Autonomy and Metacognition : A Healthcare Perspective

Levinsson, Henrik

2008

Autonomy and Metacognition

A Healthcare Perspective

Henrik Levinsson

LUND UNIVERSITY

Department of Philosophy

Autonomy and Metacognition: A Healthcare Perspective
Henrik Levinsson

Printed by Media-Tryck AB, Lund, Sweden

ISBN 978-91-628-7614-2

For my family

Acknowledgements

Since this project is multidisciplinary, several people from different academic fields have been involved in the development of the dissertation. I would like to begin with thanking three professional researchers who have been very influential in my academic development. First, I would like to express my gratitude to my supervisor Ingar Brinck, for all her help and good advice of how to think philosophically and how to become a skilled philosopher. Her experience of cooperating professionally with other scientific disciplines has also been very valuable in the process. Ingar and I have shared many meaningful and thoughtful discussions and I am very thankful for all her support and patience. She has really encouraged me to do hard but constructive work. Further, I would also like to thank my co-supervisor Margareta Östman who invited me to the psychiatric field. Margareta also invited me to join her at the 30th International Congress on Law and Mental Health in Italy 2007, which was very rewarding. Her clinical experience and constructive comments have been important motivators in my thinking and writing. Margaretas positive encouragement is a true virtue. I would also like to thank Professor Ingalill Rahm Hallberg, for all her careful thoughts, and her interest in how philosophers can contribute to medical research. Ingalill has in several respects inspired my work and development. Her academic strength is exemplary.

In the final stages of the project, Professor Erik J Olsson provided stringent comments on several parts of the dissertation. Thank you! I would also like to thank the talented philosophers who have participated in the Higher Seminar at the Department of Philosophy: Staffan Angere, Sebastian Enqvist, Bengt Hansson, Tobias Hansson, Jan Hartman, Martin Jönsson, Anna-Sofia Maurin, Johannes Persson, Stefan Schubert, Robin Stenwall, Niklas Vareman and Lena Wahlberg. Furthermore, I would also like to thank Professor Nils-Eric Sahlin and his "mini-seminar" where some parts of my theoretical reflections were presented. Thank you, Mats Johansson in Practical philosophy, Petra Björne in Cognitive Science, and Rikard Liljenfors in Psychology, for valuable comments and support.

During my four years and more as a doctoral student, I have frequently participated in workshops arranged by The Vårdal Institute (Vårdalinstitutet). These workshops have been very fruitful. I would like to thank all the inspired researchers in the Psychiatry platform for your clinical expertise and positive approach to multidisciplinary research. It was exciting to discuss psychiatric as well as philosophical topics with you. Through the Vårdal Institute I have received valuable comments from the "Anthology Group". This group made the anthology *Etiska utmaningar: i hälso- och sjukvården* possible. Thanks to all the people involved and special thanks for the healthcare discussions we had.

I am very grateful for the financial aid from the following foundations: Fredrik och Ingrid Thurings stiftelse, Stiftelsen Erik och Gurli Hultengrens fond för filosofi, Stiftelsen Fil dr. Uno Otterstedts fond för främjande av vetenskaplig forskning och undervisning, Stiftelsen Makarna Ingenjör Lars Henrik Fornanders fond, and The Vårdal Institute.

Many thanks to all my friends for their support and understanding through challenging periods. Some persons I wish to mention here: Maria Andersson, Kristina Blom, Christina Henriksson, Helena Lundberg, Marie Norgren, Anna Perbring, Therese Persson, and Anna Sjöland. Additional thanks to Annika Hedlund and Tandläkargruppen who took care of my teeth during the most stressful periods.

Finally, my family, to whom I dedicate this dissertation, has been of tremendous help and support: Bertil, Christina and Christoffer Levinsson. I have no words to describe my thoughts of gratitude to you. Nonetheless, thank you very much!

CONTENTS

PART I The Concept of Autonomy

PART I

The Concept of Autonomy

Chapter 1

Autonomy: A Metacognitive Capacity

1.0. Introduction

Autonomy is normally seen as something valuable – as something worth having, something that gives meaning to our lives. In several respects, autonomy is important (Hill 1991, p. 43). The possession of autonomy is also associated with a right: the right to be treated in certain ways. The principle of autonomy, which concerns this right, is probably one of the most highly valued principles connected with human rights. In this sense, to be or remain autonomous, or to have one's autonomy respected or protected, is in many ways important. However, what does being autonomous involve? The thesis will deal with this question.

The concept of autonomy is central in many fields. Consider, for instance, philosophy, politics, medicine, biomedical ethics, and education policy. This thesis will deal with the concept as it is understood in philosophy and medicine, concentrating especially on Swedish healthcare and psychiatry.

The contention that the concept of autonomy is one of the most central concepts in medicine is not controversial. In a fundamental sense, healthcare policies and general directions are based upon the idea of the patient's right to autonomy (Tännsjö 2008, p. 104). In the Swedish Health and Medical Service Act one of the basic, distinguishing requirements of proper healthcare is that such care should be "…built upon respect for the patient's right to self-determination" (SFS 1982:763, my translation).[1] This is reiterated in the Swedish Social Service Act (SFS 2001:453). It is important to note that the right to autonomy – in medicine, at least – is commonly equated with the right to self-determination, i.e. the right to decide for oneself. In healthcare the role of respect for autonomy has, perhaps, become even more important than it once was as a result of the role of the patient's reinforced position (Nordgren 2003). The right to autonomy is an often-elucidated goal when questions about healthcare and medical options are considered.

An important distinction, and one that is seldom made explicit in healthcare, is that between the concept of autonomy as a cognitive capacity and the concept of autonomy as a

---

[1] In Swedish: Ett av kraven enligt Hälso- och sjukvårdslagen är att den skall "…bygga på respekt för patientens självbestämmande" (SFS 1982:763). All subsequent excerpts from the Act given in this chapter were translated by the present author.

right. This makes the concept of autonomy problematic. Like the Health and Medical Service Act (SFS 1982:763) above, the principle of autonomy concerns the individual's right to make her own decisions (Beauchamp & Childress 2001, p. 12). The Act claims that "…everybody is to respect others' capacity for, and right to, self-determination (autonomy), participation and integrity" (MFR 2002, p. 17).[2] Concerning patients' right to have their autonomy respected, the principle further claims "…that they have the capacity to decide independently regarding information about alternatives" (MFR 2002, p 17).[3]

The concept of autonomy is often understood from an ethical perspective, as when it is interpreted as a right. However, and as will be argued in this thesis, it cannot be understood as a right. According to Anderson and Lux (2004a, p. 312), one should explicitly respect the distinction between normative notions of autonomy and the individual's capacity to be autonomous. As I see it, this distinction is important: it is vital to avoid misunderstandings caused by failure to separate autonomy as a moral concept and autonomy as a cognitive capacity.

In connection with both the Health and Medical Service Act (SFS 1982:763) and the principle of autonomy, there are exceptions to the right to autonomy. Consider, for instance, coercive care in psychiatry, where the right to autonomy can be hard to respect because the mental disorder reduces the patient's ability to make decisions and take care of herself. As one may note, the concept of autonomy is especially problematic and intractable in psychiatry. The thesis will focus on this problem.

It is a major problem in healthcare and psychiatry that the concept of autonomy is understood both as a right and as a cognitive, or mental, capacity of the individual.[4] I will argue that the understanding of autonomy as a right has over-shadowed its interpretation as a cognitive capacity. The focus here on patient rights may become problematic if the role of cognitive capacity is neglected. In evaluating and developing healthcare efforts, one must be explicit about this distinction. Otherwise, implementations of various healthcare efforts run the risk of demanding autonomy from patients who cannot, in point of fact, be ascribed

---

[2] In Swedish: Autonomiprincipen säger att "…var och en skall respektera andras förmåga och rätt till självbestämmande (autonomi), medbestämmande och integritet" (MFR-rapport 2 2002, s. 17).

[3] In Swedish: Autonomiprincipen innebär att "…de har förmåga att självständigt ta ställning till information till handlingsalternativ" (MFR-rapport 2 2002, s. 17).

[4] In psychiatry "mental capacity" is a term often used in connection with cognitive functioning of the individual. See, for instance, Berghmans, Dickenson and Meulen (2004) and Breden and Vollmann (2004).

autonomy. In view of this, the concept of autonomy, understood as a cognitive capacity, has to be elucidated and dealt with in more detail.

To summarize, the distinction between autonomy as a right and autonomy as a cognitive capacity has not been sufficiently investigated and elaborated. Indeed these contrasting concepts appear sometimes to be used interchangeably. Problematically, the two understandings run the risk of becoming mixed up.

The concept of autonomy is problematic in psychiatry (and other fields) because it is an ambiguous concept and can be interpreted in several ways. The reader might well become confused over the diversity of meanings when reading the literature.

