirely propositional.24

Another set of studies changed the USi with which a new CS (i.e. a brand) was associated. Instead of pictures of people having fun, taken from the International Affective Picture System (which are typically used in EC), Madelijn Strick and colleagues used humorous cartoons (as well as only implicit measurements, taken by an affective priming task). In trials that were far more realistic than the typical pairing of stimuli against a black background used in typical EC trials, 20 humorous cartoons were in one condition inserted in a magazine of which two pages were shown on a screen for ten seconds. With each page turn, subjects were asked to focus on the object with yellow lines around it, which was either a cartoon or a filler. One group saw humorous cartoon and another group saw cartoons that had been rated as not funny. Moreover, some brands were consistently shown on pages with humorous ads and others on pages with nonhumorous ads. It was shown that the priming of both an energy drink and a ballpoint brand that had been associated with humor facilitated significantly shorter response times in later affective priming tasks that involved the words beautiful and good. Interestingly, participants showed not only no awareness of the conditioning effect, but brands that had been associated with humor were also less well recognized.25 Because both funny people and humor figure prominently in advertising, it is plausible that EC mechanisms facilitate real world effects of advertising.
So far we have considered conditioning trials for brands that were new to subjects. This represents of course a small portion of the total amount of branding. We would therefore want to have evidence for effects of evaluative conditioning on brands with which subjects are highly familiar. Such evidence has been provided by Brian Gibson in an experiment in which attitudes for Coca Cola and Pepsi were changed by conditioning trials. In a pretest, subjects were asked – among many other tasks – to rate their preferences for Pepsi and Coke. Respondents with attitudes at the extreme ends and the center of the distribution were asked to return. All three groups were then confronted with five rounds in which they saw 86 diverse images that were paired with other images or words, and were instructed to hit a key whenever a certain soft drink appeared (this was always a different soft drink from Pepsi or Coke). Interwoven in these trials were four images of Pepsi paired in one condition with positive images and in another condition with negative images, as well as four images of Coke which were paired with USI of the opposite valence. Gibson reasoned that none of the three groups should show any explicit attitude change, because of the strong explicit attitude networks of the brands. However, with regard to implicit attitude measures, he found that although lovers and haters of a brand were not affected by either positive or negative stimuli, those in the neutral condition were.

As an explanation for this effect Gibson argues that evaluative conditioning can activate an implicit associative network but that this is only possible if there are both positive and negative associations. For the neutral group these associations were assumed to reside in a state of balance and so could be tipped to one side by the conditioning trials. An explanation in terms of direct affect transfer would be that those with strong positive or negative attitudes had strong affective implicit attitudes, whereas those in a neutral state did not. So on this account the newly attached affect could only have an impact if there were no strong positive or negative affective relations to be overcome. Either way, the process is a violation of autonomy because Gibson found that the level of awareness was inversely related to the effectiveness of the conditioning trials.

2.2 How Autonomy Violating Brand Attitudes Cause Sales

If autonomy violation by unconscious conditioning were to be structural, then the implicit attitudes it brings about should predict sales over and above effects of explicit attitudes. In a recent overview, Malte Friese and colleagues considered eleven studies that investigated this relation. The common finding of these studies on the consumption of fair trade products, yogurt, fruit, vegetables, restaurant meals, Coke and Pepsi was that implicit attitudes have at best a very small effect on consumption choices. That is, if subjects are to choose in the laboratory any one of these products in a survey or in the form of a real pick, then implicit attitudes do not substantially explain why they do so. However, the authors note that such results are precisely what one would expect to find if dual system theories of decision making are correct. That is, implicit attitudes are not expected to predict choices when subjects have all the time in the world to make a decision. This is also what one would expect from the nature of the measures in the first place. Survey responses are typically given without time constraints, while implicit measures tap small differences in very quick reaction times. It is therefore only natural to expect that when consumers make decisions under circumstances that reflect filling in a survey, explicit attitudes will better predict such choices. By corollary, under impaired processing conditions (which I will specify in a moment) the effect of implicit attitudes should be larger.
In their overview, Friese and colleagues report on seven studies – mostly carried out by the authors themselves – that have manipulated conditions to facilitate controlled decision making versus impulsive decision making. Which decision-making system is activated depends on four factors, which have all been put to the test. First, time pressure should facilitate impulsive decision making. One study measured subject's implicit and explicit attitudes towards two sets of similar groceries where one was branded and the other set was not. Participants were then divided into a group with congruent implicit and explicit preferences for one of the sets, and a group with incongruent implicit and explicit attitudes. When asked to choose a set and to take as much time as they wanted, 80% of the choices of subjects in both conditions were explained on the basis of their explicit attitudes. However, for subjects who had to choose within five seconds it turned out that for those with incongruent attitudes – which is the group in which the effect of implicit attitudes can be observed – 60% of choices were predicted by implicit attitudes.

Second, cognitive capacity should facilitate deliberative processing. To test this assumption, another experiment measured implicit and explicit attitudes towards fruit and chocolate and then had participants make a decision without time constraints. However, in one condition subjects meanwhile had to memorize an eight digit number (and thus had little spare capacity to make a decision) while in another condition subjects memorized only one digit (and thus had a lot of capacity left). As predicted, choices under high cognitive load were hardly related to explicit attitudes and significantly related to implicit attitudes. The impact of cognitive load on product choice has also been studied for brands. In his experiment on evaluative conditioning of attitudes for Coke and Pepsi, Gibson ended by an on screen product choice test in which he asked half of the participants to memorize an eight digit number. As predicted, under high cognitive load implicit brand-related attitudes were predictive of brand preference.

At least for new brands there is evidence – from a study by Gibson and Redker on new beer brands which were paired with the same USi (country music) – that even choices absent cognitive load might be related to implicit attitude change (resulting from EC trials). Two weeks in advance of the experiment, Gibson and Redker had subjects rate their preference for country music after which those at the extreme ends were called back and asked to participate in the main experiment. Here they saw a number of web advertisements on a computer screen among which were ads for the beer brands Fitz’s and Sparky’s (which are local beers from a different region than the one in which the experiment took place) accompanied with country music. As predicted, those who hated country music rated the beer in question less favorably on both explicit and implicit measures and those who loved country music showed the opposite pattern. More importantly, when participants at the end of the experiment were told that they had earned a beer and could choose between Fitz’s and Sparky’s on their screens (without any kind of pressure being exerted on them), implicit attitudes predicted unique variance in choice compared to explicit attitudes. More evidence for this thesis comes from the experiment by Strick and colleagues on humor in advertising. They found that implicit attitudes that resulted from simultaneous-different EC trials predicted choice for ballpoints and energy drinks.

Third and fourth, both the depletion of self-control and alcohol use should facilitate impulsive behavior. As I noted in chapter 2.2, there is evidence for thinking that self-control functions like a muscle because, among other things, it can be depleted. Under these conditions implicit attitudes should better predict behavior. According to alcohol myopia theory, alcohol impairs the allocation of attention to goals. If system 2 processes require more attention resources, then system 1 processes should more effectively influence behavior in condi-
tions of alcohol consumption. Both predictions were tested only for physical consumption. Depletion of self-control (by the suppression of emotion after watching a film clip) led to a greater predictive value of implicit attitudes towards potato chips with regard the amount of chips eaten on a subsequent taste test. And vodka consumption enhanced the predictive power of implicit attitudes towards candy which were thereafter consumed.35 The latter two studies are of course not evidence of purchasing behavior, but increased consumption at home is probably correlated with future purchases.

If it is true that implicit attitudes are predictive of sales under time pressure, cognitive load, self-control depletion and alcohol use, then we must go on to ask whether such circumstances reflect actual shopping conditions to a significant degree. If such impulse facilitating conditions are a rare phenomenon, then implicit attitude change will not cause many sales. And that would mean that there is also no structural incentive on the part of brand managers to bring it about. In this regard there is only circumstantial evidence, but this evidence points in the direction of a significant role for impulsive behavior in shopping. In US supermarkets 62% of purchases are impulsive – defined as made up on the spot – with percentages up to 80% in certain product categories.36 Impulsive purchases are less likely for cars and houses and more likely for food, with consumer electronics and the like located somewhere in between. Moreover, Friese and colleagues note that conditions which facilitate impulsive action occur chronically. When moving through the supermarket we often think about our next meeting, had a beer, make phone calls, want to catch a train, are under pressure from a waiter or do our shopping at the end of the day when our self-control resources have been depleted.37

2.3 Historical Evidence of Autonomy Violation in Brand Attitude Change

So far, my case for thinking that successful branding structurally involves processing that undermines autonomy has relied on laboratory work. This case would be considerably stronger if I were able to demonstrate that such processing is correlated with enhanced sales in the real world. After all, there might be sources of influence that annul the effects of conditioning. In order to make such an argument, one would want to identify properties of advertisements (or combinations of such properties) that signal autonomy violation in combination with specified conditions. That is, if advertisements that use evaluative conditioning are more effective in causing sales for messages that are likely to induce conflict, then this is evidence for autonomy violating mental state formation.38 One way to test these hypotheses would be to manipulate properties of advertisements in a split-cable/customer card setting, but such a research project would be very expensive. I will therefore outline another method, which relies on the historical records of advertisements and sales and is therefore easier to test. It asks whether autonomy violating properties of advertisements are correlated with higher sales (than sales for comparable products in the same market) under conditions that are assumed to favor their effectiveness.

A problem with testing whether evaluative conditioning occurs in advertising is that stimuli appear together everywhere in advertisements.39 In order to make its occurrence testable all the same, I will in the first subsection draw on research which suggests that the processing of metaphors and puns is of the associative type (as in evaluative conditioning). Studies that quantify the occurrence of metaphors and puns in US text advertisements shows that they are used in 45% of text ads, while other studies show that this percentage has gone
up steadily (at least with regard to US magazine advertisements) during the past half century. If evaluative conditioning were the only plausible effect of metaphors and puns, then we would have a very strong case for thinking that it is used widely in branding. But it is not. In the second subsection I discuss the main rival account that explains the success of tropes equally well on the basis of the enhanced elaboration it evokes. In the final subsection, I will rely upon my theoretical framework developed in chapter 3 to specify the conditions under which the two theories about metaphors make contradictory predictions.

The Associative Processing Account of Tropes
Although distinctions between rhetorical figures are not wholly uncontested, there is wide agreement on the distinction between schemes and tropes. That is, it is agreed that schemes involve an excess of regularity (e.g. rhyme, alliteration) whereas tropes are characterized by irregularity. Within the latter category a distinction is made between substitution tropes (e.g. metonymy, hyperbole) and destabilization tropes (metaphor, pun, irony and paradox). All variants of the latter category involve the entertainment of multiple interpretations at a time, and that is the feature that is presently relevant. For if several associations in relation to an attitude object are presumably spontaneously generated, then the co-occurrence of these associations with the attitude object may facilitate conditioning effects by means of associative processing. There are two relevant lines of research that support such an interpretation.

First, there is work on the generation of associations for advertising metaphors by Edward McQuarrie and Barabara Philips. Following pragmatic theories in linguistics, these authors assume that all interpretation relies on the generation of inferences about stimuli. For some stimuli there will be wide agreement in a culture on which inferences are appropriate (i.e. strong inferences). But when there is no such agreement, the mind will generate a great number of inferences until it has found an appropriate fit (i.e. weak inferences). Metaphors by definition involve the latter process. In order to test this claim, McQuarrie and Philips relied on response times in order to infer whether weak inferences had been generated or not. The basic assumption of this method is that inferences that are generated at the time at which the metaphor was processed will lead to shorter response times when people are asked to assert these inferences later (because retrieval is on average faster than online generation).

Thus, subjects were shown a number of advertising metaphors and asked to focus on the multiple meanings the producers of the advertisements might have intended. They were then asked to assert or reject possible meanings devised by the researchers in advance. On the basis of pre-testing, they had found that certain meanings are extremely unlikely, and including them arbitrarily in the set avoided automatic assertions on the part of subjects. Now, if assertions of proposed meanings of the metaphor took less response time than rejections, this was interpreted as evidence for spontaneous generation at the time of initial metaphor processing. The reasoning behind this assumption is that if spontaneous interpretations are formed, then assertions should be made quicker (these can now be retrieved instead of generated online) while rejections should take more time (because more stored interpretations are scanned). Such an effect was found for visual metaphors (without anchoring). There has as of yet not been any follow up studies on the way in which such associations influence the evaluation of the attitude object, which must involve some form of linking and may or may not involve autonomy violation.

A second line of research by Claudiu Dimofte and Richard Yalch uses implicit measures to suggest that much more interpretations are spontaneously generated by destabilizing
tropes besides the interpretations of visual metaphors that McQuarrie and Philips uncovered. They moreover measured the effect of these interpretations on brand evaluation and the way in which this effect is mediated. The main tool of Dimofte and Yalch is their Secondary Meaning Access via the Automatic Route Test (SMAART). In this test, subjects are asked to assess the literal truth of a range of sentences (e.g. cars are fast). Some of these sentences are metaphors (e.g. cars are snails) and some sentences are just nonsense (e.g. roses are insects). When subjects show longer response times for metaphorical sentences (but not for nonsense sentences) the effect is supposed to be caused by automatic generation of secondary interpretations. If this tool is valid, then one can suppose that subjects with a high SMAART score should be better at grasping puns in advertising than those with a low score. Thus, by presenting a pun with a negative secondary meaning, high SMAART subjects should rate the brand in question less favorably.

For their study Dimofte and Yalch used the slogan ‘raising the bar’ for the brand Cingular (which offered an exclusive wireless phone service). The positive primary meaning of this pun was that the technology of the product was of a particularly good and innovative quality, while the negative secondary meaning was that consumers may not have been able to meet the credit rating conditions after the ‘bar was raised’. On explicit ratings of the brand after exposure to the pun (accompanied by a powerpoint presentation) subjects with high SMAART scores evaluated the brand, as predicted, less favorably than low SMAART subjects. Even more revealing were tests of the process that mediated this outcome. For what Dimofte and Yalch found was that high SMAART subjects showed evidence of grasping the secondary negative meaning only on an Implicit Association Test and not when asked to rate explicitly how easy it would be for them to acquire the product. Hence, it must have been associative and not propositional processing that caused the negative overall evaluation of the Cingular brand. And such processing involves autonomy violation.

In sum, both the spontaneous generation of meanings for visual metaphors as well as puns lend themselves to evaluative conditioning. When visual metaphors are processed, multiple meanings are generated which can be recalled. Such meanings may have a strong valence and so release positive or negative affect which can then by means of misattribution or by associative linking be attached to the attitude object (i.e. the brand). In other words, the meanings that are spontaneously generated can function as USi that condition a CS. Moreover, if the negative secondary meaning of a pun influences the evaluation of an attitude object and these negative meanings can only be retrieved on implicit measures, then this must be due to some associative link between the negative evaluation and the attitude object. However, as I mentioned earlier, we should be careful in explaining the success of tropes in advertising in relation to evaluative conditioning effects, because of research that suggests an alternative explanation.

The Enhanced Elaboration Account of Tropes
A second explanation of the occurrence of tropes in advertising is that they enhance deliberative processing. This approach may seem to contradict the associative approach, but that is in fact not the case. The resolution of this paradox is that the elaboration effects that have been documented do not pertain to the rhetorical figures themselves, but to the content and form of the advertisement they are part of. Thus, the metaphor above an advertisement may arouse interest and cause a reader to read it and its arguments more carefully. Evidence for this theory comes mainly from experiments carried out by David Mothersbaugh and colleagues. In one experiment it was found that when subjects read a mock magazine –devised for the purpose of the study – they paid more attention to an advertisement for ball points
when it carried in its headline either a scheme (e.g. “One pen – More ink”) or a trope (e.g. “Bingley pens are full of it”), where the effect of the latter was greater than that of the former.\textsuperscript{46} The measure of greater readership was obtained by means of a questionnaire that was attached to the back of the mock magazine. When participants came across this questionnaire, they were not allowed to leave back. In it, participants had to simply relate whether they had read the ad for the pen. Subsequent manipulations revealed that the enhanced readership effect of rhetorical figures was greater when texts were longer or the relevance of the information was lower.

In another set of variations on this experiment, the researchers asked subjects to “list all thoughts, ideas, images, or reactions” they experienced as they watched the advertisement.\textsuperscript{47} Neutral raters later divided these words into two groups: one for messages related to the content of the pen advertisement and one related to the style of the advertisement. Interestingly, it was found that when schemes were used in the headline, subjects generated more thoughts related to both content and style (compared to a headline without a rhetorical figure), and that when a trope was used subjects generated more words related to just the content. Raters also coded the responses of participants for their valence. On this basis it was found that if rhetorical figures were used, more positive brand evaluations were obtained when the advertisement for the pen contained strong arguments (‘the ink cartridge contains 65% more ink’) than if weak arguments were used (‘the ink cartridge contains 2% more ink’). And tropes led to greater appreciation of argument strength than schemes, and thus – one may safely suppose – greater elaboration.\textsuperscript{48} Thus, on this reading the greater increase in tropes compared to schemes in the advertising record is explained with reference to greater message elaboration rather than evaluative conditioning.

\textit{Testing Why Tropes Have Proliferated in Advertising}

We have seen that the associative and the elaboration account both predict that destabilizing tropes will be more successful than schemes.\textsuperscript{49} Thus, it is possible that the entire effect of associative processing of metaphors does nothing to enhance the effectiveness of advertisements, and that this effect is entirely due to enhanced elaboration. In that case no autonomy violation would be involved in the targeted use of tropes. In order to rule out this possibility we would have to find a condition under which the use of tropes should increase on the associative account but not on the elaboration account. As such a condition will count conflict perception. That is, on the elaboration account, advertisements that contain propositions that are contrary to what an agent believes to be valuable will decrease the persuasiveness of the message (and thus purchase intention). On the associative account, by contrast, metaphors and puns are an excellent devise to \textit{bypass} conflict perception. All thing being equal, advertisements with such tropes should therefore lead to more sales.

This argument can be usefully operationalized on the basis of my framework on the structure of branding that was developed in chapter 3. If this framework is correct, then appeals to the affiliation, mating and status motivation modules inherently generate conflict in the consumer. Thus, metaphors and puns that appear in advertisements with cues that activate the three higher-order social motivations should generate more sales (than advertisements that appeal to other motivations) if the associative account is correct. On the other hand, if the sales-enhancing effect of metaphors and puns were due to elaboration, then no such effect should be found. In that case, metaphors and puns in advertisements with higher-order social motivation cues should lead to enhanced elaboration of their content, so that more conflict would be perceived and sales of the respective products would be \textit{impaired}. 

100
3 The Centrality of Autonomy Violation in Branding

The above cases of autonomy violation in branding only violate liberal autonomy if they are sufficiently central to persons at the receiving end. Let me for the purpose of clarity note that I distinguish the centrality of autonomy violation (as discussed in chapter 2.4) from its structural nature (as discussed in chapter 3). The structural nature of autonomy violation is determined in relation to the agent who commits the autonomy violation (e.g. an organization that acts as a controller or authenticity violator). The centrality of autonomy violation, by contrast, is determined in relation to features of the person who is the object of autonomy violation (e.g. the consumer). Thus, there can be cases in which an organization structurally violates autonomy (because of stable institutional features) with effects that are only marginally central to receiving subjects. And there can be cases in which an organization only accidentally violates autonomy, but has a central impact on receiving subjects when it does. The centrality of autonomy violation is measured objectively in terms of the sheer quantity of conscious thoughts and behaviors. Thus, if I acquire an implicit attitude towards ice skating which never impacts either my conscious thoughts or behaviors (maybe because I moved to Australia), then this attitude is not central. Below I will argue that the influence of branding goes beyond such peripheral effects and that it has a central impact on consumers.