From now on I will analyse the concept of autonomy in terms of a cognitive capacity. Thus, the discussion of autonomy as a *right* will be put aside. However, I will briefly present some other interpretations of the concept of autonomy, since these have been influential in recent decades.

1.1. Diversity of meanings

Several philosophers admit that there is little agreement, or at any rate no consensus, about the meaning of autonomy, and that the concept of autonomy is ambiguous (Hill 1991; Beauchamp & Childress 2001; Dworkin 1988; Taylor 2005). It is commonly held that autonomy can mean many things (Dworkin 1988). Different understandings of the concept of autonomy are frequently discussed in the literature. It is uncontroversial to say that the concept of autonomy is used in ways that are both ambiguous and broad.

Historically the concept of autonomy referred to self-rule by city-states (Dworkin 1988; Beauchamp & Childress 2001).[5] A city was regarded as autonomous if it ruled itself in the absence of external forces. A city-state was claimed to have autonomy if it was independent (Dworkin 1988, pp. 12-13). Over recent decades the ideas of self-rule and independence have been extended to apply to *individual* autonomy as well as liberty rights, freedom, integrity and privacy (The National Board of Health and Welfare 1991, Sweden). However, there are divergent views about how, exactly, these ideas apply to the individual (Dworkin 1988; Beauchamp & Childress 2001).

Several interpretations of the concept of individual autonomy can be found in the literature: self-rule, self-government and independence; but also self-determination, freedom

[5] In Greek, *auto* refers to the self, and *nomos* to law (see Dworkin 1988).

of the will, rationality, self-directedness, self-control, sovereignty, to be a law to oneself and self-trust (Frankfurt 1971; Lindley 1986; Christman 1989; Dworkin 1988; Lehrer 1999b; Beauchamp & Childress 2001; Buss 2004). Although overlapping and similar understandings of the concept of autonomy are often discussed, some remain more plausible than others. At the same time, I would acknowledge that no single, precise concept of autonomy exists today. In Part I, some common interpretations of the concept of autonomy are discussed.

In neither philosophy nor medicine is it sufficiently clear exactly how the concept of autonomy is to be understood. Given that great value attaches to respect for autonomy, one might reasonably expect there to be an elaborated theory of autonomy. Unfortunately, we lack such a theory. For instance, one may wonder what exactly is being defended or respected when it is claimed that autonomy must be respected, protected or retained in medicine. Another problem is whether autonomy always has to be respected, and whether autonomy today is respected to a larger degree than it should be. As Christman (2003) notes, there is scepticism and controversy over the question whether autonomy is an unqualified value that relates to all individuals.

Philosophers have tried to categorize the various views of the concept on offer. One category that is frequently discussed concerns, more or less directly, the understanding of autonomy as a cognitive capacity.

According to Beauchamp (2005, p. 310) three common claims can be distinguished in the literature on autonomy. First, there is the metaphysical status of the autonomous person.[6] The second claim concerns the understanding of autonomy based on a theory of mind, self and the person, while the third claim directs attention to the connection between moral status and the concept of a person. In this thesis the second claim will be the focus. The reason for this is that it is this claim that is in agreement with the understanding of the concept of autonomy as a cognitive capacity. Below I explain why.

Moral and metaphysical analyses are put to one side in the thesis, but this does not mean that they are unimportant. According to Beauchamp (2005, ibid) the second claim need not

[6] Consider, for instance, the Kantian notion of autonomy that refers to some kind of metaphysical freedom unobtainable in the empirical world (see Oshana 2006, p. 5).

necessarily concern moral notions. However, as I see it, it might have moral implications both in theory and in practice.[7] It is important to bear this in mind.

Hermerén (2006, p. 178) makes a distinction that is similar to Beauchamp's. According to Hermerén the concept of autonomy can be placed in three categories: a psychological one, a normative one, and one that concerns values. Consequently, Hermerén introduces and considers several interpretations of the concept autonomy. As the reader may suspect, it is the first interpretation that we shall deal with. The focus will be on the cognitive capacity of the individual.[8]

According to Feinberg, autonomy has four related meanings. It refers to a capacity for self-government, to the actual condition of self-government, to an ideal, and "…to the *sovereign authority* to govern oneself" (Feinberg 1989, p. 28). As one might guess, it is the first interpretation that will mainly be dealt with in this thesis.

As was noted, we seem to lack a satisfactory theory of autonomy as a cognitive capacity. At the same time, it is common in the literature to presuppose that some groups in society lack autonomy. For instance, infants and somewhat older children, patients who suffer from severe mental disorders and the comatose are claimed to be non-autonomous individuals because they lack the cognitive capacity needed in order to make decisions and lead their lives in a controlled way. However, in what sense do they lack the cognitive capacity required? As I see it, this question is not conclusively investigated. We lack a theory that makes explicit what kind of cognitive capacity autonomy is.

The present analysis of the concept of autonomy concerns the autonomous *individual*, not autonomous *acts*. Analyses of both concepts – that of the autonomous individual and that of autonomous action – are valuable and important. For instance, it might be fruitful to analyse autonomous action with regard to certain tests that require specific kinds of skill in particular situations, or situations where certain decisions have to be made. Analyses that emphasize autonomous action in particular situations normally understand the concept of autonomy from a local perspective. By contrast, analyses that focus on the autonomous individual over extended periods of time understand the concept of autonomy from a global perspective. I

---

[7] Moral and normative notions of autonomy concern, for instance, the question whether autonomy is an intrinsic or instrumental value; the question whether autonomy is a (human) right; and the question whose autonomy deserves respect.

[8] I will understand "cognition" from a broad perspective taking into account perception, emotion and attention. Concerning the broad perspective on cognition, see Brinck (2007).

previously claimed that the concept of autonomy is ambiguous. As I see it, in order to reduce this ambiguity, it is important to be clear about whether the analysis of autonomy one is concerned with applies to individuals or actions.

The global perspective will be dealt with in this dissertation. I shall focus on the individual because the analysis is meant to facilitate the implementation of certain healthcare guidelines and prescriptions that concern the autonomy of the individual – e.g. her right to govern herself in the everyday life.

1.2. Undermined Autonomy in Terms of Cognitive Impairment

Given that our aim is to elucidate autonomy as a cognitive capacity, it might be instructive to consider some examples of undermined autonomy. These examples are intended to illustrate the notion that whether or not an individual is autonomous depends on her cognitive capacity.

A. Cognitive impairment caused by a stroke

> Suppose an individual has spent time in a hospital following a stroke. She is now discharged from the hospital, but because of her stroke several of her cognitive functions are impaired. She is confused, her memory capacity is impaired, and she suffers from chronic fatigue. In addition, her ability to plan and to localize in time and space has weakened; so also her sense of identity. Nevertheless, her physical condition, her bodily functioning, is intact. Without difficulty, she can walk and move around using her body as earlier; but when she leaves the hospital, she cannot find the way home.

B. Anton's Syndrome

> A severe frontal injury contused the anterior portions of John's brain and at the same time shattered both of his orbits, severing his optic nerves and leaving him with no light perception at all. The resulting behavioral syndrome was quite striking in that John not only insisted verbally that he still had vision, but he also initiated behavior as if he did, trying to move about his room in the manner of a person with normal vision. As a result, he walked into walls and furniture, collided with objects in his path rather than avoiding them, and repeatedly placed himself in positions that were extremely precarious for a person who could not see. Despite his ability to initiate action in an apparently self-directed way, John's persistently mistaken assessment of his visual capacity with respect to his actions made it impossible for him to act as he intended. In this sense, many of his actions could not count as autonomous, not because he could not see – plenty of blind individuals are perfectly autonomous – but rather because his impaired self-assessment left him unable to make sense of what he was doing. At least with respect to those actions, he was deeply alienated from himself as an agent. (Anderson & Lux, 2004b, p. 280)

C. Thought insertion in schizophrenia where the mental activity is experienced as alien by the individual herself

> The content of the experience seems to be exactly that token thoughts are being generated by some other person, and, perhaps with malice, inserted into the mind of the patient, so that the patient has direct introspective knowledge of a token thought which was generated by someone else. (John Campbell, 2002, p. 36)

The three examples elucidate undermined autonomy of the kind resulting from cognitive impairment. Case A is intended to illustrate the fact that impaired cognitive functioning caused by a stroke deprives the affected individual of autonomy and in turn makes it troublesome to govern oneself in daily life. The stroke might impair cognitive functions such as learning, understanding and remembering. Case B involves lack of insight into one's condition because of neurological deficit, which in turn undermines autonomy. Case C is meant to elucidate the unusual experiences reported by patients suffering from thought insertion. In cases of thought insertion the patient experiences her own cognitive processes as alien. All of these (somewhat different) cases of impaired cognition illustrate deprived autonomy understood from a global perspective: that is, the patients are in several respects unable to make decisions about themselves. The similarities obtaining between the cases motivate an analysis of deprived cognitive capacity for autonomy.

According to Edwards, mental health and the capacity for autonomy are interrelated. For the first "…includes only those desirable mental/behavioral normalities and occasional abnormalities which enable us to know and deal in a rational and autonomous way with ourselves and our social and physical environment" (Edwards 1997, p. 53). Edwards further states that autonomy is a capacity that is actualized in order to enable the "…making one's own choices, managing one's own practical affairs and assuming responsibility for one's own life, its station and its duties" (Edwards 1997, ibid).

That autonomy is a cognitive capacity is a fairly uncontroversial claim, or so it will be argued here. Consider, also, other cases, including the case in which an individual who (in the nature of the case) has never been autonomous is born with severe brain damage. In such cases, autonomy is undermined. There are individuals who cannot be ascribed autonomy. If an individual's autonomy is to be respected, she has to be autonomous in a certain sense (or, at least, must *have been so*, as will be discussed in 7.3). This is an essential line of thought throughout the thesis. The major task is to elucidate the sense in which it is true.

1.3. Autonomy in Psychiatry

In Swedish psychiatry, and indeed in the field of medicine as a whole, a common understanding of the concept of autonomy invokes the notion of self-rule (or self-determination). Self-rule is normally understood as the ability to make decisions (Dworkin 1988, p. 14). However, it is sometimes claimed that a decision must be made in the absence of external and internal influences if it is to count as autonomous (Beauchamp & Childress 2001).

In this thesis it will be argued that the concept of autonomy, when it is understood in terms of self-rule, delivers at best an incomplete understanding of what is involved in autonomy as a cognitive capacity. The ability to make decisions is one kind of cognitive skill that is essential for autonomy, or so it will be argued. Nevertheless, the concept of autonomy as self-rule tends to put too much emphasis on decision-making.

As noted above, the concept of autonomy plays an important role in Swedish psychiatry. On the one hand, it is to be respected, protected, or retained. For instance, the healthcare efforts that emphasize participation and integration in society for individuals suffering from mental disorders today are a result of the closure of the old mental hospitals. However, these efforts seem to require the relevant individuals to be capable of taking care of themselves. As Markström (2003, p. 155) writes, some of the goals of the efforts in psychiatric care developed by the Psychiatric Investigation in the 1990s were built on the view that the patient has the same rights and duties as other groups in society. Further, these healthcare efforts were to be based on the patient's own choices and priorities.

The aim of these goals was to improve the life situation of the patient. Yet the goals also require that patient to be autonomous. More precisely, the healthcare efforts focus on the individual, and on her ability to take care of herself. However, this can be claimed to be problematic in psychiatry. Several mental disorders involve cognitive impairments that reduce the sufferer's ability to live "a normal life" outside a mental institution. In a document written by The National Board of Welfare (Sweden), one finds the following:

> Cognition is a summary designation concerning our capacity to receive and interpret impressions in order to handle them rightly, leading to purposeful actions. Examples of such fundamental functions are attention, endurance, memory, language and aspects of intelligence. The cognitive functions are therefore fundamental in order to operate in daily life. Impaired cognitive capacity is very common in schizophrenia. Several patients have such

> serious impairments that the capacity to manage the fundamental demands of daily life is compromised. (The National Board of Health and Welfare 2003b, my translation)[9]

It is an important question to what extent autonomy can be respected and protected in cases where an individual suffers from a severe mental disorder, and where consequently cognitive capacity is reduced. It is, for instance, an important question whether there are non-autonomous individuals who are treated as if they were autonomous, or (perhaps more accurately) as if they are expected to be so. In turn, this may – paradoxical as this may sound, and against the goals of healthcare – negatively influence the individual's life as a whole.

It seems, in a general sense, contradictory to respect, or protect, an individual's autonomy if she lacks autonomy. Coercion, as it is used in psychiatry, more or less explicitly supports this claim. However, it is, in an important sense, questionable whether there is an established relationship between the concept of autonomy (interpreted as a capacity) and the criteria for coercive care in psychiatry. Let me present some of the criteria governing the imposition of coercive care in Sweden.

According to The Compulsory Mental Care Act coercive care is legitimate only if the patient suffers from a severe mental disorder. Moreover, coercion is legitimate when there are reasons to suppose that care cannot be provided through the patient's consent (SFS 1991:1128). Given this I would deny that there is a clear relationship between the concept of autonomy as a cognitive capacity of the individual and the criteria of coercive care. However, it remains important to consider this relationship, since coercive care is partly based on, and justified by, consideration of the individual's autonomy. I hope that this thesis provides fruitful suggestions as to the nature of the elusive relationship; and I will suggest that the relationship must be discussed more explicitly when coercive care is considered. That is, the criteria governing the imposition of coercive care ought to be tied to an understanding of undermined autonomy.

---

[9] In Swedish: Kognition är en sammanfattande beteckning på vår förmåga att ta in och tolka intryck för att sedan behandla dem på rätt sätt så att de leder till ändamålsenliga handlingar. Exempel på sådana grundläggande funktioner är uppmärksamhet, uthållighet, minne, språk och aspekter på intelligens. De kognitiva funktionerna är därför grundläggande för att det dagliga livet skall fungera. Försämrad kognitiv funktionsförmåga är mycket vanligt vid schizofreni. Flera patienter har så allvarliga försämringar att det försvårar förmågan att klara de grundläggande krav som ställs i det dagliga livet.

1.4. Methodological considerations

The present work differs in some respects from more conventional dissertations in analytic philosophy in that it is partly applied. Therefore, I want briefly to discuss some methodological issues. Part I of the dissertation puts forward a conceptual analysis of autonomy, as that concept is understood in philosophy and medicine. In an effort to evaluate the concept I will use real-life cases of undermined autonomy. In that way I can test its limitations and possibilities. The phenomenon of undermined autonomy will be illustrated by, and substantiated with, empirical data on impaired cognition as well as functional cognition. In Part II of the thesis I will apply the analysis of autonomy put forward in Part I in the context of Swedish healthcare and psychiatry.

The conceptual analysis is inspired by Carnap (1950). I particularly emphasize the relation between explicandum (that which is to be explicated) and explicatum (that which explicates). According to Carnap, the relation between explicandum and explicatum has to be similar to each other. However, explicandum is pre-theoretical and informal. With respect to the present thesis, the diversity of understandings discussed above points to a somewhat informal account of the concept of autonomy. The explicatum, on the other hand, should be precise and well constructed. Two more conditions put forward by Carnap are emphasized in the analysis: fruitfulness and simplicity.

When he speaks of fruitfulness, Carnap is claiming that a conceptual analysis has to be useful. Methodologically, the analysis developed here meets the condition of fruitfulness. Part II of the thesis will apply the philosophical theory put forward in Part I to Swedish healthcare and psychiatry. This application of a philosophical analysis of the concept of autonomy will, I hope, both improve our understanding and offer a better way to deal with questions about autonomy in practice.

However, the analysis might also be of interest in other fields. As I see it, a fruitful theory of the concept of autonomy should be applicable in fields other than psychiatry. Consider, for instance, law, education, and fields in medicine such as paediatrics and geriatrics. Further, as the project lying behind this thesis is multidisciplinary, an analysis of the concept of autonomy might strengthen the communication between various professions confronted with issues arising in connection with autonomy.

The simplicity condition states that a theory of autonomy should be (plausibly) consistent with other concepts associated with the concept of autonomy. This is also pointed out by Dworkin:

> The concept should be neither internally consistent nor inconsistent (logically) with other concepts we know be consistent. So, for example, if the idea of an uncaused cause were inconsistent and autonomy required the existence of such a cause, it would fail to satisfy this criterion. (Dworkin 1988, p. 7)

I freely admit that it is not easy, in a conceptual analysis, to establish the necessary and sufficient conditions of autonomy. That there are necessary conditions for autonomy might be evident, but whether these conditions are together *sufficient* is more difficult to establish. The difficulty partly depends on the abundance of contextual factors that influence the exercise of autonomy, both in a substantial and an undermining sense.[10] I will give examples that show why. Nevertheless, the analysis I provide is intended to establish what is essential for autonomy.

In addition to trying to respect Carnap's criteria, we shall presuppose that a satisfactory explication of autonomy will be empirically applicable – and hence that there are in fact autonomous individuals in the sense suggested. The analysis should issue in an understanding that is neither too narrow nor too broad: whether an individual is autonomous or not cannot rely on too restricted or too allowing conditions. As Beauchamp and Childress observe, "No theory of autonomy is acceptable if it presents an ideal beyond the reach of normal choosers" (Beauchamp & Childress 2001, p. 59).