3.1 The Centrality of Autonomous Action Violation in Branding

Because purchasing and consumption actions are potentially public events, their effects in terms of conscious thoughts and other behaviors can be mediated either by the agent herself or by other agents. In the first type of case, the agent herself mediates all the effects of purchasing actions and consumption actions. Specifically, such effects of purchasing actions are mediated by two mechanisms. The first one is known in the persuasion literature as the commitment and consistency mechanism and has been studied extensively in hundreds of experiments on cognitive dissonance reduction. Once we have committed ourselves to buying a product (in this case in a way that violated autonomy) then the power of consistency guides us in the direction of future actions that are consistent with that purchase. That is, when we were led to buy a car that represents an upper middle class lifestyle, there may be pressure to buy other commodities that are in line with that lifestyle. The same mechanism may also create a threshold amount of dissonance that would lead our conscious thoughts to align with our actions, provided such an alignment would be the way of least resistance (if it is not, then people would throw away the product).

The second mechanism by which the agent herself mediates the effects of purchasing acts occurs via conscious thoughts. Recent research by Ji Park and Deborah John suggests that carrying around a product (either a Victoria’s secret shopping bag or an MIT pen) can cause personality traits that are associated with those products to ‘rub off’ on consumers. That is, women who carried around a Victoria’s secret shopping bag in a mall for one hour, later reported stronger associations between the self and personality traits that were associated with that brand (i.e. good-looking, feminine and glamorous). Likewise, men who used a MIT pen for six weeks afterwards showed stronger associations with intelligence, being a leader and working hard (traits that were previously shown to be associated with MIT).

Interestingly, this rubbing off effect only occurred for people who had a so called ‘entity theory’ about the self. This concept is derived from the so called Implicit Persons Theory
Measure, which determines one’s score along the extremes of believing that one’s personality is unchangeable (entity theory) to believing that one’s personality is malleable (incremental theory). The explanation of this effect is that self theories activate different strategies for managing public impressions. Entity theorists do not see options for improvement and therefore allocate energy to signaling their traits, while incremental theorists do see possibilities for improvement and therefore allocate their energy not to signaling abilities and traits but to improving them. Suggestive evidence for this interpretation comes from an experiment in which a threat to self-perceived intelligence led to a preference for an MIT pen over a regular pen only for entity theorists. Supposedly, only entity theorists are interested in signaling high intelligence as a means of making up for a perceived deficit in that trait.

In the second group of cases our purchasing act leads other people to act in a way that causes us to act (or think) in response. Thus, when I am led to buy a new suit, a colleague may respond by buying a classier suit for himself, which in turn may lead me to up the ante. A recent study found direct evidence for such arms races, in that it was shown that people who lived in a street in which all inhabitants who had bought a state lottery ticket won a prize spent more money on cars even if they had not bought a ticket (compared to similar people in other streets).51

In the case of consumption behavior, effects that are mediated by others include responses of disgust, approval or even imitation, which may lead to responses on the part of the agent herself in turn. Effects that are directly mediated (by the agent herself) may include – apart from consistency mechanisms – addiction circuits (although the mere act of purchasing may also trigger addiction circuits). Consumption acts may also have negative health consequences, as pointed out by Mitchell and colleagues in their study of television advertising and food consumption. They calculated that children who watched one hour of commercial television a day (which is far below average) would gain 14 pounds in weight a year (if food consumption would not decrease during regular meals). Especially this latter effect would make the effects of consumption behavior sufficiently central on a minimal account.

3.2 The Centrality of Autonomy Violation Brand-related Mental State Formation

Although it is obvious that branding influences conscious thoughts, it is presently very hard to determine the effects of autonomy violating branding. There is even hardly any evidence on mental state formation brought about by branding in general (that is not related to sales or brand awareness). Hence, the effects of branding on personality or value change are still largely uncharted territory (the reason being of course that from a marketing perspective such effects are not relevant). The exception to this state of affairs is a recent study by Agnes Nairn and Pierre Berthon.52 They measured the effects of two types of advertisements on participant’s scores on the so called Romantic-Classicist personality index, which has been specially developed for consumer research. Its questionnaire positions subjects on a scale ranging from romantic (e.g. emotional, spontaneous, impractical) to classic (e.g. rational, orderly, restrained). For the purpose of the experiment, subjects were divided into three groups. One group saw an advertisement for a vacation that used images of classicist features, another group saw the same advertisement with images of romantic features and a control group was not exposed to any advertising. After exposure to the advertising (if there was any for a group) participants completed the personality questionnaire, which – once analyzed –
showed that those who had watched the romantic advertisement scored significantly higher towards the romantic side of the spectrum than the other two groups. If so, then these subjects may also have engaged in more romantic (e.g. spontaneous) thinking styles (as well as romantic behavior of course, but such effects are discussed later on in this section). However, it is unclear from this study how lasting these effects were and it is moreover unclear whether the process that brought about these mental states violated autonomy.53

With regard to the effect of mental states that result from autonomy violating branding on behavior, there is of course a plethora of research on purchasing behavior which was reviewed in section 2.2. The effects of these behaviors were already discussed in section 3.1 and must in the present case be counted as sufficiently central.54 A different type of behavioral effect of attitude change is extra-purchase behaviors. In a fascinating set of experiments, Gráinne Fitzsimons and colleagues showed that brands can unconsciously activate brand attitude structures that in turn activate goal pursuit.55 In one study, one group of subjects was primed with a subliminal image of Apple, which in a pretest was found to be associated with creativity, while another group was subliminally primed with an image of IBM, which did not show any such association. In a follow up ‘unusual uses test’, participants had to think of as many uses of a brick as possible. In terms of both the number of uses and the creativity of these uses (as judged by independent raters) those who had been primed with Apple gave significantly more creative responses. Moreover, this effect increased in strength when the task followed the priming effort by a five minute delay. Because it is typical of goal pursuit that the delay of goal satisfaction increases motivation, this is strong evidence for thinking that goal pursuit is the underlying mechanism. As subjects had no awareness of the prime, the process operates outside of awareness and hence cannot be easily corrected.

In a second study, one group filled in a questionnaire that included, as a priming measure, questions about Disney (which in a pretest was associated with honesty) while another group filled in the same questionnaire with questions about E! Channel (which was not associated with honesty). It was hypothesized that on a questionnaire that presents subjects with scenarios in which they were to choose between acting honestly and acting in a socially desirable way, those motivated to be honest would more often admit to acting in socially desirable manners. As an extra manipulation, the experimenters added for one subgroup of Disney primed subjects a question that implied that they were already honest persons (and so induced the belief that the goal to be honest was satisfied) and for another subgroup a question that implied that they were striving to become honest. If the priming effects related to Disney (of which subjects were not aware) was mediated by goal motivation, then only participants who did not believe that their goal to be honest was satisfied should show the enhanced honesty effect, and that was what the researchers found.

Because these two studies suggest that brands can activate goals unconsciously, there are grounds for thinking that implicit attitudes are involved (although the relation between implicit attitudes and unconscious goal priming are not well understood presently). If so, then the effects of autonomy violating brand attitude formation reach beyond the domain of shopping into the domain of private interactions, where honesty and creativity are important regulatory goal states. In terms of the goal hierarchy discussed in chapter 2.4 these goals are moreover located in the top of the hierarchy (as ‘be goals’) which makes them all the more central.
4 The Retrievability of Brand Managerial Intentions

From the argument in chapter 3 it follows that there are pressures on brand managers to knowingly use messages that undermine the autonomy of consumers. That is, it was argued that there are structural features (of norms, the economy and the mind) that favor the selection of brand managers who devise messages that bypass conflict perception mechanisms and awareness. Strictly speaking, it does not follow from the fact that brand managers with such behaviors are selected that they must have the intention to act in such ways. However, it seems reasonable to assume that — all things being equal — members of an organization who know what they are doing will be more effective at what they are doing.

One might object to this argument that it disregards ethical considerations on the part of managers. It could be that there is a professional ethic that holds brand managers back from engaging in policies that they know will result in manipulation. Sadly, there is evidence to suggest that nothing of the sort is the case. A study into the ethical standards of business school students found that over the period from 1981 to 1987 on average about 50% of students indicated that they were willing to advertise a product as being “new and improved” when they knew that it was not true. A similar share of students indicated the willingness to mention on a shampoo bottle that two applications are recommended, after having learned that one application is “effective”.56 Maybe the business school on which this study was conducted was a hotbed of unethical students, and maybe the early 1980s were a period in which a temporary ethical vacuum emerged. But that is not very likely. For one thing, the trend for unethical responses on both questions is upwards, so that students became less concerned with ethical standards as the 1980s progressed, not more concerned.57 Furthermore, ethical standards can be assumed to be higher for students in their early 20s than for people who have entered the power struggle within organizations. When faced with the realities of making a living, feeding a family and paying off loans, people can be assumed to be less strict with regard to their ethical standards than students who imagine being employed in a firm.

If most brand managers indeed know that some of their messages are processed in a way that defies the recognition of conflict or intentions, then one would expect that such intentions will be readily retrievable. Unfortunately, these intentions have hardly been researched systematically in sociology, barring a recent project by Gergely Nyilasy. Because I deem it unwise to complement this research with lost quotes and self-promoting interviews from the professional advertising press, I will rely on documents from the archives of the Dutch beverage companies Amstel Beer and Van Nelle (the records of which are up to the early 1970s publically accessible). Drawing on both types of sources (i.e. sociology and archives) I will detail the extent to which there is evidence for intentions to violate authenticity and the more encompassing intention to manipulate. To repeat, the intention to violate authenticity only requires that an agent has some awareness that the means by which she wields influence violate autonomy (4.1), while the intention to violate self-control requires moreover the intention to cause a behavior or attitude change in a controlee by means of the influence attempt that is known to violate autonomy (4.2).
4.1 Managerial Knowledge about the Autonomy Violation Involved in Using Persuasion Techniques in Branding

With regard to intentional authenticity violation of consumer actions I can be brief. The effects of music, mimicry and smell are known not to cause awareness in consumers, and retail managers employing these techniques are therefore knowingly subverting authenticity. With regard to attitude change matters are more complicated. Although brand managers intentionally devise messages that induce negative affect, it is highly unlikely that they will know that a possible persuasive effect is due to introjection. But deliberate attempts to bring about evaluative conditioning must involve at the very least knowledge about the lack of deliberation involved when consumers process paired stimuli. One way to find out whether brand managers do indeed have such knowledge is to simply ask them whether this is the case.

Systematic interviews reveal that most brand managers ascribe to the view that advertising works first and foremost because of appeals to ‘emotions’ and only secondarily because of appeals to ‘rationality’. But it is not clear what they mean when they talk about emotions. Because there are hardly any tools to measure the impact of emotions, creatives can insist that making advertisements is an artful skill that defies scientific scrutiny and so maintain that they have cultivated a gut level knowledge about effective emotions. In his field investigation, Nyilasy was unable to find any agreement among brand managers about why emotional advertising works beyond banalities that there are categories of products that are bought more impulsively (e.g. fashion, beer, soft drinks) and products whose purchase is more carefully considered (such as cars). Surveys and systematic interviews moreover consistently demonstrate that brand managers do not keep track of the academic literature on marketing.

Given the present stage of sociological research into the motivations of brand managers, the most effective way to find out to what extent advertising creatives or managers know about the psychological mechanisms by which their messages have an effect is to investigate in detail the historical records of companies. Here one finds reports from creatives of advertising agencies explaining their proposed campaigns to management. One also finds notes of discussions of the board of directors on their reasons for accepting or rejecting proposed campaigns. Strictly speaking, evidence of knowledge about the autonomy violating nature of advertising messages from one of these two sources (i.e. production and endorsement) would suffice for saying that the messages were intentionally autonomy violating. For present purposes I will briefly discuss fragmentary evidence from the archives of the companies Van Nelle and Amstel, from the period 1950 to 1970, because these snippets provide the most insightful traces of what goes on in the minds of advertisers that are currently available.

Evidence of intentions of workers in advertising agencies is best preserved in the form of reports that accompany campaign proposals. These offer suggestive evidence for the idea that advertising creatives gave a lot of thought to the associations they established with their messages. One campaign that is well documented in the Van Nelle archive is the one for Supra coffee, which was presented in what was the company’s biggest promotional campaign until then – the year 1962 – as a new stage in the history of coffee due to the vacuum enclosed package. This emphasis on novelty was the result of one of Van Nelle’s advertising agencies’ (named Interad) strategic considerations that relied heavily on the Unique Selling Proposition (USP) theory developed by Rosser Reeves. Its core idea was that in order to sell a product it is best to be able to mention one characteristic of the product that no other product has. Interad considered three candidate propositions and had them tested on house-
wives. The first was the vacuum aspect – but that was dismissed, as nobody was supposed to buy a product because of a package, and on top of that they feared the negative association of “vacuous”. The second possible USP – the fact that supra coffee was freshly grinded – was also suspect, because “unconsciously” the housewives were expected to associate this aspect with the fresh grinding they still did themselves on coffee-grinders that they had so patiently acquired by collecting points on Van Nelle packages. In other words, the risk was that housewives would feel to “voluntary abolish something they collected with great efforts”. This selling proposition was therefore to be “treated carefully”.63 The final and best proposition was, much to the surprise of the advertisers, the aforementioned novelty. After all was said and done, the campaign was to encompass all three aspects, but to emphasize the last one with the main slogan “An entirely new coffee”. It also included the suggestion of high quality which was deemed “impossible to check” except by buying the product and so gave “ample opportunity to mold the women”.64

The other campaign for which there are more extensive snippets was for tobacco. In 1970, Van Nelle’s advertising agency Van Alfen informed the head of promotion that it had thought of a new slogan for the promotion of Van Nelle medium tobacco. As a variation on the full company name ‘The widow of Van Nelle’ the new slogan was to be “the daughter of the widow of Van Nelle”. The background for this proposal was that at the time Van Nelle dominated the market for heavy tobacco, which had given it a tuff image that held back sales in medium tobacco, especially to young people. But, as Van Alfen’s account manager was careful to observe, the company had to avoid the risk of gaining market share in the medium tobacco market by hurting its share in the more precious heavy tobacco market. In this context he pointed out that there was something “magical” about the “tension” between the words “heavy” and “widow”, which his agency did not understand, but which should not be interfered with by means of the new slogan.65 To this end, he had formulated a range of questions that were subsequently investigated by the market research agency IPM later that year. At its core, the research conducted aimed to establish the consistency of the proposed slogan with the existing Van Nelle branding efforts. It was therefore deemed crucial to discover the estimated age of the widow of Van Nelle, in order to see whether it would be plausible to assert that she had a daughter. What was more, two proto-advertisings (developed by Van Alfen) were shown in combination with the word “medium tobacco” (while the name “daughter of the widow Van Nelle” was covered), after which the responses given by the subjects in a context of free association were matched to target reactions. It turned out that the subjects did swallow the concept of a daughter of the widow of Van Nelle as planned, and that those who already smoked Van Nelle’s heavy tobacco did not disapprove of the family extension, although they did not ponder trying it (which was good, as heavy tobacco smokers were deemed more valuable).66

One can find a similar concern for the associations of their brand in documents from the archive of Amstel, for which there is moreover more telling evidence on the role of the directors. A report by a market research firm on the brand-image of Amstel that was submitted in 1962 observed that although Amstel was well-known with consumers (88% against 95% for its main rival Heineken) its brand-image was not good.67 Amstel was seen as high-brow, civilized, wimpy, and fit for beginning drinkers as well as women. Worse still, it was seen also as neutral, objective and too dispassionate. Heineken on the other hand was seen as cozy, smooth, genial, spicy, sturdy and beer for the masses. When the respondents were asked to link personality types to beer-brands, Heineken was superior on all counts, from sporty types, cozy housefathers to experienced drinkers, except for one character-type: a high-brow lady giving a party.68 These concerns subsequently were to inform Amstel’s promotion policy.
On August 30th 1963 Amstel director Kranenberg outlined what were to become the guiding principles for the years to come. The new advertising manager Avis and he had visited a talk by Ernst Dichter in Munich and "conditionally accepted" the latter’s claim that "a deeper penetration is brought about” by “appealing to unconscious behavioral patterns”. The “unconscious target” of the new policy was to “consistently foster trust in the company and its product”. What this meant in practice was that Amstel restyled its logo, used more “masculine” letters and added below the logo an illustration of a man-woman situation in which “the man must dominate”.69 One month later the new line was presented to the board of directors. Although the company director Egberts wondered “whether it is correct to base advertising for beer on evolution and not on tradition”, the new approach was authorized.70 This policy was carried forward in the years to come and in a 1967 report by the advertising agency FHV we read that these ideas were firmly adopted for the campaign around the slogan “This is the man, this is his beer” (the leading idea of which was that the male consumer of Amstel needed to have his beer at six o’clock, and that his wife needed to take care that he would indeed have it at six o’clock).71 The campaign also included an action in which beer corks could be sent to Amstel for a 25% reduction on “manly” tools, such as pocket-knives, fishing sets and even a Honda 600.72

I take it then that there is suggestive evidence from the snippets of these two company archives, together with field interviews, for thinking that brand managers are well aware that they use techniques that rely on attaching their brand to associations in the mind of consumers and that the nature of such associations is emotional and largely unconscious. If so, then they may intentionally persuade consumers with messages they know to have effects that the consumer is unable to perceive, let alone to reflect upon in case of conflict.

4.2 Managerial Intentions to Influence Behavior by Means of Autonomy Violating Techniques

So far I have argued that there is suggestive empirical evidence for the theory-driven expectation that advertisers deliberatively use evaluative conditioning techniques and have some understanding of these techniques operating in a way that violates autonomy. If such evidence were sufficiently strong, it would prove that intentional authenticity violation occurs in branding. However, in order to intentionally violate self-control, advertisers must also be shown to have the goal of bringing about a certain behavior on the part of the consumer.

If it is true that positive attitudes towards a brand that are brought about by misattribution EC lead the consumer to purchase that brand, then it is uncontroversial to suppose that brand managers want to bring about the behavior in question. Increasing sales, after all, is what they are supposed to do and what they tell others they are doing (although there are of course different types of marketing campaigns, some of which merely want to raise awareness for a product). To sum up, with regard to authenticity violation unconscious influence can be believed to be a byproduct, but with regard to self-control the unconscious influence must be intended to bring about a certain behavior. However, because brand managers are single-mindedly focused on causing sales, this distinction does not make much of a difference in the present context.
Consumer Consent to Covert Brand Influence

If my arguments so far are correct, then brand managers violate the autonomy of consumers with regard to actions and attitudes in a way that is both central and retrievable. However, in order to establish whether the liberal state should intervene to stop brand managers from doing so we must know whether consumers are able to give appropriate consent. Or if this would not yet be the case, whether conditions can be conceived of in which such consent is possible. In chapter 2.4 I distinguished three levels at which consent to an autonomy undermining practice can be given – that of the message, that of the setting (in which people can expect to receive a specific type of message) and that of society at large (in which messages of the specified type are common) – and argued that people secure their autonomy by giving consent that is either tacit (5.1) or explicit (5.2).

A complication with regard to branding is that it includes a number of practices with distinct consent contexts. That is, types of branding differ in the extent to which it is clear that a message is intended to be persuasive and the extent to which a set of messages is confined to a delineated setting. For present purposes we must distinguish at least six types of branding: explicit and covert interpersonal influence in shops, explicit interpersonal influence at home (door-to-door salesmen and direct marketing), covert interpersonal influence in extra-shop and extra-domestic environments, product placements (in movies and games), advertising in delineated slots (on tv, on radio, in print and on the internet) and outdoor advertising (including billboards, sponsorship and guerilla marketing). Let me repeat that I am concerned here solely with messages that occur in these branding contexts insofar as they violate autonomy.