It was previously mentioned that empirical data on cognitive functioning and impaired cognition will be considered in order to illustrate what it means to be autonomous. However, it is difficult to integrate empirical data in a philosophical analysis. For instance, philosophical method differs crucially from the methods of the empirical sciences. Moreover, concepts are differently understood in different fields, and are operationalized for certain purposes. Therefore, general aspects of a concept might be neglected. However, I think it is valuable to try to integrate relevant disciplines with an interest in questions about autonomy as a cognitive capacity. For instance, philosophers can suggest "…types of inquiry"; and philosophical "…reflections are valuable as guides to the kinds of empirical inquiry it would be useful to pursue" (Dworkin 1988, p. 162). Such guidelines may, for instance, relate to the investigation of certain social as well as psychological factors that seem to make autonomy (and the exercise of it) possible. However, they might also relate to investigations of undermined autonomy.

---

[10] That there are difficulties in trying to establish necessary and sufficient conditions for autonomy is also stated by Dworkin (1988).

As I see it, philosophical investigation of the concept of autonomy should not be isolated from the empirical sciences. If it does, it will not be fruitful for the theory. Murphy (2005) claims that we might increase our understanding of the concept of autonomy by taking into account clinical experience, and that we have reason to integrate empirical data in philosophical discussions of autonomy. He writes:

> We think that autonomous individuals critically evaluate their life and actions, endorse them as self-determined rather than imposed from without, and guide their own lives in accordance with the plans and values they have worked out for themselves. Such self-government has to depend on some psychological structures, but philosophers have tended to theorize about what these structures might be at a considerable distance from what the behavioral sciences can tell us. (Murphy 2005, p. 303)

If I understand Murphy right, philosophical discussions of the concept of autonomy often tend to ignore empirical research, and therefore there is a need to clarify what we mean when say that autonomy is a cognitive capacity. Agich further states that bioethics "…misses the more challenging and potentially fruitful collaboration that neurosurgery and neurology affords for advancing the philosophical understanding of the conditions of autonomy" (Agich 2004, p. 295). In order to clarify what kind of cognitive capacity autonomy is, I believe, the dialogue between philosophy and the empirical sciences should be more vigorous.

Next, I first sketch the general content and structure of the thesis. After that, the analysis of the concept of autonomy I favour is put forward.

1.5. Overview of the thesis

The aim of Part I is to examine the cognitive aspects of autonomy. The central question concerns what kind of cognitive capacity autonomy is. It will be argued that the concept of autonomy is best understood in terms of a metacognitive capacity of the individual. That is, the analysis put forward is an analysis of the *metacognitive* components of autonomy.

In Chapter 2 and Chapter 3, theories of autonomy that interpret the concept of autonomy metacognitively are discussed and evaluated. The major theory dealt with is Dworkin's – a variety of so-called "hierarchical theory". Dworkin puts forward the idea that autonomy is a second-order capacity of the individual. This capacity, it is argued, is a metacognitive capacity. Three common objections to hierarchical theories as such are analysed and evaluated. Thereafter a coherence theory of autonomy, which may be seen as a development

within hierarchical theory, is analysed and evaluated. Importantly, it solves some of the three leading problems that strike hierarchical theories.

What both theories have in common is that the metacognitive capacity of the individual is essential for autonomy. However, both theories tend to over-emphasize the role of second-order capacity as a conscious process and neglect the role of lower level reflexivity. Chapter 4 therefore offers a detailed analysis of what metacognition is and asks in what sense it involves lower level reflexivity. It is argued that metacognition has two components: procedural reflexivity and metarepresentation. Metarepresentation in turn can be divided into inferential reflexivity and other-attributiveness. These two components are essential for autonomy. Particular emphasis is put on procedural reflexivity. It is concluded that autonomy, understood in a global perspective, is both a procedural and a metarepresentational capacity. Further, since the essential function of metacognition is control, it is argued that the concept of autonomy, understood as a metacognitive capacity, can be interpreted in terms of control.

Chapter 4 and 5 discuss various aspects of control. Issues arising from empirical data from neuroscience on functional, as well as impaired, metacognition, and on undermined autonomy, are dealt with. It is argued that autonomy cannot be determined with respect to subjective conditions. Neurological impairments, like Anton's Syndrome, dementia, and thought insertion in schizophrenia, are put forward in support of this claim. To determine autonomy we require external conditions. It is suggested that intersubjectivity constitutes such a condition.

External circumstances, like manipulation, that prevent an individual from exercising her autonomy irrespective of her metacognitive intactness are discussed. Importantly, there is a difference between being autonomous and exercising autonomy. It is argued that autonomy is to some extent relational, and hence that being autonomous cannot be the same thing as being independent in a strict sense.

Independence, strictly speaking, would exclude, for instance, loyalty and advice from other people, as well as emotional mechanisms which, it is argued, play an important role in metacognition. However, it might be hard to separate external circumstances that violate the exercise of autonomy from those that do not. Consider, for instance, persuasion: does persuasion undermine the exercise of autonomy? While persuasion can be understood in different ways, being capable of resisting persuasion, and yet being receptive to good advice, are central to autonomy – or so it will be argued.

The conclusion drawn at the end of Part I is that if respect for autonomy is a goal of Swedish healthcare, we need to focus on the metacognitive capacity of patients. Since respect

for autonomy is, in many respects, a goal in Swedish healthcare, an enhanced focus on the metacognitive capacity of individual patients is required.

On the other hand, it is argued that a fruitful theory of the concept of autonomy must reflect two dimensions: the metacognitive/internal and the relational/external dimension. In the autonomy debate, attention tends normally to be drawn to one dimension, while the other is downplayed or neglected. But to avoid a narrow understanding of the concept of autonomy, a two-dimensional theory is called for – one that takes into account what autonomy is as well as what might hinder an individual in exercising it. By considering the relational/external dimension of autonomy in relation to the individual's metacognitive capacity, we obtain (it is claimed) a plausible and fruitful theory of autonomy that can be applied to the various areas of interest. This theory takes into account both the metacognitive capacity of the individual and her environment.

It is argued that the metacognitive components of autonomy might make it possible to determine autonomy. However, in order to determine whether an individual is autonomous, both the metacognitive status of the individual and the external setting must be considered, since they are in interplay and consequently influence each other.

In Part II, as was mentioned above, the analysis put forward in Part I is applied to Swedish healthcare and psychiatry. In chapter 6, the focus is on the principle of autonomy, the patient's reinforced position in Swedish healthcare, and the Swedish deinstitutionalization of mental hospitals. To enable those involved to live a "normal" life like members of other groups, the goals of deinstitutionalization included the patient's integration and participation in wider society. However, in chapter 7 I question what will happen to individuals belonging to vulnerable groups in the realization of these goals. Actual and potential problems, such as morbidity, isolation and passivity, among patients who suffer from persistent mental disorder are discussed.

In Chapter 7, it is also argued that the individual's right to autonomy depends on his or her metacognitive capacity. Unfortunately, this idea seems to have been neglected, and indeed there is sometimes (what might be called) a "blind" defence of the autonomy principle in Swedish healthcare. However, the defence is inadequately supported, since it is not clear what autonomy involves or requires.

In the last section of chapter 7, it is suggested that the metacognitive account of the concept of autonomy might help clarify the criteria governing coercive care. Finally, in chapter 8 some suggestions concerning developments and improvements in Swedish healthcare, especially in psychiatry, where the concept of autonomy is important but problematic, are put forward.

Chapter 2

Autonomy, Second-order Capacity and Metacognition

2.0. Thinking Thoughts About One's Thoughts

Theories of individual, or personal, autonomy often put forward the idea that to be autonomous one requires a certain capacity (Haworth 1986; Lindley 1986; Dworkin 1988, Waller 1998; Lehrer 1999b; Anderson & Lux 2004b; Taylor 2005; Oshana 2006). In Chapter 2 and Chapter 3 an analysis of the concept of autonomy as a metacognitive capacity of the individual is put forward. It will be concluded the concept of autonomy can reasonably be interpreted as a metacognitive capacity.

Metacognition is normally regarded as the capacity to reflect on one's own thoughts (Metcalfe & Kober 2005; Proust 2007). To clarify, metacognition makes it possible to think thoughts about one's thoughts (Metcalfe & Kober 1994). Alternatively, put in another but similar way, "Metacognition is defined to be cognition about cognition" (Smith, Shields & Washburn 2003, p. 318). From a global perspective, metacognitive skills are useful in everyday life. Consider, for instance, our ability to plan for future goals, to predict our capacity with respect to some cognitive goal, to retrodict (judge after the facts), or to evaluate and control our own mental states (Nelson 1996; Metcalfe & Shimamura 2000; Proust 2007).