5.1 Tacit Consent to Autonomy Violation in Branding

With regard to consent at the level of the individual message and settings in which messages occur, we noted that tacit consent cannot legitimize autonomy violation if such violation is central to the influenced person. The reason was that assuming that people tacitly consent to a practice that violates their autonomy becomes too risky if the consequences of the autonomy violation become too serious. In the absence of objective indicators of tacit consent, all autonomy undermining practices can be justified on this basis and that seems wrong (at the very least for manipulations that have a central impact). That is, when entering a shop, switching on the television or radio, entering a cinema or starting up an internet browser, consumers can be argued to give implicit consent to the manipulation that occurs via these channels. But it can also be argued that they do not. In the absence of objective criteria for thinking that they do, we are best advised to play safe with an eye to the substantial incremental effects of these messages, and so to require explicit consent.

Tacit consent at the level of society is more problematic still. It can be argued that by taking part in a market economy we tacitly consent to a number of implicit commitments that are a necessary part of it. If branding (including its autonomy violating elements) is a necessary part of a market economy, then we can be said to tacitly consent to autonomy violating branding upon becoming members of market economies. This argument fails however because branding is not a necessary part of market economies either historically or in terms of the size of its contribution (as I will show in chapter 5). If that is correct, then there is no pressure to tacitly consent to branding that is derivative of pressure to consent to market economies.
5.2 Explicit Consent to Autonomy Violation in Branding

Because tacit consent to branding is insufficient, we must investigate whether explicit consent will do. More specifically, we must ask both whether adequate consent is already given and, if this is not the case, whether it is a realistic possibility.

With regard to individual messages it can be argued that in the case of overt messages consumers can decide for themselves whether they are willing to be exposed to branding or not. This may be more difficult for product placement and some types of internet advertising, because in these cases messages cannot be avoided individually (one does not know when product placements will figure in a movie or when an advertisement will pop up on a web page). But we can certainly decline a sales pitch, turn a newspaper page when our eyes slide over an advertisement, and switch to a different channel when a commercial break interrupts a television program. That is, in these cases we can know more or less when advertisements are coming, we can recognize them because they are clearly demarcated as such, and we can avoid them. However, for explicit consent it is moreover necessary that people produce some public representation that unambiguously signifies that they consent to an autonomy violating practice. Not only is this not yet the case, it is also not realistic that individual branded messages will be subjected to explicit consent, given that the typical consumer in industrialized societies is estimated to be exposed to tens of thousands of messages a year.73

More realistic is explicit consent at the level of settings in which autonomy violating messages can be expected to occur. Because what is worrisome about branding is not attitude change brought about by individual branded messages, but rather their incremental effect, the object of consent would also have to be this incremental effect. For this reason, consent at the moment of switching on a mass media devise is not adequate. Instead, consent would have to be required when engaging in a cable television contract, an internet contract or when buying a radio – for these are the moments at which deliberation about long-term effects is realistically conceivable. However, a difficulty for such a consent procedure is that it is impossible for the consumer to know which value induction by which methods she will be consenting to. That is, because of the dynamics of fashion trends, the rate of product innovation, the wear-out of persuasion techniques, changes in demographics and shifts in popular culture, brand managers have to constantly adapt the content of their messages to retain consumers or to attract new ones. As a consequence, the themes and persuasion techniques shift so rapidly that the consumer cannot hope to know the manipulative intentions behind the hundreds of persuasive messages she is confronted with every day. That is, she does not know which values manipulators will target, nor does she know the direction in which her values will be changed. She knows only that her values will be changed in whatever way brand managers deem most prudent in order to gain or retain market share. In other words, what must be required of the consumer when she signs a cable television contract is that she give carte blanche consent.

I noted in chapter 2.4 that there is serious disagreement about whether explicit consent can justify central autonomy violation. On the present-autonomy reading it is possible to introduce consent conditions that allow consumers to give appropriate consent to branding influence, while the balanced-autonomy interpretation rules out this possibility. Which interpretation is the correct one is – as I noted earlier – a matter that would require more thorough analysis than I can hope to give in this thesis. In what follows I will therefore work out the implications of both readings.
5 Why Branding is Special in Structurally Violating Liberal Autonomy

We arrive now at the fourth step of the argumentative design that was mapped out in chapter 1. I noted there that if autonomy is violated in a host of organizational practices besides marketing, it would seem that I have been developing an argument not so much against branding, but against complex societies. Hence, the scope of my normative analysis must be expanded. Given the considerable work that went into establishing whether branding violates liberal autonomy, I confine myself to analyzing two subsequent organizational practices that are intuitively most likely to violate autonomy. My argument will be that even on the most sympathetic reading of the evidence for autonomy violation in religion and political promotion, these violations are not irreconcilable with liberal autonomy. If so, then branding would seem to be special in violating liberal autonomy.

In section 1, I turn to religion and argue that although stable ecological features may exert pressure in the direction of autonomy violation, people are able to give proper consent to religious practices. Hence, at least the consent threshold is not transgressed. In section 2, I consider political promotion, which turns out to transgress all practical thresholds, but to do so – I will argue – for a good reason (from the perspective of liberal legitimacy).

1 Liberal Autonomy and Religion

It can be argued that the assessment of autonomy violation in branding applies to religion as well because of structural similarities between branding and religion. Specifically, recent research carried out by Ron Shachar and colleagues suggests that religiosity and brand reliance may function as substitutes. To begin with, it turns out that the number of brand stores (for brands such as Apple and The Gap) negatively correlates with the number of congregations and religious attendants both on the level of states (within the US) and on the level of counties within California, Texas and New York. This is suggestive evidence for some type of similarity in appeal, since alternative explanations in terms of income, education and urbanization were controlled for. Subsequent psychological experiments corroborated this interpretation. When subjects (with diverse religious backgrounds) were asked to write about a personal religious experience they thereafter were less likely to choose national brands over equivalent products than a control group that was not primed with religion. Crucially,
this effect occurred only for products that can be used to signal something to others (e.g. watches, sunglasses and soccer socks) and not for purely functional products (e.g. bread, batteries and medicine). In a follow-up experiment it was found that self-ratings of chronic religiosity were negatively correlated with the choice for national brands (for products that have signaling functions). If all of this is correct, then it can be argued that religion might appeal to the same signaling motivations as branding, and that it will therefore violate liberal autonomy for similar structural reasons.

On the other hand, religion is also different from branding in that it is not a devise to enhance profits. Moreover, religious doctrines and rituals are constrained by tradition, from which they typically receive much of their authority. If that is true, then religious agents do not have the same elbow room and incentives as brand managers for appealing to stable motivations. As a result, there is no similar pressure to undermine autonomy. Admittedly, this does not rule out that religious traditions are the result of some type of competition. In fact, naturalistic accounts of religion would predict that the types of religion that were most effective in appealing to cognitive and motivational structures have survived. Still, this does not entail that the same pressure for autonomy violation existed for religions as it exists today for brands. If religions are typically constrained by tradition, then the form of their appeal to the affiliation module was probably shaped way back in human history when people may not have had many competing affiliations. If so, then there was no pressure on traditional religions to develop autonomy violating techniques.

What I need then in order to assess structural autonomy violation in religion is a naturalistic account of religion that takes these different ecological conditions into account. That is, if we know the psychological and ecological constraints under which traditional religions emerged, then we can assess whether these have generated pressure to undermine autonomy. This is what I do in section 1.1, where I argue that this may indeed be the case. However, in section 1.2 I go on to argue that – precisely because of the role of tradition in religion that sets it apart from branding – the practical threshold of liberal autonomy violation is not transgressed, at least with regard to consent.

1.1 Structural Autonomy Violation in Religion
Since the explanation of religion is the main focus of research of naturalistic accounts of culture there is no shortage of theories in this regard. Some of these theories consider psychological conditions for religion independent of ecological constraints. Among these theories, most aim to explain the structure of belief in supernatural beings. Part of the emerging consensus in this regard is that the mindreading module that we use to interpret behavior of others is calibrated in an oversensitive way, so that it tends to interpret nonhuman phenomena (such as the blowing of the wind) as caused by some agent. Another part of the explanation is that there is an attention optimum for religious beliefs, such that they should be sufficiently odd to arouse attention but not so counterintuitive that they cannot be related to other thoughts (and hence cannot be remembered). Whatever the details of these explanations, they are not relevant for my purpose, because these properties of beliefs do not violate minimal autonomy.

Other naturalistic theories of religion take ecological constraints into account. One type of explanation points out that the environment in which most major religions emerged was ridden with distress caused by death, deception, disease, pain, loneliness and injustice. Given some minimal assumptions about a motivation to assuage fear, religions may be expected to tap into this motivation under these ecological conditions. The main support for
this claim – besides the anthropological record on fear inducement in religion – comes from psychological experiments, carried out by Scott Atran and Aran Norenzayan. These experiments demonstrate that American subjects report stronger belief in God and supernatural intervention after they have been exposed to a story about a terrible accident, just as a group of Maya subjects was more likely to believe in the efficacy of prayer after being exposed to death primes.9 Granted, many religious rituals actually induce intense fears in practitioners, but, as Atran and Norenzayan note, this fear is always assuaged in the end.10 It is not clear though by which precise psychological process fear is assuaged, while the experiment has also not been replicated for other religions. Still, this line of research offers suggestive evidence for thinking that religions tap into a fear-reduction desire. If so, then the presentation of religious symbols and doctrines to people who experience negative affect might conceivably induce introjection.

A second type of ecological constraint that is used to explain religion is the problem of cooperation. That is, in order to extract resources from the environment as well as for the purpose of protection against threats, it is beneficial to cooperate not only with those to whom one is genetically related (i.e. inclusive fitness) and those who reciprocate, but also with people in communities whose size makes it impossible to keep track of reciprocal relations due to limited human memory capacity.11 The standard problem in explaining such cooperative behavior is that sacrificing self-interest for the collective is an evolutionary dead-end if there are free riders in the group who benefit from such efforts without contributing (the energy thus saved by the free riders translates into a host of minor benefits that increase their reproductive success). In response to this problem, some have proposed that humans signal their trustworthiness by making costly commitments to the group and that such sacrifices occur in the context of religion.

There is some empirical evidence that supports this view. First of all, the anthropological record shows that sacrifice is endemic in religion.12 Moreover, a study which compared around 200 religious and secular communes from the 19th century found that religious communities were four times as likely to outlast secular ones, while subsequent analysis demonstrated that the costs imposed by religious communities positively correlated with their lifespan. Another study found that men from a Kibbutz acted more altruistically in a resource game than secular males and females as well as female Kibbutz members, which was explained by the fact that only men were required to pray three times a day.13 However, costly commitments in themselves do not seem to guarantee cooperation from a member (since they could be made because of opportunistic calculations). It still needs to be explained then why and under which conditions costly commitments are a reliable signal of cooperative intent.

One way to account for the reliability of commitment signals, proposed by Candace Alcorta and Richard Sosis, is to point to the deep emotions involved in rituals, which are difficult to control by the executive system in the brain.14 That is, primary emotions are rooted deep in the brain where they took root early in mammalian evolution to deal with threats and opportunities. As such, they function relatively independently from higher-order mental capacities that developed much later in the neocortex. Hence, the higher-order intention to free ride would be unable to make the organism fake authentic happiness, sadness or fear.15 Of course, not any display of deep emotion will do in convincing a person that her group members will cooperate. They must be sensible displays of commitment to the group. Alcorta and Sosis point to two kinds of displays as most credible. One is the display of group identification by means of positive emotions, while another is the display of fear for justice-type gods who punish those who fail to cooperate. Positive emotions are aroused by

112
the dopaminergic reward system and motivate approach related behavior. When this system is activated – in rituals that arouse positive emotions – dopamine is released and a reward is experienced. If the system is repeatedly activated in relation to symbols that mark the group, Alcorta and Sosis argue, then incentive learning occurs via some associational network, and the symbol that stands for the group is invested with positive valence. A similar process of association – with even more lasting effects – occurs when justice Gods are associated with negative emotions processed in the amygdala, which is particularly common in dramatic rites of passage. On this account, part of the function of rituals is to create the conditions – with the help of chanting, dance and music – under which implicit attitudes are evoked.

A limitation of this approach is that all of these mechanisms serve to assess the cooperative intent of other people. As such, they can only explain cooperation in small communities and not for larger social settings in which individuals are no longer able to assess the cooperative intentions of other members. In response, Joseph Bulbulia has proposed that for the purpose of cooperation people not only assess other members for their intentions but also environmental signals that are effective in triggering cooperative behavior. Such signals must indicate three things for them to effectively reduce uncertainty of cooperation. They must signal that others are motivated to cooperate, they must signal that these motivations are synchronized and they must do both of these things reliably.

The main idea behind Bulbulia’s proposal is that the reliability of environmental signals is secured by their automaticity. He notes a number of priming studies in which pro-social behavior is triggered automatically by cues in the environment. If there are such processes, then the corresponding cues may be taken as reliable signals of future cooperative behavior (of whoever the members of the community may be), since the motivations that are triggered are not contingent on possible Machiavellian machinations of system 2 processes. The evidence for such functioning of environmental signals is not yet very strong however. Bulbulia mentions one neuroimaging study which shows that when a group of Danish Christians were praying, their dorsal stratum was activated, which is associated with delayed reward. Thus, ritual prayer may indicate that anti-social goals – that may interfere with cooperation – are suppressed and hence signal to a person that future cooperation is likely. With regard to synchronization, Bulbulia notes a study which shows that synchronous movement (passing a cup) and singing is associated with decreased rates of defection in a subsequent game. Although this argument is largely speculative, I will opt for a sympathetic reading and accept that there is a suggestive case for thinking that both in small groups and larger collectives religious music, bodily practice and prayer may function as environmental signals of cooperative goal sharing. If this is true, then such practices would – together with possible occurrences of introjection in religious fear-responses – violate minimal self-control and authenticity conditions for structural reasons.

1.2 Practical Thresholds of Autonomy Violation in Religion
There is unequivocal evidence on the central and lasting impact of religious practices on the lives of converts. Since it can therefore not be seriously doubted that religious persuasion transgresses the centrality threshold I will focus here on the practical thresholds for retrievable intentions and consent.

Retrieval Intentions to Violate Autonomy in Religion
There are two theoretical reasons for thinking that the opportunity of religious specialists to manipulate participants is constrained. First of all, if religious rituals consist in part of emo-
tional displays of cooperative intent, then cynical religious officials would have a hard time deliberatively manipulating their subjects without believing in the creed themselves as well (at least to some extent). This is of course not a safeguard against manipulation, but the argument suggests that the pool of manipulators will be restricted to those who actually believe in the religion. Such manipulation, in which religious officials use free slots in rituals to deliberatively change behavior and attitudes is by itself of course still unacceptable if it involves the structural activation of unconscious pathways. Secondly, next to constraints upon the formation of manipulative intentions there are constrains upon the ability to carry out such intentions. Put simply, performing a religious ritual is not doing whatever the official feels like doing at that moment. Religious rituals are typically rigid, repetitive and highly scripted, while ritual practitioners tend to be obsessed with strictly carrying out rituals, thus limiting the discretionary space in which religious specialists can act upon possible manipulative intentions.

One objection to this argument is that new religious groups are not constrained by ties of tradition, so that their leaders have the discretionary wiggle room to intentionally implement autonomy violating practices. In response, a pragmatic argument can be constructed for granting small groups some leeway. For what sociological studies demonstrate is that small cult groups have a chronic shortage of money due to their time-intensive and demanding group activities (once you are in, you have to take on lots of commitments and have less time to work and earn money). Hence, members are forced to go out and collect money for the group, which brings them into contact with different opinions and leads to high defection rates. So although the leader of a new cult may wield extensive power in designing ritual practices, financial constraints keeps expansion of such groups in check.

Consent in Religious Conversion

With regard to consent in religion what we need to know is whether potential converts are in a position to properly assess what they are to consent to. For one, because religions are typically constrained by tradition and therefore highly stable, it would seem that potential converts can be held responsible for informing themselves about the ritual practices of the religion they want to join. If such practices are realistically knowable, then the crucial question is whether conversion processes are structured in such a way that seekers can take steps to acquire this information before conversion. For the purpose of such an assessment I will work with the taxonomy of conversions that has been developed by John Lofland and Norman Skonovd. They identify six conversion motifs, of which one half is classified as gradual conversion and the other half as quick conversion. The gradual intellectual, affectional and experimental conversion motifs royally grant participants the opportunity to inform themselves about the religion as they move through various stages. This is not necessarily the case for mystical, revivalist and coercive conversion motifs, which entail very short conversion periods in which the to-be-converted are subjected to various mental and physical pressures.

There are good reasons not to be all that worried about the latter cases though. First, the most serious impediment to proper consent is so called sudden or instant conversion. However, although early twentieth century conversion research was obsessed with such conversions, it has systematically failed to come up with hard evidence. Notions of instant mass conversion are even more problematic, as there are no reliable eyewitness accounts of any such events. Its assumed role in the rapid spread of Christianity in the Roman Empire for example was the result of sloppy calculations. Secondly, even quick conversions (which need not be instant conversions) are a relatively rare phenomenon. A recent study based on
the Lofland and Skonovd model underscores this point: of 70 cases of conversion to Islam in the United Kingdom, only 10 cases could be classified as Mystical, 3 as Coercive and none as revivalist (by contrast, intellectual motifs were reported by 60 converts, experimental by 42 and affectional by 46). Furthermore, statistical analysis of self-reported experiences of Christians shows that these map rather nicely on a model of gradual conversion developed by Lewis Rambo, in which the convert goes through no less than seven stages. Thirdly, participatory research by Rodney Stark and Norman Lofland in small religious groups demonstrates that conversion typically occurs by establishing personal contacts (which facilitates a gradual conversion process).

In sum, religion has a central impact on believers and its officials may have some room for pursuing manipulative intentions. But when it comes to consent, religions do not transgress the minimal practical threshold because they are stable and because conversion is typically gradual.

The central argument of this section has been that religion is different from branding when it comes to autonomy violation mainly because it is constrained by tradition. In this regard it must be noted that this conclusion does not apply to the peculiar development of religion in the United States over the past century. Since the late 19th century, a number of US churches have begun to apply marketing principles for the purpose of attracting new members and retaining existing ones. This development has accelerated from the 1950s onwards because of the emergence of so called mega churches (defined as having on average over 2000 attendees) which may have as much as 43,500 members per ‘church’ (these churches are in fact huge stadiums). At present there are 1384 of such mega churches in the US alone, with huge protestant mega churches now arising in South Korea as well. As a result of massive sales of Christian popular music, movies and books, mega churches now command considerable resources – and hence marketing budgets – with an average income in 2008 of $6.5 million per church.

With this money religious officials are actively devising strategies to increase growth, for which they rely on the assistance of about 5000 religious consultants (as estimated by the Church Executive Magazine in 2005). These include denominational consultants (who are employed by churches and have titles such as ‘church growth consultant’ or ‘director of church development’) and nondenominational consultants (who are hired for fees on a project basis). In order to facilitate the work of such consultants, churches now regularly order demographic assessments, surveys and focus group session to keep track of changes in demand. Some of the consultants have been reported to advice their clients to add contemporary music, modern presentation styles and more drama to the sermons. Indeed, a 2008 Hartford Institute survey found that 96% of mega churches make use of electrical guitars or bass while 97% use drums and visual projection techniques. According to a 2005 Newsweek report, a new residence of the leading Willow Creek mega church is totally free of stained glass, bibles and crosses because market research found that attendees do not like such traditional features. Thus it is fair to conclude that constraints of tradition hardly count for anything in mega churches. As a result, practical consent and retrievable intention thresholds are transgressed, and that means that ‘brands of faith’ must be put in the same evaluative category as commercial brands.
2 Liberal Autonomy and Political Promotion

There are several *prima facie* reasons for thinking that the conclusions about autonomy violation in branding apply also to political promotion. First of all, political promotion is developed by the very same public relations firms and advertising agencies that develop branding campaigns. If the same people serve two competitive markets then it stands to reason that they draw on the same talents and techniques. Secondly, branding campaigns and political promotion campaigns are evaluated on the basis of the same criteria by the professionals involved. As such, Barack Obama and his team defeated the teams behind Nike and Apple to win the 2008 marketer of the year award. Finally, a study using the Implicit Association Test showed that voting behavior of Italian swing voters (i.e. the ones who could not tell on who they were going to vote before the election) can be better explained when implicit attitudes are taken into account besides explicit attitudes. Thus, the type of attitude that is most vulnerable to autonomy violation (i.e. implicit attitudes) is predictive of relevant behavior in the political context as well.

However, for my normative assessment of branding to carry over to political promotion, it must be true that stable ecological and psychological features of the type that make autonomy violation in branding structural exert pressure on political promotion as well. This is obviously not the case with regard to economic conditions because political parties do not operate in a market and because voting behavior is *not public* (so that it is at least not evident that voting is subject to the same three motivational constraints, which have as their main output signaling behavior). Hence, it must be investigated whether there are psychological and ecological features of the political domain that make autonomy violation in political promotion structural. This is what I do in section 2.1 where I argue that this is indeed the case. In section 2.2 I go on to establish that practical thresholds for liberal autonomy are transgressed in political promotion as well. However, I then claim that political promotion is different from branding in that there is a compelling reason for the liberal state to allow for autonomy violation in the former but not in the latter.

2.1 Structural Autonomy Violation in Political Promotion

As in the case of branding, we need to ask whether there are structural pressures for political campaigners to engage in autonomy undermining activities. And we must do so both for influence on actions and influence on mental states.

**Triggering Voting Behavior**

The feedback mechanism that I postulated for branding works differently for political promotion, since the latter is not produced in the context of a market. The amount of promotion that a political organization is able to produce is not even simply a function of the amount of votes that the organization receives. Instead, this amount is determined by the quantity of money that the organization manages to acquire from donors or (depending on national regulation) the state. All the same, there are several reasons for assuming that there will be a strong correlation between campaign funds and election results. First, funders are a subgroup of voters for a party, so that one may assume that – all things being equal – when the group of voters increases, the group of funders will increase as well. Second, funders will want their contributions to be effective, where effectiveness will be assessed in terms of the
share of votes that a party (or a proposal) gains. Thirdly, public funding of political campaigns is often contingent on the amount of votes that a party received.

If this is correct, then there are structural pressures to use persuasive techniques that directly influence voting behavior. As direct influence on behavior will count – as in the case of branding – goal activation. Recent research on the influence of locations on voting has provided evidence for precisely such an effect. Jonah Berger and colleagues for example found an influence of voting in schools on support for a ballot initiative on education spending in the state of Arizona. Even after they controlled for demographics and the political views of voters, people who voted in schools were still significantly more likely to vote in favor of higher education spending. The researchers reasoned that if this effect was due to priming, they should be able to replicate the effect in a laboratory setting. Thus they had subjects rate either images that were related to schools (such as classrooms) or images that were not related to schools (e.g. office buildings). It was found that on a subsequent task subjects who had rated school-related images were more likely to vote for the education spending initiative. What is more, two weeks before the experiment the attitudes of the participants on a range of issues – including school reform – had been obtained. Interestingly, the effect of these attitudes was the same between the two priming conditions, so that the voting effect is probably not due to previous relevant attitudes rising to the surface of consciousness. Similarly, subsequent analysis of election results suggests that voting in a church may have facilitated voting for conservative candidates as well as for a ban on same-sex marriage.

Changing Political Attitudes

With regard to mental state formation due to branding I identified two features that exert pressure in the direction of triggering associative processing over deliberative processing. One was psychological and concerned stable motivations and the other was cultural and concerned stable social norms. I will now argue that there are similarly, on a sympathetic reading of the evidence, both motivations and norms that drive political messages in the direction of autonomy violation.

If the structural nature of autonomy violation in political promotion were to be rooted in motivations, then there would have to be stable motivations that consistently pull citizens in an identifiable political direction. For political campaigners will structurally craft their messages such that these motivations are mobilized in favor of their position only if they are confident about the effects of their messages. Thus, the first crucial question is whether there are stable positions in political contests (to which motivations could relate in a stable way). In this regard, the two core dimensions equality and openness to change can be argued to capture most of the left/right (or liberal/conservative) divide. These dimension by no means go together always for everyone, but they appear to go together for most voters most of the time across societies. John Jost and colleagues propose to call these two dimensions core principles of ideology and to distinguish this core from peripheral principles. If there are indeed relatively stable ideological positions, then motivations and cognitive styles that correlate with such a position would be the type of features that can make autonomy violation structural. Interestingly, a meta-analysis by Jost and colleagues indicates that there may be two relevant features in this regard.

The first group of motivations that correlate with political orientation thus defined is epistemic in nature. To begin with, experiments in a number of countries show an association between political attitudes and the Need for Cognitive Closure scale, which measures people’s motivation to have simple, clear and certain views. That is, it is consistently found
that those who score high on the Need for Cognitive Closure scale report having more conserva-
tive attitudes.48 Similarly, those who have a high need to evaluate – i.e. people with the propen-
sity to perceive the world in good/bad terms – report having more conservative atti-
tudes. Conversely, people who score high on the Need for Cognition scale (which measures,
among other things, how much people enjoy thinking) have more liberal views.49 Taken
together, research in political psychology suggests that more cognitive elaboration is associ-
ated with more liberal views. If that were true, then cognitive load manipulations should
lead subjects to shift their attitudes to the right. This is indeed what has been found in a
number of experiments. In one, the propensity to blame persons for their problems (i.e. a con-
servative view) was shown to increase if subjects were prevented from undertaking cogni-
tive correction efforts. Likewise, if subjects consumed alcohol, their opinions were consider-
ably more conservative.50 If this is correct, then conservative political campaigners may have
a structural interest in impeding cognitive elaboration efforts (although not of course on all
and every issue).

A second group of motivations is related to the motivation for group affiliation. In par-
ticular, the mechanism studied in Terror Management Theory by which threat cues trans-
late into greater allegiance to one’s in-group appears to be more conducive to conservati-
ves. That is, showing an existential threat prime led voters to shift their preferences from the
liberal presidential candidate John Kerry to the conservative candidate George W. Bush.
And a study that tracked a group of republican and democratic voters who were highly
exposed to the September 11 attacks for 18 months after the event showed that these voters
shifted their attitudes disproportionally to the right.51 If the affiliation motivation indeed
informs political preferences, then at least conservative political campaigners have an inter-
est in appealing to it. However, in doing so, they face the same constraints that brand manag-
ers have to deal with. As a result, a similar structural pressure in the direction of autonomy
violation that characterized branding may be operative in the political domain.

Besides motivations, stable social norms might also exert structural pressure in the direc-
tion of autonomy violation in the political domain. In particular, social norms might make
subliminal influence more effective than explicit influence. In an internet-based study, Joel
Weinberger and Drew Westen have recently found evidence of explicit attitude change with
regard to political candidates due to subliminal influence.52 In one experiment, subjects
rated an unknown political candidate more negatively on a number of seven point scales if
they had been subliminally exposed beforehand to the word RATS, but not if the word
STAR or the letters XXXX were used. In a second experiment, the effect was replicated for
the well-known Democratic politician Gray Davis, who was in 2003 – at the time of the
study – negatively evaluated by voters across the political spectrum. It was hypothesized that
since Democratic and Independent voters evaluated Bill Clinton positively, subliminally
associating him with Davis would boost their explicit rating of him. Since Republicans held
a very negative opinion of Clinton, their already negative rating of Davis was expected to
worsen still further after the subliminal association. Both predictions were validated. Thus,
subliminal influence might work both for new and familiar political figures in a real world
setting (i.e. that of the internet users that participated in the study).

However, all this may – given that subliminal attitude change violates minimal authenti-
city and self-control – at best prove the possibility of autonomy violation. For it to be a case of
structural autonomy violation it would have to be true that subliminal messages are more
effective than explicit messages in changing voting-relevant attitudes. Hence it must be indi-
cated at least why political campaigners would have an incentive to use subliminal influence.
In this regard, I propose to say that campaigners have a reason to use subliminal messages if
they know the message to be effective if it were not for negative explicit reactions due its violation of a social norm. In the realm of at least the political arena of the United States, racist messages have proven highly effective in political campaigns. However, openly playing the race card is a dangerous strategy, because it violates strong norms of equality. This may alienate voters from issues of the campaign and it may alienate them from the politicians who authorized the campaign. In that context, subliminal (or associatively processed) racist messages may be assumed to be more effective than explicit racist messages.

2.2 Practical Thresholds of Liberal Autonomy in Political Promotion

When it comes to practical thresholds of liberal autonomy, there is no reason to assume that political promotion will be any different from branding. Although political campaigns are waged only before elections, changing political attitudes and behavior has major ramifications so that the centrality of the presumed autonomy violation cannot be disputed. Similarly, retrievable intentions to change attitudes and behavior will be similar to those that motivate brand managers, especially since the same people and organizations develop political promotion campaigns. And because political messages are in flux and transmitted via the same media, consent conditions are similar to those in branding as well.

As such it would seem to follow that the liberal state should act against the types of political promotion that violate liberal autonomy. Although this argument would uncontroversially apply to priming efforts in the voting booth, I want to argue that there is a compelling reason not to restrict efforts aimed at political attitude change. This compelling reason is that political communication is constitutive of liberal legitimacy. This is the case because political legitimacy is contingent on democratic involvement, and democratic involvement is possible only by virtue of free political communication. Thus, if political communication is restricted, then liberal legitimacy may decline. So although autonomy violation in political attitude change is objectionable, restricting political communication – and all the risks that it brings with it – is far more objectionable from the perspective of liberal legitimacy. If this is a sound judgment, then autonomy violation in political promotion should not be confined.

One can of course argue that although branding violates liberal autonomy this fact is – just as in the case of politics – trumped by a good reason, which is the efficient functioning of markets. But such an argument is not plausible. Whether branding increases market efficiency is first of all an empirical question, and one on which economists continue to disagree. That is, the market power theory of advertising argues that advertising leads to more artificial product differentiation, and thus to less perceived substitutability on the part of consumers. The result would be reduced price elasticity. The information theory of advertising, by contrast, argues that advertising informs consumers about substitutes and thus increases price elasticity. There are empirical studies backing both sides, so that the debate is now about which overarching theory best accounts for the mixed data. Hence, the net effect of advertising on market efficiency is not clear, and if this effect would turn out to be positive estimates do not exceed a gain of around 1% of Gross Domestic Product. That would hardly be a compelling reason for justifying autonomy violation.

To sum up, religion is different from branding because it is highly stable in content and practice and hence offers realistic opportunities for consent, while political promotion is different from branding because it is a necessary condition for liberal legitimacy. In other words, branding is special in violating liberal autonomy because it is not stabilized by tradition and because there is no good reason that justifies its autonomy violation.
Conclusion

My main argument in this thesis has been that the liberal state ought to either ban branding that undermines liberal autonomy or create practical opportunities that allow citizens to properly consent to it (depending on one’s take on the type of consent that liberal autonomy demands). In these concluding remarks I want to explore the political relevance of this argument. In doing so, I am going to assume that in order to be politically relevant an argument must be both convincing and able to motivate agents to advance it in the political domain.

In order to assess the robustness of the overall argument it is convenient to break it down into three claims. The first claim is that the minimal theory of autonomy that I have developed is a non-perfectionist account that legitimizes the liberal state. I take it that the argument for thinking that autonomy is the most plausible basis for liberal legitimacy is sound, provided that it is possible to develop a non-perfectionist theory of autonomy. With regard to the theory that I have developed for this purpose, the self-control component is certainly less controversial than the part about authenticity. Fortunately, these components of autonomy imply very similar minimal empirical standards – if my account is correct – so that very little hangs on the validity of my conception of authenticity. I am furthermore fairly confident about the intuitive plausibility of my practical thresholds of liberal autonomy (with respect to both self-control and authenticity), given the outcome of the test that consisted in applying them to the fields of religion and politics.

The second claim is that branding violates this minimal conception of autonomy at least some of the time. If intentions or conflicts (or both) are indeed not recognized in associative processing or automatic goal activation, then there is – I think it is fair to say – solid evidence from consumer psychology for thinking that branding violates liberal autonomy at least sometimes. Finally, the third claim is that branding undermines minimal autonomy for structural reasons, and this claim is certainly more contentious. To begin with, the naturalistic account of culture that forms the backdrop of my analysis is very controversial and has not yet been corroborated extensively. Moreover, my postulation of higher-order social motivations is contested both theoretically and empirically. It has furthermore been tested only incidentally in the context of consumer psychology and not at all in the context of the socio-historical reality of branding. The same holds for my claim that the conflict that such claims generate would make associative processing more effective from the perspective of brand managers. In short, my claims about the structural nature of autonomy violation in branding rely on a host of assumptions and ought to be put to the test, for the purpose of which I have developed a number of hypotheses. Absent empirical corroboration along these lines, it is not sufficiently clear how widespread autonomy violation in branding is and which type of state intervention would be a proportional response.

This brings me to another reason for skepticism about the political relevance of my argument. It can be objected that even if all the claims that I have made in the course of my
argument would be empirically corroborated, my argument would still not have any practical implications because there is no political support for the issue. That is, people simply do not seem to care much about covert influence in branding, especially not if it gives them a reduction on, say, the price of entertainment on television. They presumably know at some level that they get this reduction (on the price of a television cable contract) because advertisements are effective in inducing viewers to buy products. And given that so little actual product information appears in commercials, they presumably also know at some level that such influence is unlikely to take place via deliberative processing. Thus, people seem to willingly sell some of their autonomy to a cable television provider in exchange for a price cut. If there is indeed so little concern for autonomy violation, then it is unlikely that philosophical arguments to this effect will have much of a practical impact.

If we want to assess the force of this argument, then we must know what causes the lack of concern about preserving autonomy. In this regard, one of the most plausible explanations may be what the psychologist Robert Levine has called ‘the illusion of invulnerability’. As part of this illusion, most people (at least in the West) not only think that they are less prone than average to diseases and earthquakes, but also that they are less prone than average to manipulation. For example, Levine and colleagues found that whichever feature of magazine advertisements was changed (e.g. product, style or content), subjects were adamant in claiming that they were not influenced by it. However, subjects readily agreed that others might be influenced by the advertisements. If there is indeed such an illusion of invulnerability, then it will indeed be very hard for people to be concerned about violations of their autonomy. Even if they accept an argument to this effect intellectually, they may not worry too much about this because of a deeply held conviction that they are invulnerable.

If the illusion of invulnerability is indeed what blocks the emergence of a concern for autonomy violation, then I think that there is reason for optimism. For under the influence of research in neuroscience, notions of conscious control are now under serious pressure. With regard to persuasion for example, researchers recently managed to predict the effectiveness of persuasive messages in terms of triggering behavior up to a week in the future more accurately on the basis of measuring brain activity than on the basis of self-report measures. That is, when they asked participants one week after exposure to a number of persuasive messages how much sunscreen they had used, the reported behavior was explained significantly more accurately on the basis of earlier measurements of activity in the medial prefrontal cortex (upon seeing the persuasive messages) than by what subjects themselves had predicted (after exposure to these messages). As doubts about conscious control – that are fueled by this experiment and many similar ones – are trickling down in society, people may be expected to become more open to the possibility that they are being influenced, and hence to become more interested in protecting their autonomy.
English Summary

If the liberal state is to protect autonomy and if branding undermines autonomy, then it follows that the liberal state should protect the autonomy of its citizens against branding. In my thesis I claim that the first premise of this argument is to be accepted and that the second is conditionally true. On this basis I conclude that there is presently an obligation for the liberal state to protect the autonomy of its citizens from several types of branding. Such protection can be realized either by consent procedures or by a ban on autonomy violating branding. I remain uncommitted as to which of these two forms of protection is most appropriate.

In support of the first premise I rely on a recent argument by Ben Colburn for thinking that all anti-perfectionists must be in favor of autonomy promotion and protection (provided they are not against any state action) and that all perfectionist defenses of liberalism are seriously unattractive. I claim that this argument is incomplete though in that Colburn does not give guarantees for the lack of a spillover of first-order values (which make sense in relation to states in the world) into the second-order value autonomy (which makes sense in relation to a value generation process). As such a guarantee, I argue, would count a conception of minimal autonomy that is inclusive of most theories and intuitions about autonomy, since such a concept can reasonably be assumed to be sufficiently free of perfectionist bias. In the course of my argument for the conditional truth of the second premise I develop such a concept of autonomy.

The merits of this second premise – that branding undermines autonomy – have been discussed extensively in the business ethics literature. However, those engaged in this debate have so far cherry picked concepts of autonomy, discussed only individual advertisements and failed to compare autonomy violation in branding with autonomy violation in similar institutions. In response to these three deficiencies I have mapped out an alternative strategy. First of all, the most relevant type of autonomy, at least from a political perspective, is the one that is minimally required to secure liberal legitimacy. Hence, a normative analysis of branding requires that one first construct an independent argument for such a minimal conception, which can then be applied to branding. Second, although it would be useful in itself to have a framework for evaluating individual branded communications, the state cannot realistically regulate such communications on a case by case basis. A comprehensive strategy must therefore identify structural features of branding and relate these to minimal liberal autonomy. Third, these structural features allow one to systematically compare autonomy violation in branding with autonomy violation due to similar institutions such as religious organizations and political parties.

As the first step of this strategy I identify, on the basis of a review of the relevant literature, seven core intuitions around which philosophical theories of autonomy have been built. Because most of these intuitions are highly controversial, I am careful to avoid making theo-
tical commitments where I don’t have to and eager to use empirical research in order to limit the range of reasonable controversy. In doing so, I aggregate five intuitions into the conception of self-control, understood as the ability of the agent to appropriately regulate her actions and the absence of control by other agents. I then argue against a number of philosophers who believe that this is all there is to autonomy, on the basis of strong intuitions in favor of including the notion of access to relevant ‘self’-related mental states (at both the encoding and the retrieval stage), which I refer to as authenticity. For both self-control and authenticity I detail minimal empirical conditions with regard to action and mental state formation. On the basis of the resulting four types of thresholds of autonomy I am able to identify several types of autonomy violating processes, which include unconscious goal activation, negative affect induction, subliminal attitude change and covariation learning (such as evaluative conditioning). However, these processes violate liberal autonomy only if they are widespread (because state intervention in general must be proportional) and if they are retrievable without violating basic privacy rights and have not been consented to (because state intervention that aims to protect one aspect of autonomy should not bring about more serious violations of another aspect of autonomy).

Instead of applying these standards to a sample of branded communications directly, I identify structural conditions of branding and investigate whether they alone or in conjunction generate pressure for brand managers to develop communications that violate minimal autonomy. Conditions are structural if they are historically stable elements of the environment in which branding occurs. A structural economic condition is the pressure to generate profits, which can explain by itself the use of unconscious goal and trait priming in shopping contexts. More important for my argument though are structural psychological conditions, such as stable higher-order human motivations. There are theoretical reasons and there is some empirical evidence for thinking that such stable higher-order motivations exist. Moreover, work in consumer psychology shows that consumers consistently respond to cues in branding in a way coherent with what one would expect to observe if there were stable higher-order motivations. These generate pressure for autonomy violation insofar as appeals to three of these higher-order motivations (i.e. mating, affiliation and status) induce conflict in the consumer and so give brand managers a structural incentive to use especially evaluative conditioning.