The concept of autonomy, understood in terms of a metacognitive capacity of the individual, requires detailed discussion. The working definition of metacognition put forward in the previous paragraph is tentative. A more precise understanding of metacognition will be developed stepwise throughout the thesis. The end of Part 1 will summarize these steps in a general theory of autonomy as a metacognitive capacity of the individual. This is the step taken in the thesis.

In the present literature on the concept of autonomy some theories deal with autonomy as a metacognitive capacity, since they claim that autonomy is the capacity to have higher-order thoughts (see Meyers 1987; Dworkin 1988; Ekstrom 1993; Christman 2003; Buss 2004; Mele 2005; Taylor 2005; Oshana 2006). The concept of metacognition is not used explicitly by most of these theories, but they all concern the individual's capacity to reflect on her lower-level mental states, like desires or beliefs, on a higher level of thought. Among the most elaborate and discussed such theories are Dworkin's hierarchical theory and Ekstrom's coherence theory (Dworkin 1988; Ekstrom 1993). These will be examined below.

In understanding the concept of autonomy as the capacity to have higher-order thoughts the two theories suggest something important. However, they do not draw any further conclusions from this. They do not take the further step of understanding this capacity in terms of metacognition, nor do they examine in what sense, exactly, autonomy might be metacognitive. The metacognitive account of the concept of autonomy is fruitful, because it clarifies the kind of capacity required for autonomy.

This chapter will explain in what sense autonomy is a metacognitive capacity with reference to Dworkin's hierarchical theory. Chapter 3 is devoted to Ekstrom's coherentist analysis of autonomy, which is a development of, but also an alternative to, Dworkin's theory.

2.1. Dworkin's Hierarchical theory

Among hierarchical (sometimes described as structural) theories of autonomy, Dworkin's variant is one of the most developed and fully elaborated. His theory is described as a hierarchical theory of autonomy because it is influenced by the higher-order thought theory. Before going into a detailed presentation and analysis of Dworkin's theory, let me briefly explain the essential features of the higher-order thought theory.

According to the higher-order thought theory, humans are able to observe their first-order mental states on a higher level of thought – a level at which the mental state of the higher level represents the first-order mental state. According to Dienes and Perner (1999) this requires

> …explicit representation of the content of the first-order mental state. For example, to consciously know that the banana is yellow, I must explicitly represent that it is a present fact that the banana is yellow, that this fact is known, and I must be able to explicitly represent that it is I who know it. (Dienes & Perner 1999, p. 741)

In higher-order thought theory, the understanding put forward in the above quotation is common (Dienes & Perner 1999). In several respects, it has influenced hierarchical theories of autonomy, and today it constitutes a cornerstone of the autonomy debate.[11]

There are several variants of hierarchical theory. Why, then, is Dworkin's theory the one to be examined in this thesis? The answer is that Dworkin's theory, unlike other hierarchical

[11] See Frankfurt (1971) and Dworkin (1988).

theories, explicitly concerns the autonomy of the *individual*, not autonomous acts or desires.[12] He writes: "I am not trying to analyze the notion of autonomous acts, but what it means to be an autonomous person, to have a certain capacity and exercise it" (Dworkin 1988, pp. 19-20). In this sense, his theory is in line with the present enquiry. Of course, hierarchical theories, including Dworkin's, are in some other respects similar and share basic similarities. I will now present the essential features of Dworkin's theory.

Dworkin holds that it is characteristic of autonomy to involve the capacity to adopt second-order attitudes to one's first-order motivations (Dworkin 1988, p. 15). This requires the individual to have the second-order capacity for higher-order thought.

A first-order motivation is, according to Dworkin, simply a desire, preference, wish or intention to do x. To have a second-order attitude, on the other hand, is to have the desire, preference, wish or intention to have the desire, preference, wish or intention to do x. Hence, a second-order attitude is a thought about a first-order motivation. For instance, to give a simple example, the desire to have the desire to have a cup of coffee.

Dworkin has developed and refined his theory of autonomy over the years. However, in early versions he argued that identification was a necessary condition of autonomy, i.e. that an individual, in order to be autonomous, must be able to identify with a first-order motivation on a second level of thought (Dworkin 1970, 1976).

The identification requirement was, as I understand it, the essential characterization of the second-order capacity. Consider, for instance, a second-order attitude such as "I desire to have the desire to leave the room". In such a situation, the individual, on a second level, identifies with the first-order desire to leave the room. As long as an individual has the capacity to identify with the relevant first-order motivations – i.e. where the individual's "…second-order identifications [are] congruent with his first-order motivations" (Dworkin 1988, p. 15) – autonomy obtains. On the other hand, incongruence between the levels would be a case of undermined autonomy.

Lack of the capacity to identify with first-order motivations will make the individual non-autonomous. Consider the heroin addict who, at the second level, does not identify with the first-order desire for the drug, but acts on that desire. Her first-order desire makes it irresistible to take the drug, and she acts on that desire, while her second-order desire is the desire to refrain from the desire to take the drug. In such a case, autonomy is undermined.

---

[12] It might sometimes be difficult to determine, in the theories, whether autonomy concerns a feature of individuals, of actions or of desires (see Oshana 2006).

A short parenthesis here: hierarchical theories often claim that the capacity to identify with a first-order motivation on a second-order level is a basic feature of autonomy or personhood. This can be seen in Frankfurt's published work from 1971. Although Frankfurt did not develop an account of autonomy, but rather of personhood and freedom of the will (at the same time as Dworkin himself began to develop his own theory of autonomy), his work had a strong influence on hierarchical theories of autonomy (Christman 2003; Oshana 2006). Frankfurt's theory has been especially influential with respect to his understanding of the concept of identification. However, Christman claims that the theory Frankfurt puts forward is not an account of autonomy. Frankfurt's theory rather concerns an analysis of the freedom of the will. Yet, as was previously claimed, Frankfurt's theory has influenced the analyses of the concept of autonomy (Christman 2003). However, Frankfurt used the concept of autonomy explicitly in his later works. For instance, in one textual note one reads as follows: "Autonomy is essentially a matter of whether we are active rather than passive in our motives and choices− whether, however, we acquire, they are the motives and choices that we really want and are therefore in no way alien to us" (Frankfurt 2004, p. 20). It is true, though that Frankfurt's initial theory, and the one that has influenced the autonomy debate, did not deal with the concept of autonomy explicitly. Therefore, to extrapolate from Frankfurt's theory and to derive conclusions about the concept of autonomy might run the risk of confusing the debate. Therefore I will only discuss theories the major aim of which is to elucidate the concept of autonomy.

Back to Dworkin and the concept of identification. Dworkin revised his early view, and the identification requirement was argued not to be sufficient for autonomy in later versions (Dworkin 1988, p. xi). The argument was that the second-order capacity involves more than mere identification, i.e. autonomy cannot solely be a matter of whether the individual is able to identify with a first-order motivation on a second level of thought. Dworkin writes: "It is not the identification or lack of identification that is crucial to being autonomous, but the capacity to raise the question of whether I will identify with or reject the reasons for which I now act" (Dworkin 1988, p. 15).

With regard to the above quotation, the second-order capacity is to be understood from a global perspective, because "…it is a feature that evaluates a whole way of living one's life and can only be assessed over extended portions of a person's life, whereas identification is something that may be pinpointed over short periods of time" (Dworkin 1988, p. 16). Dworkin further claims, for instance, that the lobotomized individual "…is not having his

identifications interfered with, but rather his capacity or ability either to make or reject such identifications" (Dworkin 1988, ibid).

Understood from a global perspective, the concept of autonomy is a second-order capacity an individual has over time, but which is exercised only if needed. Dworkin understands this capacity for autonomy as follows.

> …autonomy is conceived of as a second-order capacity of persons to reflect critically upon their first-order preferences, desires, wishes, and so forth and the capacity to accept or attempt to change these in light of higher-order preferences and values. By exercising such a capacity, persons define their nature, give meaning and coherence to their lives, and take responsibility for the kind of person they are. (Dworkin 1988, p. 20)

An individual must have the second-order capacity in order to be autonomous, but that does not mean that she constantly needs to exercise it. According to Dworkin this is, as I understand the matter, the major reason why autonomy, as a second-order capacity, has to be understood from a global perspective.

In addition, it is important to note that second-order reflection does not have to be a conscious, explicit and articulated process (Dworkin 1988, p. 17). Dworkin adds this to defend his view against intellectualist approaches to autonomy, which, on his view, seem to require too much in respect of the complexity of reflection. Dworkin does not develop this claim in detail, but we shall later see how the claim harmonizes with the understanding of the concept of autonomy as a metacognitive capacity.