With minimal empirical standards for evaluating autonomy and relevant structural conditions of branding in place, I turn to work in consumer research in order to see whether the expected autonomy violation processes actually occur, both at the psychological and the socio-historical level. Experiments in consumer psychology concerning purchasing behavior and attitude change provide ample evidence for mental state and behavior change that is induced by means of evaluative conditioning and unconscious goal and trait priming. However, the evidence from the real world of branding – in particular, the historical record of advertising – is more difficult to interpret. I therefore generate hypotheses that at least make the claim that evaluative conditioning occurs on a large scale in the real world of branding falsifiable. Research in consumer psychology and history also inform the assessment of the three conditions for liberal autonomy violation. Whereas conditions for the centrality of effects and the retrievability of intentions are met, I remain undecided as to whether consent by consumers to autonomy violating branding can absolve the state from its obligation to protect the autonomous life slices of its citizens.

In the final stage of the argument a case is made for thinking that branding may be unique in violating autonomy for no good reason. I argue that the two main institutions that pose a threat to this status are political parties and religious organizations. Although political
parties deploy techniques similar to those used in commercial branding, such communications are legitimate in virtue of their support for political legitimacy which is generated by political communication. Branding offers no such indispensable contribution to liberal legitimacy, since market economies (which could be argued to be a necessary condition for the existence of a liberal state) will not collapse when branding efforts cease. Religion is different from branding because of a different consent context. That is, autonomy violating communications in religion are stable for structural reasons and can hence be properly consented to, whereas communications in branding are changing constantly.

Because I remain undecided as to which type of consent to autonomy violation should be required by the liberal state I offer two conclusions. Those who think that limited consent conditions suffice should demand that such minimal consent procedures are integrated into branding contexts. Those who think that extensive consent conditions are required should be skeptical about the possibility of such consent procedures in branding, due to its rapidly changing content and form. If this skepticism is justified, then they must support a ban on autonomy violating branding.
Als de liberale staat autonomie behoort te beschermen en merkenbouw autonomie ondermijnt, dan volgt daaruit dat de liberale staat de autonomie van haar burgers moet beschermen tegen merkenbouw. In deze dissertatie beweer ik dat de eerste premisse van dit argument geaccepteerd moet worden en dat de tweede voorwaardelijk waar is. Op basis van deze beweringen concludeer ik dat de liberale staat de autonomie van haar burgers behoort te beschermen tegen verschillende vormen van merkenbouw. Deze bescherming kan worden gerealiseerd door middel van toestemmingsprocedures en door middel van een verbod op merkenbouw die autonomie ondermijnt. Welke van deze twee maatregelen het meest gepast is laat ik open.

Als bewijs voor de eerste premisse voer ik een argument op van Ben Colburn. Hij beweert dat anti-perfectionisten voorstanders moeten zijn van de bevordering en verdediging van autonomie (wanneer ze niet tegen iedere vorm van staatsinterventie zijn) en dat alle perfectionistische onderbouwingen van het liberalisme onaantrekkelijk zijn. Ik stel dat dit argument onvolledig is omdat Colburn geen garantie geeft voor de afwezigheid van contaminatie van de tweede orde waarde autonomie (die betekenis heeft in relatie tot een waardengeneratie proces) door eerste orde waarden (die betekenis hebben in relatie tot toestanden in de wereld). Zo'n garantie zou gegeven kunnen worden door een concept van minimale autonomie dat een maximale hoeveelheid intuities en theorieën over autonomie in zich bergt, omdat we redelijkerwijze kunnen aannemen dat een dergelijk concept van autonomie voldoende vrij is van perfectionistische vooroordelen. In de loop van mijn argument voor de waarheid van de tweede premisse ontwikkel ik een dergelijk concept van autonomie.

Het vaststellen van de merites van deze tweede premisse – dat merkenbouw autonomie ondermijnt – is het onderwerp van een uitvoerig debat in de bedrijfsethiek. De bijdragen aan dit debat zijn echter niet overtuigend. Zo wordt de keuze voor een concept van autonomie niet onderbouwd, worden reclameboodschappen per geval geanalyseerd en wordt de ondermijning van autonomie in reclame niet vergeleken met de ondermijning van autonomie in andere domeinen. Als antwoord op deze drie gebreken stel ik een alternatieve argumentatiestrategie voor. Ten eerste is het meest relevante type autonomie – in ieder geval vanuit politiek perspectief – het type dat liberale legitimatie genereert. Een normatieve analyse van merkenbouw moet daarom allereerst een onafhankelijk argument voor een minimaal concept van liberale autonomie ontwikkelen dat in een volgende stap kan worden toegepast op merkenbouw. Ten tweede moeten structurele eigenschappen van merkenbouw worden geïdentificeerd, omdat de liberale staat niet voor bijvoorbeeld iedere reclameboodschap kan bepalen of autonomie al dan niet wordt ondermijnd. De identificatie van structurele eigenschappen van merkenbouw maakt het, ten derde, mogelijk om de ondermijning
van autonomie in merkenbouw te vergelijken met de ondermijning van autonomie in vergelijkbare domeinen zoals religie of politiek.

Ik zet de eerste stap van deze strategie door zeven centrale intuïties over autonomie te isoleren waaruit filosofische theorieën van autonomie zijn opgebouwd. Omdat de meeste van deze intuïties zeer controversieel zijn, leg ik me zo min mogelijk vast in theoretisch opzicht en gebruik ik empirisch bewijs om de ruimte voor onenigheid te beperken. Ik voeg vijf intuïties samen onder de noemer zelf-controle, dat ik begrijp als de mogelijkheid van een persoon om haar handelingen op gepaste wijze te reguleren en de afwezigheid van controle door een andere persoon. Ik keer me tegen een aantal filosofen die denken dat hiermee alles gezegd is over autonomie, omdat er sterke intuïties bestaan om toegang tot relevante mentale toestanden van het ‘zelf’ (zowel tijdens de codering als het afroepen van mentale toestanden) als onderdeel van een minimaal autonomiebegrip op te vatten. Ik noem dit aspect van autonomie authenticiteit. Voor zowel zelf-controle als authenticiteit geef ik minimale empirische condities aan met betrekking tot handelingen en de formatie van mentale toestanden. Op basis van deze vier soorten criteria kan ik verschillende processen waarin autonomie ondermijnd wordt onderscheiden, waaronder onbewuste doelactivatie, het oproepen van negatieve emoties, subliminale invloed en covariatie leren (zoals evaluatieve conditionering). Deze processen ondermijnen liberale autonomie echter alleen dan wanneer ze voldoende centrale effecten hebben (omdat de liberale staat in het algemeen proportioneel moet handelen) en wanneer de communicatie te reconstrueren is zonder basale privérechten te schenden en de ontvanger niet ingestemd heeft met de invloed (omdat de liberale staat een bepaald aspect van autonomie niet behoort te beschermen wanneer daardoor een ander aspect van autonomie meer gecompromitteerd wordt).

Om de aard van autonomie-ondermijning in merkenbouw te bepalen heb ik onderzocht of de structurele condities van merkenbouw druk uitoefenen op merkenbouwers om boodschappen te ontwikkelen die autonomie ondermijnen. Condities zijn structureel wanneer ze historisch stabiele onderdelen zijn van de omgeving waarin merkenbouw plaatsvindt. Een structurele economische conditie is de druk om winst te genereren, en deze volstaat op zichzelf om te verklaren waarom onbewuste doel- en eigenschapactivatie worden gebruikt in winkels. Belangrijker nog zijn structurele psychologische condities, zoals stabiele complexe motivaties. Er zijn theoretische redenen om aan te nemen dat zulke motivaties daadwerkelijk bestaan, en onderzoek in de consumentenpsychologie suggereert dat mensen in een consumptiecontext zich gedragen zoals men zou verwachten wanneer de theoretisch gespecificeerde motivaties bestaan. Mijn argument is dat de consumentenconflicten met stabiele sociale normen of eigen mentale toestanden waarneemt wanneer merkenbouwers openlijk aan drie stabiele complexe motivaties appelleren, en dat dit merkenbouwers een structurele prikkel geeft om communicatievormen zoals evaluatieve conditionering te gebruiken die autonomie ondermijnen.

Op basis van de minimale empirische standaards voor de evaluatie van autonomie en de relevante structurele condities voor merkenbouw bepaal ik of autonomie daadwerkelijk ondermijnd wordt in merkenbouw aan de hand van bevindingen uit de consumentenpsychologie. Experimenten in deze discipline laten zien dat het mogelijk is om mentale toestanden en gedrag van consumenten te veranderen met behulp van evaluatieve conditionering en onbewuste doel- en eigenschapactivatie. De feiten over de socio-historische ontwikkeling van merkenbouw zijn minder eenduidig en ik ontwikkeld daarom niet-triviale hypotheses die in ieder geval mijn stelling kunnen falsificeren. Op basis van consumentenpsychologie en socio-historische data stel ik vast dat de effecten van merkenbouw op consumenten een centraal karakter hebben en op legitieme wijze te reconstrueren zijn. Of
toestemming van consumenten voor de ondermijning van hun autonomie in merkenbouw de liberale staat van zijn plicht ontslaat om autonomie te beschermen is een vraag die ik open laat.

In het laatste deel van mijn argument beweer ik dat merkenbouw uniek is in het ondermijnen van autonomie zonder dat daarvoor een goede reden bestaat. Mijn argument is dat politieke partijen en religieuze organisaties de meest plausibele kandidaten zijn voor dezelfde status, maar dat beide van merkenbouw verschillen in een cruciaal opzicht. Hoewel politieke partijen dezelfde technieken gebruiken als merkenbouwers zijn deze technieken in het eerste geval legitiem omdat ze onderdeel uitmaken van het politieke discours van een democratische samenleving en aldus aan liberale legitimiteit bijdragen. Er is geen soortgelijke goede reden die merkenbouw legitimeert, omdat de economische basis voor een liberale democratie niet zal instorten wanneer merkenbouw zou verdwijnen. Religie is anders dan merkenbouw vanwege de context waarin toestemming wordt gegeven aan autonomie-ondermijning. De autonomie-ondermijnende boodschappen in een religie zijn stabiel vanwege de rol van traditie in religies zodat mensen hiermee kunnen instemmen, terwijl boodschappen in merkenbouw telkens veranderen.

Omdat ik open laat welk type van toestemming tot blootstelling aan autonomie-ondermijnende praktijken als standaard moet dienen voor de liberale staat kom ik tot twee conclusies. Zij die denken dat beperkte toestemmingscondities volstaan moeten ervoor pleiten dat minimale toestemmingsprocedures worden ingebed in de context van merkenproliferatie. Zij die denken dat uitgebreide toestemmingscondities nodig zijn moeten sceptisch zijn over de mogelijkheid van zulke procedures vanwege de snel verschuivende vorm en inhoud van merkenbouw. Als dit scepticisme gerechtvaardigd is, dan moeten zij een verbod op autonomie-ondermijnende merkenbouw ondersteunen.
Notes

INTRODUCTION

1 I understand branding as the sum of activities that are intended to communicate a preconceived brand image. These activities include advertising, public relations, direct marketing, sales promotion, product placement, sponsorship and guerilla marketing.

2 The section on autonomous mental state formation can be read as a response to skeptical remarks about the possibility of a normative analysis of autonomy and advertising (which is a subset of branding that influences consumers via the formation or reinforcement of mental states) that one finds in for example Joseph Heath, “Liberal Autonomy and Consumer Sovereignty”, in: John Christman and Joel Anderson ed., Autonomy and the Challenges to Liberalism (Cambridge 2005), 204-225, there 214. Heath argues that it would be “deeply implausible” that one could make a reasonable distinction between authentic and inauthentic desires, and that even if some such distinction could be made, there is no serious prospect of finding a consensus on how this is to be done. At this point I want to observe only that one ought to at least consider the proposed theories about authentic desires before making any judgments about their plausibility, let alone the prospects for reaching a consensus about any one proposal (or about shared aspects among proposals). This is what I will do in chapter 2.

CHAPTER 1

1 This means that the overall argument in this thesis does not speak to those who are committed to fascist, communist or a number of communitarian legitimations for state action.

2 See Ben Colburn, Autonomy and Liberalism (New York 2010), 45 ff.

3 See for example Galston, who says that “the promotion of personal autonomy, understood as choice based on critical rationalism, is not among the shared liberal purposes. Autonomy is one possible mode of existence in liberal societies – one among many others”, Liberal Pluralism. The Implications of Value Pluralism For Political Theory and Practice (Cambridge 2002), 24.

4 See Ibidem, 50 ff.

5 I will note one crucial condition under which this assumption is reasonable later on.

6 Ibidem, 58.

7 He also discusses relativism as a justification for anti-perfectionism, but this position is sufficiently implausible that nobody defends it, and I will therefore not discuss it here. See Ibidem, 61-62.

8 Weak pluralism cannot hope to do so, because it only maintains that there can be two options that are both worth pursuing. This does not get one to the position that perfectionism is wrong. Ibidem, 63.

9 Ibidem, 66.

10 In a similar vein, John Christman writes that “the claim that just principles must be accepted as legitimate by those living under them assumes a kind of being who makes such judgments of legitimacy. The precise conception of the autonomous person, then, will be crucial in the determination of the plausibility of principles of justice.” See Social and Political Philosophy. A Contemporary Introduction (London and New York 2002), 119.

11 If this argument is correct than Rawls’ proposal not to construct a theory of justice as fairness on the basis of (comprehensive) autonomy because that value might not be accepted by all would amount to saying that the liberal state should not promote and protect (comprehensive) autonomy because by doing so it would fail to protect (comprehensive) autonomy. See for Rawls’ argument Political Liberalism (New York 2005), xlii-xliii and 78.


13 Colburn, Autonomy, 119.

14 One can also argue that diversity is valuable because of other values such as tolerance. However, if tolerance is said to be intrinsically valuable, then there is no way in which this value can be explicated without involving autonomy, and if tolerance is said to be of instrumental value, then we need another value. See Ibidem, 111.


16 Ibidem, 154.

Colburn’s own account of autonomy as hypothetical reflective endorsement of values certainly does little to disconfirm the suspicion that it is based on a first-order value, *Autonomy*, 25. He allows – officially – for unreflective lives to count as autonomous as well if people would endorse such a life if they were to reflect on it. However, this insertion does not make a practical difference because he fails to identify state policies that can distinctively foster hypothetical endorsement. Ibidem, 38. Thus, if it is true that the state should promote autonomy and it can do so only by fostering reflection (which may or may not foster hypothetical endorsement), then it will promote a controversial ideal of the contemplative life. Many will balk at this, and they may do so not only on the basis of first-order values, but also on the basis of a different account of (the second-order value) autonomy.

In this regard it will not do to claim that autonomy is different from rugged individualism, as in Robert Young, *Personal Autonomy. Beyond Positive and Negative Liberty* (London 1986), 5 and Gerald Dworkin, *The Theory and Practice of Autonomy* (New York 1988), 8. Dworkin takes his ‘ideological neutrality’ constraint on theorizing about autonomy to be a weak one, which in practice means that the theory should not appeal only to those who hold an individualist worldview. At least for a liberal conception of autonomy this constraint is too weak.

Although the main debate on the issue over autonomy and advertising is played out in this journal, there are noteworthy contributions in the business ethics literature that have appeared elsewhere. Richard Lippke argues in *Radical Business Ethics* (Lanham 1995) that persuasive advertising (as distinguished from informational advertising) “suppresses” autonomy because it keeps consumers from reflecting on choices by “implicit” persuasion (108). Although Lippke lists some intrusive phenomena such as group pressure and emotional appeals, this list is argumentatively fruitless, as he has not specified what level of autonomy ‘suppression’ is normatively relevant, nor how this level is to be defined. Moreover, the distinction between informational and persuasive advertising is underdetermined, which impairs the practical relevance of his argument.


A better defense of Arrington’s position would be to claim that heuristic decision making can be rational as well (and hence autonomous), and that puffery offers valuable input to such mechanisms.

Ibidem, 10.


Arrington elsewhere argues that subliminal influence need not worry us because “most of us have a benevolent subconsciousness which does not overwhelm our ego and its reasons for action”, see his “Advertising”, 10. I have no idea what ‘subconsciousness’ refers to here, but I will show later on that scientific psychology demonstrates that unconscious influences on action are real (see especially chapter 2.2).


Crisp could alternatively make the definition of a good reason dependent on normative standards for judgment and decision making in a specific domain. Making good aesthetic decisions may not be hampered by the unconscious arousal of desires whereas good economic decision-making may be hampered by manipulation. However, it seems problematic for a theory of autonomy to stipulate which type of decision making is appropriate in a certain domain and what counts as appropriate input to it.


Ibidem, 21.

Ibidem, 22.

I discuss the merits of this critique in chapter 2.


Ibidem, 233.

See chapter 2.4.

Ibidem, 234.

The latter point counts not so much against Cunningham’s assertion as against Noggle’s original theory.

These doubts about the desirability of this standard of autonomy are discussed in chapter 2.4.
CHAPTER 2

1 It can be objected that not all intuitions about autonomy should be weighed equally. Although I agree with this objection in principle, I also believe that a minimal approach should not take differences in weight as a starting point. Instead, it should first distinguish intuitively plausible conceptions of autonomy from obviously implausible ones that are confused with it. The plausible conceptions, among which some are prioritized by different lines of thought, should be treated equally, even if lines of thought have not been equally influential. Only in case of conflict between such conceptions should different weights be allocated. This should be done by means of thought experiments in which variables are switched until conceptions make contradictory evaluations about such cases. The evaluation (and its supporting conception) that is best in line with intuition should win out. In sum, differences in weight between plausible conceptions of autonomy are in order, but they should be the outcome of analysis, not its starting point. Note also that my approach is different from the approach to minimal liberal autonomy taken by Veit Bader, who maintains that all conceptions about which controversy exists (such as the split-self view of autonomy) are to be avoided. I have decided against such an approach because the resulting notion of “minimalist individual agency” that figures in Bader’s account is so minimal that I am afraid that few would still recognize it as being about autonomy. Bader, Secularism or Democracy? Associational Governance of Religious Diversity (Amsterdam 2007), 75.

2 These two criteria – of consistency and empirical feasibility – were introduced as conditions for the adequacy of a theory of autonomy in general by Dworkin, The Theory and Practice of Autonomy, 7. Another one – ideological neutrality – was already discussed in chapter 1 and is understood stricter in the form of the first principle of maximal inclusiveness outlined above. The subsequent relevant criteria that Dworkin mentions are already met. The normative relevance of the theory is guaranteed by its role in legitimizing the liberal state; the value of the concept of autonomy (for this end) was demonstrated in chapter 1; and ‘judgmental relevance’ (or reflective equilibrium) is part of my first principle. Judgmental relevance is used throughout this chapter but especially and explicitly so in section 4.


5 Daniel Dennett argues that such brainwashing thought experiments are inconsequential, because they might be psychologically impossible. And even if they were possible, he continues, we might be able to quickly regain the capacity to relate to the world autonomously. If so, then history might matter only “because in the actual world only certain sorts of historical processes can create what really matters: the set of competences implicit in the total structure of an embodied nervous system.” Dennett, “Natural Freedom”, in: Metaphilosophy, vol 36 no 4 (2005), 449-459, there 450. I address these concerns extensively in section 2.2 and argue that they are not justified.

6 Christman describes this view in “Introduction”, 13.


8 Let me briefly elaborate this claim. First, there are conceptions that clearly carry first-order values into the definition of autonomy and hence fall outside of the scope of this chapter. In this category one finds conceptions of the heroic person (Nietzsche), the rationalist person (Kant), the creative-expressive person (Von Humboldt) and the individualistic person (Mill). Second, in non-liberal conceptions of autonomy, such as anarchism, no authority can be acknowledged except for one’s own, and this is incompatible with liberalism’s commitment to state institutions. See Marina Oshana, Personal Autonomy in Society, 15-19 on types of political autonomy. Finally, there may be some relation between moral autonomy (understood as responsibility) and whatever liberal autonomy may turn out to be, but such a relation should not be presumed at the outset.

9 Intuitions 3, 4 and 5 are identified as the three central ideas for thinking about autonomy in the overview given by Sarah Buss; Marina Oshana identified 1, 5, and 7 as central intuitions (among others that I will discuss shortly), Personal Autonomy in Society, 76 ff; on the list compiled by Nomi Arpaly in “Which Autonomy?” we find the intuitions 4, 5 and 7 (again, among several others that are not relevant to liberal autonomy). The intuitions that autonomy is about a point of view – intuition 2 – is shared by Gary Watson, Harry Frankfurt,
Michael Bratman, David Velleman and others but does not figure explicitly in the above reviews. Similarly, that autonomy involves self-control understood as impulse control – intuition 6 – is taken for granted in all accounts (but most extensively so in Mele, *Autonomous Agents*) and cannot be reduced to one of the other intuitions.


12 Such an argument has been developed by Joel Anderson and Axel Honneth, “Autonomy, Vulnerability, Recognition and Justice”, in: Joel Anderson and John Christman, *Autonomy and the Challenges to Liberalism*, 127-140. Anderson and Honneth in fact distinguish three types of self-esteem – drawing on earlier work by Honneth – which are yet to be empirically validated.

13 A more comprehensive assessment of branding and liberal autonomy could pursue this line of argument and might significantly strengthen my conclusion on the nature of autonomy violation in branding.

14 Marina Oshana for example argues that self-esteem (or “self-respect”) understood as the “inherent dignity or worth” of humans would be constitutive of autonomy and that a number of social-relational properties such as security and financial self-sufficiency would do so as well, see *Personal Autonomy in Society*, 81 and 86-87 respectively.

15 Intuitive support for thinking of positions in social relations as being constitutive of autonomy comes from cases such as voluntary slavery which are supposed to show that such restrictive social relations impede autonomy. This intuition is not decisive however, because one can imagine a slave who resists her fate by the small opportunities for resistance that are open to her. Thus, it seems more parsimonious to say that autonomy is what an agent exercises given the social relations that she finds herself in (as long as the seven core components are not interfered with by others). See Oshana, *Personal Autonomy in Society*, 43 f. The resisting slave example is presented in Berofsky, *Liberation From Self*, 22. Diana Tietjens Meyers makes a related point in *“Personal Autonomy in Society (review)”, in: Hypatia, vol 23 no 2 (2008), 202-206*.

16 Oshana believes that this is not a problem, because she thinks that the liberal state ought to be legitimate on the basis of a perfectionist conception of autonomy, *Personal Autonomy in Society*, 105. I have rejected this view in chapter 1.

17 See on the problems of defining which social relations are relevant for autonomy also Meyers, “Personal Autonomy in Society (review)”. For an account of the complexity involved in specifying social roles see Veit Bader and Albert Benschop, *Ongelijk-heden. Sociale ongelijkheid en sociaal handelen* (Groningen 1988), 120 ff.

18 See for other accounts of such a distinction between freedom and autonomy Berofsky, *Liberation From Self*, 16 ff and Christman, “Autonomy, History and the Subject of Justice”, 20.

19 Some intuitions in my taxonomy may be seen as having nothing to do with autonomy (which is ultimately a matter for experimental philosophy to deal with) and it is of course also possible that I failed to see an intuition underlying conceptions of autonomy (or that the philosophers who figure in reviews have yet to come to grips with such an intuition). The best I can do is point out in response to the former concern that I include intuitions on the importance of which at least philosophers agree (which is admittedly a highly unrepresentative sample) and that I rely on thought experiments to bring out intuitions for the reader in case there is disagreement.


23 Before I move on, it is instructive to stress that I use self-control differently from its typical meaning in social psychology. The latter meaning, which also figures in Alfred Mele’s discussion of the concept for example, refers to the opposite of weakness of will. Thus, self-control in this sense is but a component of the way in which I use the term here. A much more closely related meaning is ascribed to self-control by Julius Kuhl. In his theory, self-control refers to two subsystems that he locates in the left hemisphere and that I describe in more detail in section 3.1. Although self-control in his theory refers to similar processes that I describe here my use of self-control leaves open the precise instantiation of these functions.


25 Watson in fact speaks of free actions instead of autonomous actions, but what he refers to is what I take to be autonomy (and so is his theory interpreted generally in philosophical discussion of autonomy).


27 Ibidem, 208.


30 Ibidem, 158 and 396-397.
31 Ibidem, 391.
33 Ibidem, 234.
34 Ibidem, 232. Bratman moreover argues that such feelings must have the support of higher-order attitudes in favor of treating a feeling as a reason for action, but that is a claim that we discuss in the next section.
35 Berofsky, Liberation From Self, 82 ff.
36 Ibidem, 93 ff.
38 Ibidem, 16.
40 The general response of hierarchy theorists has been to say that only second-order desires can be said to speak for the self. However, this assertion is at least not self-explanatory. Berofsky for example mobilizes a host of intuitions against thinking that the second-order level has a special claim to speaking for the self. See Liberation From Self, 99-106. If this is correct, then the hierarchical theorist ought to explain what it is about second-order desires that allows us to say that they speak for the self. The only argument to this effect has been made by Bratman. He assumes that due to their stable role in referring to the agent’s past and future, self-governing policies give rise to a feeling of identity and can therefore be said to speak for the self. Hence, if desires are treated as justified reasons because of self-governing policies, then they can be said to receive support from the self. Bratman, “Reflection, Planning, and Temporally Extended Agency”, in: Taylor ed., Personal Autonomy, 33-57, there 46. However, Bratman acknowledges that such identity conferral is a property of policies in general (and not a prerogative of self-governing policies). Thus, on Bratman’s account a policy that tells us which considerations are justified reasons for action (as opposed to which desires are justified reasons for action) would also confer autonomy. Bratman, “Planning Agency, Autonomous Agency”, 48. And such policies do not have a conative hierarchy. Bratman’s reply to this objection is to point to the role of self-governing policies in self-management, which would exert pressure in the direction of the formation of self-governing policies and hence autonomy. But I argue below that this pressure does not amount to much.
42 Ibidem, 75.
43 Ibidem, 86.
48 See Carruthers, The Architecture, 393-394 for a more extensive critique of Bratman along these lines.
50 If this correct, then Bratman’s argument for “a kind of identification” of the two pressures in “Autonomy and Hierarchy”, Social Philosophy and Policy, vol 20 (2003), 156-176, there 167 is not relevant.
51 Dworkin, The Theory and Practice, 15. He actually requires the capacity for reflection that is to result in the identification with desires, but that extension is inmaterial to the general approach that he takes.
52 See Colburn, Autonomy, chapter 1.
53 See Also Oshana, Personal Autonomy in Society, 78 and for a more extensive discussion Berofsky, Liberation From Self, chapter 6. The precise limits of minimal rationality do not concern me here because there is no relevant disagreement on the issue in the literature on autonomy and because we saw in chapter 1.2 that a critique of branding that is based on standards of rationality does not have realistic prospects of success.
54 Carruthers, The Architecture, 391-393.
55 It is not clear what is going on when rival desires are suppressed, but social psychologists have quite recently found that when people exert self-control – in terms of inhibition of action programs – across very diverse tasks (e.g. attention control, thought suppression, emotional regulation and pro-social behavior) we may tap into a single resource. That is, if we exert self-control in resisting the temptation to eat a pudding, then that effort reduces our ability to exert self-control in solving a puzzle in a follow-up task. See Mark Muraven and Roy Baumeister, “Self-regulation and Depletion of Limited Resources: Does Self-Control Resemble a Muscle?”, in: Psychological Bulletin, vol 126 (2000), 247-259 for a review of evidence for the limited resource model of self-control.
56 Mele’s account in *Autonomous Agents* is of no use in this regard, as he is interested in the criteria for perfect self-control rather than a threshold value.


58 Ibidem, 343.

59 Ibidem, 344.

60 To be precise, the authors do not require every neurochemical to be in a preset range, but merely require a control space condition to be satisfied, so that there are many combinations of levels of neurochemicals that fall within this range.

61 I do not think that self-control, understood as the opposite of weakness of will, is a phenomenon that is usefully defined without any reference to causality. This is the proposal of Dorothee Horstkötter, who argues that if we want to maintain a viable concept of self-control in light of automaticity research – which shows that much of our behavior is triggered outside of our awareness – we must define self-control normatively. That is, if our behavior happens to cohere with a goal that we have set ourselves, then we exercise self-control irrespective of the causal chains that led to the behavior. Although setting a goal is certainly a necessary condition for exercising self-control, so are the causal links between the goal state and the activation of the motor schema. Otherwise we would have to say that we exercise self-control if a (benign) mindreading genius reads our goals and activates the relevant motor schema with the help of some futuristic remote controlling device. And that seems wrong. See Dorothee Horstkötter, *Self-control Revisited. Varieties of Normative Agency* (Enschede 2009).

62 Dennett, *Elbow Room*, 52 ff. I diverge substantially from Colburn’s definition of independence, which does not allow for agents to act on reasons they are not aware of. See his *Autonomy*, 28. This position is not tenable if the research on awareness discussed in section 2.3 is correct.


64 This is the case at least until consciousness is understood.

65 See Ibidem, 95 ff. for a discussion of this experiment and similar experiments.


68 Mele, *Autonomous Agents*, 182-185 even talks about manipulating the reasoning process by instilling habits of thought by means of conditioning.


71 See Kathleen Taylor, *Brainwashing. The Science of Thought Control* (Oxford 2004), 204 on the threats of manipulating reactance. She understands reactance as a feeling of not being in control of actions, which refers to a different cluster on questionnaires. The point on manipulation carries over however.

72 Berofsky, *Liberation From Self*, 185 ff. See also Dennett, *Elbow Room* and *Freedom Evolves*.

73 I also understand epistemic competence differently from Oshana, *Personal Autonomy in Society*, 76-78, who stresses the supportive role of motivations for epistemic competence with regard to self-assertive agency.

74 Alfred Mele has helpfully distinguished three types of such control. A controller may influence another agent by controlling information, by instilling or changing pro-attitudes or by impairing the agent’s reasoning system.

75 See Daniel Dennett, “Natural Freedom”, 450.


77 The journalist Naomi Klein reports on these experiments extensively in *The Shock Doctrine. The Rise of Disaster Capitalism* (New York 2007), 25 ff.

78 Attitude change effects by means of subliminal influence have been reported with regard to the mere exposure effect – where we like an object, in this case a subliminally presented one, more when we are confronted with it more often – and to evaluative conditioning. Dijksterhuis et al., “The Power of the Subliminal”, 84 and 90 ff.


See on this assumption Betram Gawronski, “Ten Frequently Asked Questions About Implicit Measures and Their Frequently Supposed, But Not Entirely Correct Answers”, in: Canadian Psychology, vol 50 no (2009), 144-150, there 144.


Other evidence for this view comes from eye tracking measurements and a number of manipulations. The pattern of eye movement expected to facilitate EC effects would, according to the IMM, be one in which the eye gaze was fixed on the US and then shifted to the CS (or the reverse) without intermediate fixation. If eye gaze can be assumed to indicate the internal representation of stimuli, then this pattern should lead to a temporally proximate representation of US and CS, which should in turn facilitate EC effects. The latter effect did indeed correlate with the described eye pattern. Christopher Jones, Russell Fazio and Michael Olson, “Implicit Misattribution as a Mechanism Underlying Evaluative Conditioning”, in: Journal of Personality and Social Psychology, vol 96 no 5 (2009), 933-948, there 936 f. More evidence for the IMM comes from several manipulations in which the US and CS were flashed in an alternating pattern (facilitating eye shifting between the two), in which the space between US and CS was decreased and in which the temporal overlap was increased, all of which correlated with greater EC effects. Ibidem, study 2 and 3; Jones et al., “Evaluative Conditioning”. It not clear whether alternative EC mechanisms could not also explain these effects however.

Jones et al., “Implicit Misattribution”, 942 f.


Jones et al., “Evaluative Conditioning”.


Alan Ware argues that for manipulation to occur it must be true that the manipulator “could reasonably have been expected to know” that her actions would have manipulative effects. See “The Concept of Manipulation: Its Relation to Democracy and Power”, in: British Journal of Political Science, vol 11 (1981), 163-181, there 175.

Whether scientifically less well informed controllers can be said to have such knowledge as well would have to be established on a case by case basis.


Jones et al., “Evaluative Conditioning”.


People are motivated to process deliberatively when the need for accuracy is high, when a potential cost is identified while the need to avoid it is high, and when there is a strong need to avoid prejudiced reactions.


101 See Mele, Autonomous Agents, 153-154 on the concept of sheddability. Mele does not detail the nature of the objects of such shedding, nor the conditions under which shedding can be successful.

102 Bertram Gawronski, Roland Deutsch, Sawsan Mbirkou, Beate Seibt and Fritz Strack, “When ‘Just Say No’ is Not Enough: Affirmation Versus Negation Training and the Reduction of Automatic Stereotype Activation”, in: Journal of Experimental Social Psychology, vol 44 (2008), 370-377, there 374. It must be noted that this study also shows that implicit attitudes can be changed by building rival associations. This may work to some extent for stable stereotypes, such as racism, but it is not a practical strategy if one does not know which implicit attitudes have been changed (as is the case, I will argue, in branding).


104 Ibidem, 1104.

105 Surprisingly, the effect occurs mainly in the case of tasks that were ascribed by the experimenter (and not in the case of the left over items). It is not clear why this should be so.


107 See for example Ibidem, 170, 171, 178 and 186.


109 Berofsky, Liberation From Self, 99 ff.

110 This overview is based on Peter Carruthers, “How We Know Our Own Minds: The Relationship Between Mindreading and Metacognition”, in: Behavioral and Brain Sciences, vol 32 (2009), 121-138. See 138-182 for extensive open peer commentary on Carruthers's discussion and the latter's rebuttal.

111 Ibidem, 137. Autists are for example worse at assessing their own success in a ball game.

112 Ibidem, 125.


117 Does this mean that consciousness can no longer play a role in a theory of autonomy? Matt King and Peter Carruthers go a long way in this direction by arguing that the lack of consciousness settling anything in decision-making would be evidence against the endorsement approaches of Frankfurt and Watson. It is intuitively plausible to argue that second-order desires and best judgments make an agent autonomous on the basis of cases such as the addict who rejects her addiction. On Frankfurt's account, if the addict acts on the desire to take drugs against her best judgment (or second-order desire), then she does not act autonomously. King and Carruthers assume that this intuition is plausible only if the decision not to take drugs was made consciously. For if the decision-making itself were unconscious, then what remains is the conscious action of rehearsing in inner speech that it is not a good thing to take drugs. And if that would suffice, then as long as the unconscious decision-making process is caused by some conscious action we are autonomous. This would imply that if the addict's choice to take another shot would be caused by a conscious thought about her problems, she acted autonomously. And that, King and Carruthers argue, seems wrong. The problem with this argument is that it treats all actions as equal whereas they are not. If rehearsals that result in a best judgment can cause actions (when they are coupled with the desire to act on one's best judgments), then it can be argued that what makes a person autonomous are decisions which are the result of (conscious) best judgments that together with auxiliary innate desires outcompete others motivations. If this is correct, then there is a role for consciousness in a liberal theory of autonomy, be it a much more restricted one than is commonly assumed. Matt King and Peter Carruthers, “Moral Responsibility and Consciousness”, in: Moral Philosophy, vol 7 (2010), forthcoming; http://www.philosophy.umd.edu/Faculty/pccarruthers/Accessed September 22, 2010.


120 Berofsky makes such a point in relation to endorsement approaches to autonomy. He argues that if unconscious mental states are out of sync with a person’s explicit endorsement, and such unconscious states are thought to be a bona fide part of the self, then the endorsement theorist must take recourse to unconscious endorse-ment. However, it would according to Berofsky be impossible to distinguish cases in which unconscious endorsement occurred from ones in which it did not, so that this approach is a dead end. See *Liberation From Self*, 100. The same argument can be made if we replace endorsement with retrieval mechanism. At least with regard to unconscious retrieval I will argue that there is evidence from psychology for distinguishing cases of valid unconscious retrieval from invalid ones (as well as for conditions under which valid retrieval is inhibited). Whether such unconscious retrieval can count as a case of endorsement is a matter that I leave open.


122 In an earlier account, Velleman proposed that the spontaneity of the generation of predictions would account for the autonomy of agents. However, randomly generation is unlikely to count as a source of autonomous behavior on most people’s intuition, while this account is also unable to distinguish regular action from autonomous action.

123 Velleman, “The Self as Narrator”, 69.

124 Ibidem, 70.


127 See section 2.5.


129 This is in fact a disconcerting finding from the perspective of autonomy, as experienced difficulty in effectively carrying out an action like an experimental task is related to a readiness to take on higher level identities of the act, such as participating for selfish or altruistic reasons. Ibidem, 12.


132 The psychologists Richard Ryan and Edward Deci suggest that nonautonomous decisions which are unreflected would be followed by “guilt or shame”, but such emotions seem to be able of indicating many different things, see “Autonomy is No Illusion. Self-Determination Theory and the Empirical Study of Authenticity, Awareness and Will”, in: J. Greenberg, S. Koole and T. Pyszczynski ed., *Handbook of Experimental Existential Psychology* (New York 2004). 449-479, there 468.


134 This point is made by Berofsky, *Liberation From Self*, 102.

135 Note that claiming that a liberal theory of autonomy should not privilege the reflective processing mode (against nonreflective processing modes) is different from claiming that it should not privilege reflective utterances (against body language). The truth of the second claim does not imply that the first is true as well, since a reflective utterance can be informed by a reflective and a nonreflective self-consistency check. In fact, I will indicate how a nonreflective self-consistency mechanism may send an affective signal to the reflective processing system in section 3.2 where I discuss the work of Sascha Topolinski and Fritz Strack. The second claim is found in Christman, *The Politics of Persons*, 152.


138 Northoff et al., “Self-referential Processing”.


140 Ibidem, 100. It remains unclear what distinguishes a force within from one outside of the self however.

141 Ibidem, 104.

142 Ibidem, 105.


144 Berofsky, *Liberation From Self*, 94 ff. As Beate Rössler has recently pointed out, Frankfurt treats the externalization of wishes and normative imperatives as if they were convictions, but they are not, because wishes and normative imperatives do not lose their validity after a conflict in the way convictions do after they are either falsified or validated; see “Autonomie und Ambivalenz”, 372 ff.

145 An alternative mechanism is proposed by the psychologists Tim Kasser and Kennon Sheldon in “Nonbecoming, Alienated Becoming, and Authentic Becoming: A Goal-Based Approach”, in: Jeff Greenberg et al ed., *Handbook of Experimental Existential Psychology*, 450-493. They speculate that there is an Organismic Valuing Process (OVP) that gives the organism internal information to aid it in decision-making. Its output would be what we typically experience as intuition. Most importantly, the leading goal of the OVP is said to be the health of the organism (457). There are subsequent problematic assumptions about its hostility towards extrinsic goals (such as status and looks) that we need not go into that here. At this point I want to point out only that the postulation of the OVP is not only entirely speculative, but also unlikely to operate under the goal of health, because this is too general a goal to have evolved. It is possible that an innate fear of snakes is selected for, but unlikely that an abstract goal such as survival evolved.