As the reader can understand, identification alone is not enough to constitute the second-order capacity. Other skills are also required. In this sense, Dworkin's later version is less narrow than his earlier one. Moreover, the concept of identification, as I understand it, is downplayed in the later version. The second-order capacity comprises, except identification, also the rejection and revision of one's first-order motivations. First-order motivations would otherwise not be effective in action, according to Dworkin.

In order to make her first-order motivations effective in action, the individual's second-order attitudes must originate from herself. However, this requires in turn, according to Dworkin, procedural independence, i.e. the absence of influences that disturb second-order reflection. He writes: "Spelling out the conditions of procedural independence involves distinguishing those ways of influencing people's reflective and critical faculties which subvert them from those which promote and improve them" (Dworkin 1988, p. 18). Manipulation or hypnotic suggestion exemplify failure of procedural independence, while

advice from, say, relatives, need not. However, and as will be discussed in 2.3, the conditions for procedural independence might be difficult to spell out.

It is time to summarize the essential features of Dworkin's theory. In order to be autonomous, the individual must possess the second-order capacity to reflect upon her first-order motivations and develop second-order attitudes to them. This capacity is understood from a global perspective.

The next section is devoted to an evaluation of Dworkin's hierarchical theory. Thereafter, three common problems commonly raised against hierarchical theories are dealt with.

2.2. Evaluation of Dworkin's theory

The way in which Dworkin understands the concept of autonomy, as a second-order capacity of the individual, is in line with the tentative definition of metacognition presented in 2.0. Because metacognition is normally understood as the capacity to think thoughts about one's thoughts, it is reasonable to interpret the second-order capacity described by Dworkin as metacognitive. It should now be clear, I hope, why the second-order capacity could be interpreted as a metacognitive capacity.

There are good reasons to downplay the concept of identification as happens in Dworkin's later work. For instance, mere appeal to the idea that one identifies with one's first-order motivations is not sufficient in understanding what the second-order capacity is. Since more skills are involved, Dworkin later puts forward a broader view of the second-order capacity than he had done before. The more developed second-order capacity has, as I understand it, three components: identification, rejection and revision.

Autonomy is the second-order capacity to identify with, reject and revise one's first-order motivations. Through second-order reflection, the individual can develop certain attitudes to her first-order motivations. However, and as Dworkin argues, that does not necessarily mean that she identifies with them. Here I agree with Dworkin. For instance, the individual might reject or revise her first-order motivations for good reasons. Hence, it is not reasonable to suggest that the second-order capacity must be understood solely in terms of identification. Next I will suggest that the concept of identification has to be, not downplayed as Dworkin seems to argue, but jettisoned from the theory as such.

The concept of identification is unfortunate, because it is ambiguous (Christman 2003). It can be interpreted in more than one way. Consider, for instance, the distinction between

identifying with a first-order motivation and just identifying it. The latter can refer to the *detection* of a first-order desire. On the other hand, on the first interpretation, the individual is able to identify *with* a first-order motivation.

It is not enough to disambiguate the concept of identification. We also need to clarify what "identify with" really means. Dworkin's theory does not provide a properly elaborated account of the concept of identification. As I see it, these problems constitute a drawback in understanding the concept of autonomy as it is described by Dworkin. One suggestion is that the concept of identification should be removed from the theory. Since it is ambiguous, it runs the risk of confusing the autonomy debate. On my view, it is a concept that is easy to misinterpret. Moreover, Dworkin does not seem to be willing to specify what the concept of identification means.

If the concept of identification has to be abandoned, what is the alternative? Recall the quotation of Dworkin presented earlier: "Putting the various pieces together, autonomy is conceived of as a second-order capacity of persons to reflect critically upon their first-order preferences, desires, wishes, and so forth and the capacity to accept or attempt to change these in light of higher-order preferences and values" (Dworkin 1988, p. 20). As one can see, Dworkin does not explicitly use the concept of identification in his characterization of autonomy.[13] It is reasonable to believe that, in the later versions of his theory, Dworkin downplays the concept of identification. As we saw in the quotation above, Dworkin does not explicitly use the concept of identification but rather stresses the role of acceptance. Yet it is difficult to know whether "identify with" and the concept of acceptance are used synonymously in Dworkin's theory, or whether the concept of acceptance is regarded as elucidatory. However, it is reasonable to assume the latter.

Again, I suggest that the concept has to be removed from the debate as such, rather than downplayed. But even if it is probable that Dworkin wants to downplay, but retain, the concept of identification in his theory, in many other respects I agree with the basic characterization of the concept of autonomy as a second-order capacity put forward in his theory.

My suggestion is that if we are to eradicate the problematic concept of identification, that concept must be replaced by the concept of acceptance, since this concept is less ambiguous. However, in addition to acceptance, rejection and revision are regarded as two essential features of the second-order capacity. To clarify, then, the second-order capacity for

[13] This has also been pointed out by Christman (2003)

autonomy comprises *acceptance, rejection and revision* of one's first-order motivations. In my view, the individual, through her second-order capacity, is able to accept, to reject, and to revise her first-order motivations.[14]

To summarize, the concept of identification must be replaced by that of acceptance; but acceptance is only one component of the second-order capacity. Let me now briefly discuss some examples that illustrate acceptance, rejection and revision.

An autonomous individual is able to reject an irrational or unwanted desire: that is, she can mobilize her second-order capacity in order to refrain from acting on a first-order desire. An individual might have a first-order desire that she feels is alien to her, but be able, through her second-order capacity, to reject it.

Consider also a situation where the individual is required to decide between two or more competing first-order motivations. In such a situation, the individual has to make a selection, or choice. This is possible through second-order acceptance and rejection. For instance, the individual may check the compatibility of the first-order motivations she is confronted with in order to get them to fit with her overall values and preferences. The individual can accept or reject a certain first-order motivation depending on whether or not it is in line with her overall set of desires.[15] Finally, consider situations where the individual, through her second-order capacity, revises her attitude to her first-order motivations.

The distinction between rejection and revision has to be made clear here. What the examples of the first sort aim to illustrate is, for instance, that a desire that becomes rejected is not replaced by another desire. In cases of revision, by contrast, the abandoned desire is replaced by another one (e.g. the desire to do the opposite of what the first desire dictates). To clarify with some more examples: to decide not to have the desire to eat a hamburger is a typical instance of rejection, while to decide to desire not to eat a hamburger is a typical instance of revision.

Let me summarize the discussion so far. First, it has been argued that the concept of identification has to be abandoned. The concept of acceptance was suggested to replace it. In addition, Dworkin's global understanding of autonomy as a second-order capacity seems, at least, to downplay the concept of identification. I will deal with this claim in Chapter 3. I shall explain why the concept of acceptance is less ambiguous, and hence less problematic, than that of identification. Finally, the concept of autonomy as a metacognitive capacity is in line

---

[14] The concept of acceptance is also discussed in Frankfurt's papers on identification. E.g. see Frankfurt (1999).

[15] This issue will be dealt with in more detail in Chapter 3.

with Dworkin's theory insofar as it involves the capacity to develop second-order attitudes to one's first-order motivations (desires, preferences, wishes and intentions). Thus, the second-order capacity of the individual is a metacognitive capacity of the individual.

Nevertheless, and with respect to the concept of autonomy, if the second-order capacity is a metacognitive capacity, which it is reasonable to argue, then, plausibly, this capacity must involve more than reflection upon one's first-order motivations (desires, preferences, wishes and intentions). It is uncontroversial to claim that the second-order capacity of the individual concerns not only connative mental states, but also epistemic mental states, like beliefs.

Autonomy, understood in terms of metacognition, must take into account the ability to reason about one's beliefs and not just about one's first-order motivations. Consider the examples of undermined autonomy presented in the introduction of this thesis (stroke, Anton's syndrome and thought insertion in schizophrenia). They illustrate the fact that being able to form and to reason about one's beliefs is essential to autonomy. They further illustrate that reasoning about one's beliefs can go awry as a result of metacognitive impairment.

Like first-order motivations, first-order epistemic states are objects of second-order reflection. The reflective individual is able to develop second-order attitudes to her first-order beliefs – e.g. "I believe that I believe that p". Hence, a belief of the second order is a belief of a belief of the first order. The individual, through her second-order capacity, is able to accept, reject, or revise her first-order beliefs.

Dworkin understands the mental states on the first level in terms of motivation only. However, metacognition also comprises epistemic states of the individual, like beliefs. Dworkin's theory seems to neglect this, focusing solely on motivations, i.e. the connative mental states of the individual.

Reflecting upon one's beliefs, and forming second-order attitudes to them, is essential to autonomy understood in terms of metacognition. It is a weakness in Dworkin's theory that it does not take into account the role of epistemic mental states, like beliefs, in second-order reflection.