149 Baumann and Kuhl in this regard found that state oriented individuals who reported a negative mood were less proficient in identifying coherent against incoherent word triplets (e.g. salt, deep and foam is coherent because all words are related to ‘sea’) than action oriented individuals with the same level of negative mood. Nicola Baumann and Julius Kuhl, “Intuition, Affect and Personality: Unconscious Coherence Judgments and Self-Regulation of Negative Affect”, in: *Journal of Personality and Social Psychology*, vol 83 no 5 (2002), 1213-1223.

150 Ideally one would want to see this research extended to the measurement of latencies on actual action choices instead of memory tasks. That is, one would like to see experiments in which it is shown that state oriented individuals in simulated choice situations are more likely to make choices for options that they previously rated as unattractive (especially when they are subjected with negative affect) and also show longer response latencies. Although there is no such research to date, that does not seem such a giant leap.


152 Specifying the level of negative affect is a task that I do not take up in this thesis.


155 Christman responds to this concern by noting that on his account of sustained critical reflection over a variety of circumstances (which I discussed in section 3.1) the assessment of the background values would be more realistic because there is no need to assess fleeting affective and cognitive reactions, while the assessment would moreover have to be applied only to basic organizing values which are presumably more easily detected. See, *The Politics of Persons*, 158. This may make the task of gauging a person’s mental states of (non-)alienation somewhat less improbable, but that is not the same as this task being realistically achievable.


157 What about cases in which we reflectively come to endorse a mental state (or the way in which we appropriated that mental state), but the mental state in question slipped past initiatory mechanisms? This question was posed to me by Joel Anderson. I think that both cases would indeed properly authenticate the mental state in question (because of the involvement of either coherence or the reflective point of view of the agent) and that retrospective reflective authentication would hence have to be added as a subsequent condition. This case is irrelevant to the present discussion though, because the stimuli that slip past the initiatory mechanisms (which I will discuss later) are brought about by evaluative conditioning of the misattribution type and
stored mainly as implicit attitudes. As such they are unlikely to ever become the object of reflection, and when they do, they are unlikely to be corrected since – as was discussed in section 2.2 – negating implicit attitudes at the explicit level often has the opposite effect.


159 Walther et al., “Changing Likes and Dislikes Through the Back Door”.

160 I should note at this point that a rival to the APE model is less categorical on this issue. On the Meta Cognitive Model (MCM) attitudes are not established on fine but read from memory. Moreover, in contrast to dual systems model of implicit influence, there is on the MCM but a single type of memory in which attitudes are stored. However, in this memory store, attitude objects are not just associated with an evaluation, but the evaluation itself may in turn be the object of a so called meta-cognitive tag. If the MCM is correct, then associative processing is simply less effortful processing and not a qualitatively different kind of processing (i.e. where one ascribes truth values and the other does not). However, the MCM should predict that conflict detection should be seriously impaired in associative processing, since fewer representations are taken into account so that less meta-cognitive tags – including ones with truth value – are considered. R. Petty, P. Briñol, and K. DeMarree, “The meta-cognitive model (MCM) of attitudes: Implications for attitude measurement, change, and strength”, in: Social Cognition, vol 25 (2007), 657-686.

Evidence for thinking that associative processing is at least subject to consistency constraints comes from a study by Bertram Gawronski and Fritz Strack that uses induced compliance. In such studies subjects are typically asked to write an essay that runs counter to one of their attitudes (e.g. blacks should pay more tuition). As a consequence, the subject is expected to experience a conflict between the position she actually holds (e.g. blacks are equal) and her behavior (i.e. writing that they are not). In the classical set up, one group is then exposed to high pressure to comply with writing the essay and another group to low pressure. The high pressure is reasoned to induce the belief that the behavior is only performed to please the experimenter. Combined with a belief such as ‘in order to please others you may temporarily engage in behavior that runs counter to your convictions’, this proposition resolves the conflict. Hence, under high pressure the initial proposition should not change. However, in the low pressure condition there is no such justification. The subject is therefore forced to either give up her behavior or to change her attitude, depending on what causes least resistance. Cognitive dissonance theory predicts that it is typically easier to change the initial belief, and this is what is generally found and what Gawronski and Strack found when they measured explicit attitudes. However, there was no change on implicit attitudes after dissonance reduction in the low pressure condition (while implicit and explicit attitudes were highly correlated in the high pressure condition). Betram Gawronski and Fritz Strack, “On the Propositional Nature of Cognitive Consistency: Dissonance Changes Explicit, But Not Implicit Attitudes”, in: Journal of Experimental Psychology, vol 40 (2004), 535-542. These and other experiments strongly suggest that automatic corrections for consistency do not carry over to implicit attitudes. For a discussion of these subsequent experiments see Bertram Gawronski, Fritz Strack and Galen Bodenhausen, “Attitudes and Cognitive Consistency. The Role of Associative and Propositional Processes”, in: Richard Petty, R.H. Fazio and Pablo Briñol ed., Attitudes: Insights From the New Implicit Measures (New York 2009), 85-117, there 94 ff.

161 Nicola Baumann, Julius Kuhl and Miguel Kazén, “Left-Hemispheric Activation and Self-Infiltration: Testing a Neurophysiological Model of Internalization”, in: Motivation and Emotion, vol 29 no 3 (2005), 135-163, there 147. An alternative explanation would be that subjects who were all right handed were made uncertain because of the left-hand ball squeezing and for this reason were more attentive at encoding. This explanation can be ruled out however, because retrieval was significantly better only for ascribed tasks.

162 Ibidem, 155.


166 On this inadvertent effect of discrepancy reduction, see Ibidem, 152. According to Petty and colleagues this effect explains why implicit-explicit discrepancy is so persistent.


170 See Beate Rössler, Der Wert des Privaten (Frankfurt am Main 2001) for an argument for thinking that we value privacy because of a concern with autonomy. See Robert Noggle, “Manipulative Actions: A Conceptual and Moral Analysis”, in: American Philosophical Quarterly, vol 33 no 1 (1996), 43-55, there 51 for another – practical – argument in support of the intuition that trivial cases of autonomy violation should not be remedied with state action. If the commitment to protect autonomy on the part of the state is part of the terms of a (hypothetical) contract, then at least one appropriate standard for determining what this commitment should amount to is to say that it must be enforceable (because that is a universal requirement of contracts). This can in turn be taken to mean that the state should be able to realistically intervene in cases of autonomy violation and that it is simply practically impossible for a state to monitor all interactions for the presence of manipulations.

171 Whether consent to autonomy violation is possible in principle is contested, most notably by Marina Oshana. She argues not only against the possibility of autonomously giving up what I call freedom (as in voluntary slavery) but also against the possibility of autonomously giving up (psychological) autonomy. Let me stress that these arguments are not equivalent. (This is not even true if Oshana’s conception of (social) autonomy is correct, since her argument is not that autonomy is only to be found in objective social relations, but that such social relations must be added to psychological states.) Thus, if it is legitimate to autonomously sell yourself to a slave holder, then this need not imply that you may autonomously hand your life to a hypnotist. Furthermore, it is at least pace Oshana – not obvious that a person cannot voluntarily sell herself into slavery. If this is correct, then any attempt to argue against ceding autonomy with reference to a continuum between voluntary slavery cases and cases in which autonomy is given up must fail. As an independent argument against the possibility of autonomously giving up autonomy, Oshana claims that in doing so an agent would always underestimate her moral status. Oshana, Personal Autonomy in Society, 57. (This idea, she notes, is also found in Jean Jacques Rousseau and Thomas Hill Jr.) If appreciating your moral status is a condition for autonomy and if that moral status is contingent on not ceding autonomy, then it follows that it is impossible to ever autonomously give up autonomy. This argument is also not convincing. Not only is it not clear why every autonomous act must be based on the (implicit) premise that one is a moral agent, since nothing in the concept of autonomy seems to call for such a requirement. It is also not clear why on a minimal conception of moral status it would be irreconcilable with giving up autonomy. There is no unambiguous intuition to this effect. Looking forward for a moment, I note that much hangs on this argument. If I am mistaken, and it is not possible to autonomously give up autonomy, then my argument against branding would have greater ramifications.

172 See chapter 5 on the psychological details of this process.

173 Sadly, the producer (Objective Productions) would not send me a copy.

174 See on this ambiguity also Dworkin, who initially held that one cannot be allowed to autonomously decide to give up autonomy. However, he later noted that thinking that autonomy can be given up by an autonomous choice is also coherent so that he claims to be unable to resolve the dilemma. The Theory and Practice, 128-129.

175 Note that this is strictly speaking not a case of paternalism. The state does not place consent demands on the agent because it want to secure some good for the agent that it thinks is in the interest of the agent. Instead, it places restrictions on the choice of the agent only for the sake of her future (autonomous) decision-making capacity.

176 The proposal of his that I am about to discuss is actually intended as an analysis of manipulation, but it carries over smoothly to the present discussion of authenticity.


178 In order to explain these effects, Noggle refers to evidence from learning theory for the existence of two networks, one operating on the basis of stimulus-response linkages and one operating on the basis of an interconnected semantic network. Ibidem, 63.

179 Ibidem, 64. The advantage of this criterion, Noggle suggests, is that the argument for qualifying a case of conditioned motivation as alien does not rely on it occurring against the victim’s will (this latter criterion does not work, he argues correctly, because we need a principle that accounts for the formation of an authentic will in the first place, and the will – as the outcome of that process – cannot be invoked to account for its own authentic formation).

180 Noggle might propose to specify a level of concordance in response to this particular objection, but I have no idea on the basis of which criteria he could plausibly do so.

181 Noggle ends his article by stating that “alien desires are repugnant because they involve information that does not go through the person’s rational faculties”, 67. This is confusing because he earlier defined alien desires on the basis of their concordance with beliefs (see quote above), as well as implausible as a theory of autonomy, because so much of our mental processing bypasses ‘the person’s rational faculties’.
182 For an introduction into this phenomenon, see Marco Iacoboni, Mirroring People. The New Science of How We Connect With People (New York 2008).


184 For accounts of why autonomy does not trump all other values see Oshana, Personal Autonomy in Society, chapter 6; Berofsky, Liberation From Self, chapter 10.


CHAPTER 3

1 I am grateful to Agnes Nairn, whose comments forced me to clarify this passage.

2 See Dan Sperber, Explaining Culture. A Naturalistic Approach (Oxford 1996). I take it that there is no controversy about which theories of culture count as more naturalistic in the sense of being closer to biology. Picking Sperber’s theory from among more naturalistic approaches may be less obvious. However, once it is accepted that memesics is seriously flawed because of its assumption that ideas would replicate – which it generally is – then there is no rival theory in terms of scope to Sperber’s account. And because my argument requires a theory of that scope, Sperber’s theory is the only viable account. Let me note also that social theories towards the humanities side of the spectrum tend to exclude insights from psychology which makes them unhelpful for my purpose in any case.

3 Ibidem, 60.


5 Ibidem, 134. See chapter 2.3 for an account of these principles.

6 In his discussion of the spread of myths Sperber notes that the incentive to retell a story can be either institutional or an intrinsic incentive. As an example of the latter kind he mentions ‘attractiveness’ which is related to the structure and themes of the story, but he does not explain what these themes are or how they relate to motivations. Nowhere in Explaining Culture, in fact, do we learn how motivations work in the first place. Ibidem, 95.

7 For Sperber’s discussion on feedback loops see Ibidem, 84.


13 Srinivasan et al., “An Approach to the Measurement”, 1444 and 1447


15 See also Les Binet and Peter Field, “The Conflict Between Accountability and Effectiveness”, in: Admap (June 2007), 22-24.


18 Ibidem, 150.
26 Ibidem, 29.
27 Carruthers, 24. There are limitations to such sharing though. It will not occur if two modules need the same information at the same time or if they are too far removed from one another in the brain, because processing would slow down in these cases.
29 Larry Bernard, Michael Mills, Leland Swenson and Patricia Walsh, “An Evolutionary Theory of Human Motivation”, in: *Genetic, Social and General Psychology Monographs*, vol 131 no 2 (2005), 129-184, 161. The rubric of memetic challenges involves the motivations to explain things and to leave a legacy. The argument appears to be that memes lead to social contexts beyond real groups to future groups to which people can be motivated to leave a legacy. It not clear in what the challenge exists however (even less so for the supposed motivation to explain things – it may be that the existence of memes is a necessary condition for question as to the meaning of thing arises, but that does not explain why this poses a challenge).
33 Kipling Williams in “Ostracism”, in: *Annual Review of Psychology*, vol 48 (2007), 425-452 distinguishes, next to the need to affiliate, another response to ostracism that consists of desires to acquire existential needs by engaging in antisocial behavior. The precise ostracism cues that activate the affiliation motivation module against for example the aggression module will depend on many situational factors and these have not yet been investigated in sufficient detail.
34 Mark Leary and Ellen Tambor, “Self-Esteem As An Interpersonal Monitor”, 524.
35 Williams, “Ostracism”, 338-441.
41 Bad genes would of course quickly be rooted out of any one population if it were not for processes such as mutation (of which the negative effect is too weak for it to be selected against immediately) and host-pathogen co-evolution.
43 Ibidem, 169-171. Female preferences for men are also related to their muscle mass and masculine features of the face (strong jaw, strong brows) which result from testosterone levels in adolescence. At this point males face a trade-off between allocating energy to mating effort (for example by muscle mass growth) and somatic functioning (such as that of the immune system). Ibidem, 176. Because there is thus a real cost to testosterone induced male features, they honestly signal the health of the organism to women.
44 Ibidem, 137.
45 Ibidem, 113 and 125-127. Interestingly, if android fat is deposited in for example breasts, it is not rated as attractive by men. Ibidem, 125.
46 Ibidem, 137-141.
50 Good hunters with high prestige have more surviving offspring. Ibidem, 66. Women also pay more attention to pictures of men with high status features. That is, when pictures of men, of which some display features of high status, are displayed briefly, women overestimate the number of high status men if they have to estimate their frequency under high time pressure. Because there was no such effect with regard to pictures of high status women, and because women displayed the same bias in a follow-up eye tracking study, this study suggests that women have a subtle bias for paying attention to high status men. Nathan DeWall and Jon Maner, “High Status Men (But Not Women) Capture the Eye of the Beholder”, in: *Evolutionary Psychology*, vol 6 no 2 (2008), 325-341. David Buller has contested the existence of a female preference for high status men in *Adapting Minds*, chapter 5. However, his alternative explanation of studies which show that women prefer high status men – that status homogamy predicts mate preference – does not stand up. That is, female university students indicate a preference for high income men, but this preference does not take the shape of an inverted U, as Buller’s homogamy hypothesis would predict. Andrew Delton, Theresa Robertson and Douglas Kenrick, “The Mating Game Isn’t Over. A Reply to Buller’s Critique of the Evolutionary Psychology of Mating”, in: *Evolutionary Psychology*, vol 4 (2006), 262-273, there 269. The authors also discuss studies which show that female preference for dominant men is moderated by the danger of the higher aggression being turned on women themselves, so that more dominance is not unconditionally preferred by women.


52 Thornhill and Gangestad, *The Evolutionary Biology*, 209, 211, 218, 220. Fertile female preferences for symmetrical male faces (when rating for short-term mating) are – as opposed to consistent preferences for men with more symmetrical bodily features, such as wrists – mixed. This may be because facial symmetry is a poor proxy for overall symmetry. Ibidem, 217.

53 Moreover, women may shift their preferences for a long-term mate in relation to their perceived mate value. It is found that women who are rated as more physically attractive have higher standards with regard to good genes, parenting, investment and partner indicators. David Buss and Todd Shackelford, “Attractive Women Want it All: Good Genes, Economic Investment, Parenting Proclivities, and Emotional Commitment”, in: *Evolutionary Psychology*, vol 6 no 1 (2008), 134-146.


56 Ibidem, 254.

57 Ibidem, 255.


61 What is more, with regard to behavior that is triggered by the mating motivation Geoffrey Miller has proposed that art, sports, creativity, morality, personality traits, intelligence and language evolved because of sexual selection pressures. Miller, *The Mating Mind. How Sexual Choice Shaped the Evolution of Human Nature* (New York 2000). Evidence for his sweeping hypotheses is still in short supply however.


64 Mazur, *The Biosociology of Dominance*, 66.

65 Ibidem, 67. There is also evidence that dominance is wired up similarly with innate human emotional systems for pride, shame, embarrassment, contempt and awe, which is another reason for thinking that the dominance motivation is innate. D. Keltner, J. Haidt and M. Shiota, “Social functionalism and the evolution of emotions”, in: M. Schaller, J. A. Simpson, and D. T. Kenrick ed., *Evolution and Social Psychology* (New York 2006), 115-142.


67 Ibidem, 118 ff.

Ryan Johnson, Joshua Burk and Lee Kirkpatrick, “Dominance and Prestige as Differential Predictors of Aggression and Testosterone Levels in Men”, in: Evolution and Human Behavior, vol 28 (2007), 345-451. The relation between prestige and aggression was more ambivalent than predicted, but that may be due to the fact that high prestige persons are provoked more often due their high status. Experiments that induce aggressive behavior in the laboratory could settle this matter.

There is suggestive evidence that status-seeking behavior is caused by testosterone from double blind studies (using a placebo), but due to small sample size and a small percentage of participants accounting for the entire effect, more studies are needed. See Mazur, The Biosociology of Dominance, pp 117 ff and 128-129.


This is a point that the authors do acknowledge. But if it is true, then one would need stronger arguments for allocating positions in a ‘pyramid of needs’ than saying that a certain motivation is really important.

Dan Ariely and Michael Norton, “Conceptual Consumption”, in: Annual Review of Psychology, vol 60 (2009), 475-490. What I call ‘unmediated consumption’ is called ‘physical consumption’ by the authors, but this is bound to lead to confusion.

Ibidem, 479.

See Mercier and Sperber, “Why do Humans Reason?”. Furthermore, a special category of public consumption consists of acts that are caused by the norm module. For the motives that are produced by this module are about adhering to abstract principles, not about impressing specific individuals. I exclude this special category of public consumption because of a lack of evidence about their influence (they re-enter only in section 5 and are discussed only insofar as they interact with the three social modules).

Angelo Chao and Juliet Schor, “Empirical Test of Status Consumption: Evidence From Women’s Cosmetics”, in: Journal of Economic Psychology, vol 19 (1988), 107-131, there 117. Results for mascara were more ambiguous, but the authors note that the quality test for this product differed from the rest.

See Gad Saad, The Evolutionary Bases of Consumption (Mahwah and London 2007) for an overview of evolutionary approaches to explaining consumer behavior. Let me note at this point also that several authors speak casually – especially in the context of consumption – of a presumed need for people to express themselves. See for example Jonah Berger and Chip Heath, “Where Consumers Diverge From Others: Identity Signaling and Product Domains”, in: Journal of Consumer Research, vol 34 (2007), 121-134. What is meant by this is that decisions are made on the basis of sufficient access to memory (as in PSI) but that individuals would have a need to stand out from the rest and show who they really are. I am skeptical about the sophistication of such a need to be different. It is uncontested that people often want to stand out from the rest, but such behavior seems to reduce either to simple heuristics or to the mating or status motivation. As to the heuristics, it turns out that people who order beer publicly seek greater variety (compared to table friends) than a control group in which subjects communicate their choice privately. Dan Ariely and Jonathan Levav, “Sequential Choice in Group Settings: Taking the Road Less Traveled and Less Enjoyed”, in: Journal of Consumer Research, vol 27 (2000), 279-290. There is also evidence that consumers in narrow store aisles display more varied and unique product choices. Jonathan Levav and Rui Zhu, “Seeking Freedom Through Variety”, in: Journal of Consumer Research, vol 36 (2009), 600-610. The underlying motivation of such behavior seems to be in the beer case best explained by evidence that people who make more varied choices gain more status, while the variety seeking in confined spaces may be better explained by reactance (see chapter 2.2) with which it was correlated in that experimental setting. R. Ratner and B. Kahn, “The impact of private versus public consumption on variety-seeking behavior”, in: Journal of Consumer Research, vol 29 (2002), 246–57, and Levav and Zhu, “Seeking Freedom”, 605. Hence, the desire to be distinct may not be the result of a deep urge to show one's true nature, but rather a mechanism for gaining status and making one's behavior less predictable when one feels pressured.