Consider the following claims, which seem to convey the true story of autonomy as it emerges from Dworkin's hierarchical theory:

(A) An individual x is autonomous if x is able to accept, reject or revise a first-order motivation through a second-order attitude

However, it was previously claimed that epistemic mental states, like beliefs, are relevant to the possession of autonomy. The following, then, seems correct:

(A´) An individual x is autonomous if x is able to accept, reject or revise a first-order epistemic state through a second-order attitude

In addition, since the second-order capacity to be autonomous concerns connative as well as epistemic mental states, a more general understanding is required:

(A´´) *An individual x is autonomous if x is able to accept, reject or revise a first-order mental state through a second-order attitude*

Since the concept of autonomy, treated as a metacognitive capacity, must include reflection upon one's first-order epistemic states, the idea of first-order motivation should be replaced with the general concept of a *first-order mental state* – a concept which covers both connative and epistemic states. To be able to accept, reject, or revise one's first-order connative and epistemic mental states is a necessary condition of autonomy. To clarify, in order to be autonomous one must be able to accept, reject and revise one's connative as well as one's epistemic mental states.

From now on I will use the term 'first-order mental state' (FOMS) whenever I refer to connative and epistemic states of the individual. Further, I will sometimes use 'second-order mental state' (SOMS) when I refer to an individual's attitude to a FOMS. The next section will consider three problems that are commonly raised against Dworkin's theory and hierarchical theories as such.

2.3. Objections against Dworkin's theory

Some critical comments on hierarchical accounts of autonomy have already been pointed to above. The critique mainly focused on Dworkin's theory and, in particular, the concept of identification, and the neglected role of epistemic mental states.

In this section, three other common objections to hierarchical theories will be dealt with: the infinite regress problem, the degree problem and the historical problem. I will give a general presentation of these problems and evaluate how threatening they really are to hierarchical theories. Moreover, I will explain in what sense they are relevant to Dworkin's

theory, and indicate how he has replied to some of them. Finally, possible solutions to the problems are suggested.

In Chapter 3 I will analyze the coherence theory of autonomy, which seems to avoid at least some of the problems that strike the hierarchical theory. The coherence theory suggests a more detailed and, in some respects, a more plausible account of the concept of autonomy as a metacognitive capacity. Let us now turn to the first objection.

(i) The infinite regress problem

It is sometimes argued that it is hard to determine at what level of reflection autonomy is secured (Taylor 2005). Dworkin's theory claims that an individual is autonomous if she has the capacity to reflect on her first-order motivations on a second level of thought. However, one may now ask what is so special about mental states on the second level, as opposed to, say, first- or third-order mental states. As Taylor asks: "…how is it that a person's higher-order desires possess any authority over her lower-order desires?" (Taylor 2005, p. 6).

It is uncontroversial to claim that humans do in fact sometimes reflect on a third level. Thus suppose I can have a desire on the third level that is incongruent with my second-order desire (Juth 2005, p. 136). Then, one might ask again: what is so special about the second level? The problem presented here is commonly referred to as the problem of authority. However, this problem appears to lead to another problem that we shall endeavour to deal with here, namely: the infinite regress problem.[16] Next, I will explain the infinite regress problem.

The two-level theory advocated by, for instance Dworkin, has been questioned because further levels might be required by autonomy (Watson 1975; Ekstrom 1993; 1999; Taylor 2005). Taylor writes that, according to hierarchical theories:

> …a person is autonomous with respect to her effective first-order desires if she endorses them with a second-order desire. Because this is so, the question arises as to whether this person is autonomous with respect to this second-order desire and, if she is, why this is so. If she is autonomous with respect to this second-order desire because it is, in turn, endorsed by a yet higher-order desire, then a regress threatens, for the question will then arise as to whether she is autonomous with respect to this *third*-order desire – and so on. (Taylor 2005, p. 6)

[16] With the problem of authority in mind, we will return to the question about the function of first-order mental states in Chapter 4.

Suppose an individual accepts a FOMS as she develops a SOMS toward it. Since the individual is able to do this, she is autonomous according to hierarchical theory. But must not the SOMS in turn be accepted through the development of a third-order mental state directed toward it? Moreover, must not this third-order mental state be accepted through a developed fourth-order mental state, and so on ad infinitum? This line of thought explains why the infinite regress problem arises.

The regress problem is one of the most frequently discussed arguments against hierarchical accounts (Taylor 2005). Because of it, critics claim that hierarchical theories of autonomy fail. However, is the critique successful, and must hierarchical theories of autonomy therefore be regarded as false? Below, I will discuss Dworkin's, as well as Frankfurt's, responses to infinite regress. In addition, solutions to the problems will be suggested. For instance, on the assumption that autonomy is a metacognitive capacity of the individual, the infinite regress will be mitigated – or so it will be argued.

Is it the case that there are an endless number of levels? If so, what would the consequences be for the hierarchical analysis of autonomy? Dworkin's first reply to the regress problem is as follows: "As a matter of contingent fact human beings either do not, or perhaps cannot, carry on such iteration at great length" (Dworkin 1988, p. 19). Frankfurt argues in a similar but, as I see it, more detailed way:

> Another complexity is that a person may have, especially if his second-order desires are in conflict, desires and volitions of a higher order than the second. There is no theoretical limit to the length of the series of desires of higher and higher orders; nothing except common sense and, perhaps, a saving fatigue prevents an individual from obsessively refusing to identify himself with any of his desires until he forms a desire of the next higher order. The tendency to generate such a series of acts of forming desires, which would be a case of humanization run wild, also leads toward the destruction of a person. (Frankfurt 1971, p. 16)

A problem with the idea of an endless number of levels is that human beings seem, as a rule, capable of reaching a terminus, or end-point, in reasoning. For instance, humans normally do not reason in the following sense: "I wish to have the desire to have the desire to have the desire to have the desire to have the desire to quit smoking". Of course, reasoning on a third level can happen in cases, such as: "I wish I had the desire to have the desire to go to the gym". However, if there were an endless number of levels, would it be plausible to assume that there is a correct level on which autonomy is secured – say, level 234? The search for a "true" level would be time-consuming and not in line with the way in which human reasoning

works. As is implicit in Frankfurt's and Dworkin's replies, it does not seem possible for humans to reach' let alone stop at, level 234 of thought.

Dworkin's second reply to the regress problem is that the regress arises only with respect to whether acts are autonomous or not. He writes that the objection

> …concerns the acts of critical reflection themselves. Either these acts are themselves autonomous (in which case we have to go to a higher-order reflection to determine this, and since this process can be repeated an infinite regress threatens) or they are not autonomous, in which case why a first-order motivation evaluated by a nonautonomous process *itself* autonomous. My response to this objection is that I am not trying to analyze the notion of autonomous *acts*, but of what it means to be an autonomous person, to have a certain capacity and exercise it. (Dworkin 1988, pp. 19-20)

If the individual has the prerequisites, she is autonomous. Dworkin states: "There is no conceptual necessity for raising the question of whether values, preferences of the second-order kind would themselves be valued or preferred at a higher level, although in particular cases the agent might engage in such higher-order reflection" (Dworkin 1988, p. 20).

What are we to make of Dworkin's second reply? Since humans are normally equipped with a coherent cognitive system that is stable enough to enable rational thought, and at the same time is designed to produce action as an output of the reasoning process, they are also able to reach an end-point in reasoning. Therefore, the regress does not arise and this prevents the individual from being stuck in an endless chain of reasoning. To give an example, reasoning ends when the individual accepts or rejects the FOMS and regards it as compatible (or incompatible) with her larger set of, say, desires and beliefs (Ekstrom 1993; Christman 2003).[17] The regress problem is therefore irrelevant.

The argument that humans in fact reach an end-point in reasoning relies on the assumption that the individual's overall cognitive system and reasoning capacity do not suffer from such deflectors as idiosyncratic learning history or cognitive impairment. In such cases, reasoning might go astray. Thus consider patients, suffering from frontal lobe damage, who are unable to conclude in their reasoning and have serious difficulty reaching, or making, a decision (Damasio 1994). They seem to be stuck in an endless chain of deliberation and if they act, they will do so poorly. According to Damasio, they act in an irrational way.

To Dworkin's pair of replies we might add a third comment – one that questions the relevance of the infinite regress problem. The regress objection is in a certain sense incompatible with what is claimed by hierarchical theories. These theories presuppose, on my

[17] We will return to this issue in Chapter 3.

understanding, that human cognition is limited because humans have a finite intelligence. For instance, in a decision-making situation the individual has a limited range of options to reason about and a finite amount of time for processing. It is unlikely that there are great numbers of higher-order levels that the individual has go through in a finite period of processing. Such "super tasking" seems implausible. As Frankfurt claims, that would be a sign of humanization running wild (Frankfurt 1971, p. 16).