Aric Rindfleisch, James Burroughs and Nancy Wong, “The Safety of Objects: Materialism, Existential Insecurity, and Brand Connection”, in: Journal of Consumer Research, vol 36 (2009), 1-16. The researchers in fact make distinction between communal brand connections and self-brand connections, but it is not clear whether these types of connections are distinct. Communal brand connections consists of responses to the following items:

I really identify with people who use this brand.
I feel like I almost belong to a club with other users of this brand.
This brand is used by people like me.
I feel a deep connection with others who use this brand.
A closer look at the items of the self-brand connection measure reveals that there is nothing about these questions that a person could not report as an element of her feeling part of a brand community:

This brand reflects who I am.
I can identify with this brand.
I feel a personal connection to this brand.
I (can) use this brand to communicate who I am to other people.
I consider this brand to be “me.”


81 Ibidem, 7.
85 Interestingly, in a final study it was established that the act of eating a nostalgic cookie satiated the need to belong but not the choice of a nostalgic product. For this studies participants engaged in another game of cyberball which was interrupted three times by an implicit recognition test that measured their need to belong. First to ensure that there was an effect due to the social exclusion of the cyberball game, then after a (nostalgic) cookie choice test and finally after a (nostalgic) cookie eating test. Only the cookie eating test eliminated the implicitly measured need to belong.
87 But in a follow-up task they, interestingly, were more likely to choose voluntary work over leisure time when primed with images of sexually attractive men than a control group, and they preferred conspicuous volunteering options over non-conspicuous ones. On this task men showed no such effect.
89 Yet another study found that men switch to a much higher discount rate when they have seen pictures of beautiful women against a control group. That is, when primed with attractive women, men are less likely to defer consumption for the benefit of a higher pay out in the future. Looking at cars has no such effect. Women showed no change in discount rate after looking at pictures of beautiful men. Miller, Spent, 106-111.
93 Moreover, this effect could not be explained by the expected happiness that subjects ascribed to the future use of the product in an accompanying survey.
99 This is what Emily Stone, Todd Shackelford and David Buss find in “Sex Ratio and Mate Preferences: A Cross-Cultural Investigation”, in: European Journal of Social Psychology, vol 37 (2007), 288-296. The set of countries involved is small however, and the significance of the effect (i.e. the female adjustment of mate preferences) is marginal. Note that more demanding female long-term mating preferences in low sex-ration contexts could also be an evolved strategy of up-trading in ancestral contexts of polygyny.
102 Ibidem, 530.
104 See Miller, Spent, 85-86 for a similar argument.
106 Miller, Spent, 85 and 105.
107 Ibidem, 86.
108 Miller does by the way mention conditioning as a means of making effective sex appeals in ads on page 98 of Ibidem.

CHAPTER 4

1 There is an interesting but small literature on the effect of cognitive load on persuasion attempts. Anick Bosmans and Luk Walrop in How Vulnerable are Consumers to Blatant Persuasion Attempts? Research Report (University of Leuven 2005) demonstrate that playing background music seriously impairs consumer skepticism toward persuasion attempts by salespersons. This impairment occurs even if sales intentions are explicit because of the salesperson wearing a promotional t-shirt and cap and talking about a “promotional drink” (20). These results are explained with reference to the impairment of working memory by the background music, which inhibits the application of persuasion knowledge to the source of the persuasion attempt.
4 See Dennett, Elbow Room, 10-11, for a more detailed description.
7 Ibidem, 764.
8 Ibidem, 764.
14 Ibidem, 192.
15 Ibidem, 195.
16 As Chartrand and colleagues note, in-store displays are a growing trend. Ibidem, 198.
17 Ibidem, 197.
20 Ibidem, 407.
21 Ibidem, 409.
22 Ibidem, 411.
24 Furthermore, on implicit attitude measures there was no correction effect in either condition. But this suggests only that implicit brand attitudes are more resistant to propositional negations.
27 Ibidem, 183.
28 See chapter 2.3.
30 Ibidem, 347.
34 Strick et al., “Humor in Advertisements”, 40 and 42.
36 Strack, Werth and Deutsch, “Reflective and Impulsive Determinants of Consumer Behavior”, 206.
37 Friese, Hofmann and Wänke, “The Impulsive Consumer”, 355. Note that Agnes Nairn and Cordelia Fine suggest that autonomy in the realm of consumption requires that people be able to “overcome implicit consumer attitudes in choice situations” in “Who’s Messing With My Mind? The Implications of Dual-Process Models for the Ethics of Advertising to Children”, in: International Journal of Advertising, vol 27 no 3 (2008), 447-470, there 459. If Friese and colleagues are correct in thinking that shopping environments chronically facilitate impulsive decision-making, then this requirement may be too demanding.
38 Similarly, if advertisements for unattractive products are correlated with more sales if they induce negative affect, then – all other things being equal – we have suggestive evidence for thinking that introjections occurred.
39 Note that testing whether autonomy violating instances of negative affect occurred in advertisements is relatively straightforward. In company archives one can find historical records of advertisements, sales figures and even investigations into how much consumers like a brand compared to rival brands in its market. Given that introjection is observed for the 50% of subjects who score below the median on the action orientation test, introjection effects should show up in the sales record.
43 Ibidem, 16.
45 Ibidem, 518. This may have been because of metacognitions or because of subsequent explicit attitudes that suppressed such an evaluation.
48 Ibidem, 598.
49 That is, if we assume that learning processes are at work in advertising so that advertisements become more effective over time. A third explanation of the data would be that the increase in the use of tropes is due to motivations of the advertising professionals. It could be that the use of metaphors serves as a signal of clever-
ness and creativity in the advertising industry and so functioned as a devise to enhance prestige. This is not a very convincing argument however, because schemes seem to be no less reliable as a signal of creativity.


53 On the first point see Ibidem, 94.

54 Note the difference between autonomy violation in purchasing actions, discussed in section 3.1, and autonomy violation of attitudes that in turn cause purchasing actions.


57 Ibidem, 970.

58 Ibidem, 86.


60 Ibidem, 90.


62 If my assumptions are correct, future studies should also find evidence of learning processes that show that awareness of the workings of advertisements increases at both levels over time.


64 Ibidem, 3.

65 VN, item 1556. Letter by Hesselmans to Oppenheim, September 16, 1970.


68 Ibidem, 19.

69 AM, item 582. Memorandum by Kranenberg, August 30, 1963.

70 AM, item 13. Meeting of the board of directors, October 2, 1963.


72 Ibidem, 6.


CHAPTER 5

1 See Taylor, Brainwashing, for a preliminary analysis of organizational practices that might also structurally violate autonomy. She notes that education is not inherently hostile to autonomy because it aims to provide us with critical capacities for exercising autonomy, such as knowledge and the ability to ‘stop and think’, Ibidem, 64 and 66; similarly, mental health professionals aim primarily to help people to understand themselves, Ibidem, 77-78; and the military requires discipline, but soldiers are nowadays often given considerable responsibility, Ibidem, 82-83. Possible autonomy violations in such orchestrated emotional group formation are moreover legitimate if the army can be said to defend the national and international order that is necessary for liberal freedom to persist. Much more is to be said about these cases of course.


3 In fact, the researchers correlated these two measures of religiosity to three types of brand reliance, all which negative correlated with it: the number of Apple stores per million, the number of other brand stores per million and the brand-discount store ratio.

4 Shachar and colleagues went on to manipulate the type of religious prime and found that the negative correlation with national brand reliance (for signaling products) only occurred when participants were asked to think about religion in terms of self-worth and not when they were to think about religion in terms of security or about the religiosity of others. They take this to be evidence that the shared motivational source of religion and brands is the need to express self-worth. This is highly debatable though. First of all, it is not clear what expressing self-worth refers to. The term self-worth is open to all kinds of interpretations and may thus hide
a wide variety of motivations that may have been activated in subjects. Secondly, insofar as Shachar and colleagues invoke a need to express the self, I am skeptical for the reasons given in chapter 5, where I criticized that notion.

5 The development of so-called mega churches is more similar to branding in this regard. It will be discussed in section 1.2.


8 A far-fetched argument can be developed for thinking that the structure of religious beliefs exerts pressure in the direction of autonomy violation though. That is, one could argue that concepts that confute properties from multiple modules (e.g. the folk psychology and folk biology modules) and so achieve a memory optimum (e.g. a speaking snake) cannot be validated either by induction or by deduction. As a result, thinking hard about such concepts is not going to strengthen faith. Hence, the existence of supernatural agents must be validated at an emotional level, and this is done most effectively during synchronized and more or less ritualized activity. In other words, on this argument the nature of religious beliefs drives ritual practice in the direction of autonomy violation by means of synchronic states induced by music, chant and dance. My judgment is that there are too many speculative assumptions involved in this argument for it to be plausible. But nothing in my overall argument in this chapter hangs on this objection being correct.


10 Ibidem, 717.


15 It has shown for example that involuntary movements in the face are harder to control in response to music, Sosis and Alcorta, “Ritual, Emotion, and Sacred Symbols”, 337.

16 Ibidem, 333.


19 Ibidem, 540.

20 Ibidem, 539. See also Atran, *In Gods We Trust*, 13.

21 See Raymond Paloutzian, James Richardson and Lewis Rambo, “Religious Conversion and Personality Change”, in: *Journal of Personality*, vol 67 no 6 (1999), 1047-1079. On the basis of the very imperfect available evidence the authors conclude that conversion has a minimal impact on Big Five personality traits, which are very stable over the life course, but that it has a profound effect on goals, feelings, attitudes and behaviors, while it also profoundly affects midlevel narrative personality structures that guide orderings of life goals.

22 See Bulbulia, “Meme Infection or Religious Niche Construction? An Adaptationist Alternative to The Cultural Maladaptationist Hypothesis”, in: *Method and Theory in the Study of Religion*, vol 20 (2008), 1-42, there 36, where it is argued that because of costly testing of commitment “the risks of manipulation are significantly reduced”.

23 Moreover, this structural drive for intention validation by emotional signaling offers a structural antidote to religious Machiavellians only insofar as these are regularly involved in rituals. And the larger the organizations of a religion, the more need for specialization and the less ‘ritual contact hours’. James Hunter, “Religious Elites in Advanced Industrial Society”, in: *Comparative Studies in Society and History*, vol 29, no 2 (1987), 360-374, there 371. For this reason, large scale doctrinal religions leave more room for manipulative intentions than smaller scale emotionally arousing types of religion.

24 Recently, Pascal Boyer and Pierre Liénard have argued that there may be a Hazard Precaution System in the brain that is receptive to the repetitive, scripted and goal demoted behavior (but not always) seen in rituals. In a nutshell, this mechanism makes the organism carry out precaution procedures that may prevent future harm (e.g. washing hands prevents infection and counting stones makes one attentive to possible visits by predators that passed by). Repetitive behavior may thus appeal to humans by activating this fear-relieving system, so that over the course of cultural evolution rituals with rigid repetitive procedures had better survival chances. Boyer and Liénard, “Why Ritualized Behavior? Precaution Systems and Action Parsing in Devel-


31 Stark and Finke, Acts of Faith, 114. Stark and Lofland noticed that all new members of the cult they observed joined after establishing personal contacts. See also Chana Ullman, The Transformed Self. The Psychology of Religious Conversion (New York 1989) for evidence from interviews on the importance of personal contacts for religious conversion.

32 Note that proper consent is absent in the case of children who convert to a religion. However, parents by no means influence their offspring by religious thoughts exclusively, as atheist parents likewise pass on a wide variety of beliefs about the good life to their offspring. For this reason the question we are dealing with is one of parental autonomy in general. In this regard, only a radical liberal version of parental autonomy would clash with allowing for the conversion of children, since parents on this conception may not bias their children towards any idea of the good life, see Colin MacLeod, “Conceptions of Parental Autonomy”, in: Politics and Society, vol 25 no 1 (1997), 117-140, there 128. This position is highly problematic, since, as MacLeod points out, it undermines the social conditions for children's autonomy that are secured by parenthood in the first place. That is, parents have the right to decide on most matters for their children (as long as some minimal basic rights are respected) because they can best provide the intimacy required for a healthy mental development of a child, have a privileged epistemological access for establishing the needs of the child and are most motivated to care for the child. Ibidem, 129. All three conditions are undermined when parents are no longer able to guide their parenting behavior by the goals in life they deem of the greatest importance: it hampers their motivation, it hampers the emotional basis for intimacy and, as result of the weakening of the bond with the child, the ability to access the child's needs. Hence, the unreconstructed version of liberal parental autonomy is inconsistent (it erodes the psychological basis for the autonomy it wants to preserve) and does as such not undermine the case for allowing infant conversion. For a more extensive discussion of these issues, see Bader, Secularisms or Democracy?, 211 and 268.

33 See R. Laurence Moore, Selling God. American Religion in the Marketplace of Culture (Oxford 1994) for an account of this development.

34 The exact number is regularly updated by the Hartford Institute for Religious Research at their website, it was accessed by me at February 15, 2010 http://hirr.hartsem.edu/cgi-bin/mega/db.pl?db=default&uid=default&view_records=1&ID=&sb=3&so=descend

35 http://hirr.hartsem.edu/mega_church/megastoday2008_summaryreport.html


37 Ibidem, 61.


39 http://www.businessweek.com/magazine/content/05_21/b3934001_mz0001.htm

40 What is more, religious practice is moving out of churches and into the household, where religion is experienced in interaction with music records, movies, tv preachers, and merchandise. Ibidem, 60. As a result, brands of faith become part of everyday life, just as commercial brands, and are thus not subject to the gradual conversion process that characterizes traditional religions.

41 The marker of the year award is a prize that is awarded every year at the annual gathering of the Association of National Advertisers. See Matthew Creamer, “Obama Wins! … Ad Age’s Marketer of the Year”, in: Ad Age, October 17 2008. http://adage.com/moy2008/article/article_id=131810

CONCLUSION

1 Robert Levine, The Power of Persuasion. How We’re Bought and Sold (Hoboken 2003), 3. It can alternatively be argued that the reason for the lack of concern for autonomy violation is the fact that the mass media in which such debates would have to occur earn most of their money from advertising, and that this exerts structural pressure against criticizing advertising. There may be something to this argument, but at least in Western Europe there are public broadcasting venues some of which do not figure any advertising at all – such as the BBC – but still do not extensively discuss the nature of the influence of advertising. It is possible that different institutional or cultural biases are responsible for the lack of attention for the subject on advertising-free public broadcasting channels, but I know of no evidence which indicates that this is the case.

2 Ibidem, 21 ff.

3 It may be objected that work in psychology has long since demonstrated illusions of control and that these have not changed public opinion about agency. But this objection is unconvincing because earlier psychological experiments have not been part of major research programs such as those pursued in neuroscience. Moreover, there is experimental evidence to suggest that non-neuroscientists are easier to persuade when neuroscientific insights are used even if these are not relevant. See Deena Weisberg, Frank Keil, Joshua Goodstein, Elizabeth Rawson, and Jeremy Gray, “The Seductive Allure of Neuroscience Explanations”, in: Journal of Cognitive Neuroscience, vol 20 no 3 (2008), 470-477.

4 Emily Falk, Elliot Berkman, Traci Mann, Brittany Harrison and Matthew Lieberman, “Predicting Persuasion-Induced Behavior Change from the Brain”, in: The Journal of Neuroscience, vol 30 no 25 (2010), 8421-8424.
References


Anderson, Joel and Axel Honneth, “Autonomy, Vulnerability, Recognition and Justice”, in: Joel Anderson and John Christman, Autonomy and the Challenges to Liberalism (Cambridge 2005), 127-149


Atran, Scott, In Gods We Trust. The Evolutionary Landscape of Religion (Oxford 2002)


Bader, Veit, Secularism or Democracy? Associational Governance of Religious Diversity (Amsterdam 2007)

- and Albert Benshop, Ongelijk-heden. Sociale ongelijkheid en sociaal handelen (Groningen 1988)


Bronmley, David and Richardson ed., The Brainwashing/Deprogramming Controversy: Sociological, Psychological, Legal and Historical Perspectives (New York 1985)


Buller, David, Adapting Minds. Evolutionary Psychology and the Persistent Quest For Human Nature (Cambridge 2005)


- , “How We Know Our Own Minds: The Relationship Between Mindreading and Metacognition”, in: Behavioral and Brain Sciences, vol 32 (2009), 121-138


Chartrand, Tanya, Joel Huber, Baba Shiv and Robin Tanner, “Nonconscious Goals and Consumer Choice”, in: Journal of Consumer Research, vol 35 (2008), 189-201


Colburn, Ben, Autonomy and Liberalism (New York 2010)

Dennett, Daniel, Elbow Room. The Varieties of Free Will Worth Wanting (Cambridge 1984)
- , Breaking the Spell. Religion as a Natural Phenomenon (New York 2006)
Einstein, Mara, Brands of Faith. Marketing Religion in a Commercial Age (London 2008),
Falk, Emily, Elliot Berkman, Traci Mann, Brittany Harrison and Matthew Lieberman, “Predicting Persuasion-Induced Behavior Change from the Brain”, in: The Journal of Neuroscience, vol 30 no 25 (2010), 8421-8424
Frank, Robert, Choosing the Right Pond. Human Behavior and the Quest For Status (Oxford 1985)


Lippke, Richard, Radical Business Ethics (Lanham 1995)


Mazur, Allan, The Biosociology of Dominance and Deference (Oxford 2005)


Mele, Alfred, Autonomous Agents. From Self-control to Autonomy (Oxford 1995)


Oshana, Marina, Personal Autonomy in Society (Aldershot 2006)

Park, Ji and Deborah John, “Got to Get You into My Life: Do Brand Personalities Rub Off on Consumers?”, in: Journal of Consumer Research, vol 37 (2010), pre-published online on July 6, 2010


Rambo, Lewis, Understanding Religious Conversion (Yale 1993)


Roccato, Michele and Christina Zogmaister, “Predicting the Vote Through Implicit and Explicit Attitudes: A Field Research”, in: Political Psychology, vol 31 no 2 (2010), 249-274

Rössler, Beate, Der Wert des Privaten (Frankfurt am Main 2001)


Saad, Gad, The Evolutionary Bases of Consumption (Mahwah and London 2007)


Sperber, Dan, Explaining Culture. A Naturalistic Approach (Oxford 1996)


Thornhill, Randy and Steven Gangestad, The Evolutionary Biology of Human Female Sexuality (Oxford 2008)


Weinberger, Joel and Drew Westen, “RATS, We Should Have Used Clinton: Subliminal Priming in Political Campaigns”, in: Political Psychology, vol 29 no 5 (2008), 631-651


Westen, Drew, The Political Brain. The Role of Emotion in Deciding the Fate of the Nation (New York 2007)


ARCHIVES

AM Archief van de Amstel Brouwerij N.V., (1734) 1970-1969, toegangsnummer 1566
(Stadsarchief Amsterdam)
VN Archief De Weduwe J. Van Nelle, (1837) 1837-1989, toegangsnummer 944
(Gemeentearchief Rotterdam)