Finally, if it were true that there are an endless number of levels, and that it cannot be known at which level autonomy is secured, it would be impossible, in a sufficiently stringent way, to determine whether an individual is autonomous. From a practical perspective, this would render the concept of autonomy insignificant. The claim that autonomy cannot be determined seems incorrect, because if the concept really could not be determined it would be hard to defend the idea that autonomy is as important as is so often claimed.

The extent to which second-order reflection is considered by hierarchical theories raises the question of whether autonomy is a matter of degree. I will deal with this question in the next section.

(ii) The degree problem

Are all individuals autonomous to the same extent, or are individuals autonomous to different degrees? Alternatively, is autonomy a property an individual either has or does not have? In the autonomy debate, it is often claimed that autonomy, whether it concerns actions, desires or individuals, is a matter of degree: actions, desires or individuals are claimed to be more or less autonomous (Lindley 1988, p. 69; Beauchamp & Childress 2001, p. 59; Oshana 2006, p. 31).

Below I shall suggest that the concept of autonomy, when it is understood as a metacognitive capacity of the individual, is not to be understood in terms of degree, but rather as an all-or-nothing concept, i.e. as a concept picking out a property one either has or does not have. However, let me first present three arguments for the degree view. Each will be criticized in favour of an all-or-nothing understanding.

(i) As I understand him, Dworkin holds that, because individuals can be more or less reflective, they can exercise their second-order capacity to a lesser or higher degree. He states:

> If we think of the process of reflection and identification as being a conscious, fully articulated, and explicit process, then it will appear that it is mainly professors of philosophy who exercise autonomy and that those who are less educated, or who are by nature or upbringing less reflective, are not, or not as fully, autonomous individuals. But a farmer living in an isolated rural community, with a minimal education, may without being aware of

> it be conducting his life in ways which indicate that he has shaped and molded his life according to reflective procedures. (Dworkin 1988, p. 17)

Following the comment made in the above quote, we might say that whether autonomy is a matter of degree can reasonably be understood in terms of how frequently an individual reflects. However, it can also be determined by the detail in which an individual reflects, i.e. how deeply and thoroughly she reasons about her first-order mental states. These two facets of second-order thought views – how often and how deeply an individual reflects – can be seen as arguments for the view that autonomy is a matter of degree. However, they can be criticized.

Severe cognitive impairment might seriously undermine autonomy. However, an individual who suffers from it may often reflect and develop SOMSs that are directed on her FOMSs. Yet, plausibly, she does not possess a high degree of autonomy. As I see it, the frequency and depth of second-order reflection does not guarantee autonomy. An individual might exercise a high degree of second-order reflection in respect of both frequency and depth, but at the same time lack autonomy because she has impaired cognition. Consider, for instance, severe psychiatric symptoms, like delusions, that cause the individual to suffer from false beliefs about her own self and about reality.

(ii) A second argument for understanding autonomy as a matter of degree relates to the idea of full autonomy. It is often claimed that autonomy is just an ideal (Dworkin 1988, p. 10). On ideal views, individuals can be claimed to be more or less autonomous with respect to their dispositions, attitudes, desires and reasons. However, they are never able to become fully autonomous – or so it argued. Thus, to have full autonomy is just an ideal. This argument can also be challenged.

An individual who is claimed to be more or less autonomous is still autonomous. This argument speaks in favour of a conception of autonomy as an all-or-nothing concept. Compare bicycling: irrespective of whether you are more or less skilled in bicycling, you either can bicycle or cannot.

To claim that an individual cannot have full autonomy, and that autonomy is just an ideal, is therefore to presuppose that the concept of autonomy is understood in terms of degree. So, unfortunately, these ideal formulations presuppose what they aim to show.

The concept of autonomy, interpreted in terms of metacognition, is better understood as something one either has or does not have, i.e. as an all-or-nothing concept. It seems to be a category mistake to hold that the concept of autonomy is to be understood in terms of degree.

In the current autonomy debate it has not been clarified what autonomy really is, and there are no convincing arguments showing that the degree view is more plausible than the all-or-nothing view. Even if we assume the degree view, we are still need to explain what autonomy is.

With the degree view of autonomy, it would also appear to be difficult to establish a threshold value that distinguishes autonomy from non-autonomy. Where is the limit to be drawn in the degree spectrum? One might think that a threshold could be set as a guideline. However, Lindley (1986, p. 69) reasonably compares this difficulty with the hard task of determining baldness.[18]

(iii) A third argument for the degree view of autonomy is that individuals are not born autonomous. Rather autonomy develops in parallel with the development of the cognitive abilities of the kind needed in reasoning (see Lindley 1986; Noggle 2005). To clarify, autonomy is not innate but develops gradually, i.e. in degrees.

Noggle (2005) points to the developmental aspect of autonomy by referring to the development of an individual's belief system, preference structure and reflective capacity to evaluate. Because these develop over time, so does also autonomy. Thus, the individual is not born autonomous; she becomes so, as she matures. Infants, for instance, are commonly regarded as non-autonomous because they lack several reflective skills that are required in order to achieve self-government. However, they develop these skills gradually and become autonomous. On my view, these reflective skills are metacognitive and develop over time.

However, that autonomy is not innate, and that it develops over time, are causal arguments. They give an account of *how* autonomy arises but not of what autonomy *is*. It is the latter question with which we are concerned in this thesis, and the answer to this question is not to be given in terms of degrees.

Nevertheless, although causal arguments about the way in which autonomy arises might not be philosophically relevant, the question of how autonomy develops over time, and in concert with one's cognitive development, is interesting from an empirical perspective. For instance, the skills required do not necessarily develop with age. We can illustrate this by observing simply that, as the result of impaired neurological development, a 50-year-old individual might have less developed metacognitive skills than an 18-year-old. Thus defenders of the degree view must be aware, or be willing to acknowledge, that autonomy will not necessarily develop gradually over time – their claim is just that it does so in normal

[18] An alternative approach to the determination of autonomy is discussed in 5.4.

cases. Moreover, by looking at developmental impairment in metacognition we obtain an improved understanding of autonomy as a metacognitive capacity. However, and as has been argued, the degree claim presented by, among others, Lindley and Noggle is a developmental one.

Whether autonomy is a matter of degree or an all-or-nothing concept will in part determine what kinds of intervention, according to what principles, maintain dignified care. If autonomy is a matter of degree, it might, for instance, be hard to respect a low, as opposed to a moderate, level of autonomy. A question that can be put to those who understand autonomy as a matter of degree is this: do various degrees of autonomy imply that there are different ways in which one can have one's autonomy respected? Conversely, does the all-or-nothing view imply equal respect for autonomy (as long as individuals are autonomous)? I will not develop answers to these questions here. I would nevertheless suggest taking the all- or nothing perspective because such a view makes it possible to implement guidelines and prescription in a similar way in a variety of contexts.

I conclude this section by claiming that an all-or-nothing concept of autonomy captures what autonomy is better than the idea that autonomy is a matter of degree. Even if one can exercise autonomy more or less as a result of differing empirical circumstances, autonomy as a metacognitive capacity of the individual is a property one either has or does not have.[19] Moreover, discussion of the concept of autonomy understood as a matter of degree does not seem able to give an account of what autonomy is. The question of whether autonomy is a matter of degree rather tends to concern how autonomy arises – how it develops in humans. These are two different questions. What autonomy is and how it arises are distinct issues. The present thesis deals with the latter question.[20] It is now time to consider the third objection to hierarchical theories: the historical problem.

(iii) The historical problem

Below I shall present and discuss the historical problem as it is normally understood in the literature. Since one must take into account how both FOMSs and SOMSs were learned and acquired, I will argue that historical conditions for autonomy must be established.

---

[19] I will develop an account of metacognition in Chapter 4.

[20] In Chapter 4 it will be argued that the metacognitive skills of autonomy are in line with an all-or-nothing concept: either you have them, or you do not.

The historical conditions, as they are understood below, concern the causal history of mental states. Importantly, the causal history is not to be mixed up with the causal development of how autonomy arises (as was previously discussed with respect to the degree problem above). These two issues, the causal history and the causal development, are not similar. The causal history concerns cognition as such, independently of its development.

Hierarchical theories that highlight the role of identification do not take into account the circumstances under which an individual identifies with (say) a first-order desire or belief (Christman 1988; 1991; 2005). According to the theories, it is sufficient that an individual has the capacity to identify with it. However, as Mele (1995) objects, autonomy requires that the causal history behind an individual's desires and beliefs is adequate. According to Christman, "Autonomy must be seen in light of a person's history – the various conditions that have gone into the shaping of her present character" (Christman 2005, p. 279). The historical objection put forward by Christman is well known, but it emphasizes something important.

To neglect the historical conditions of autonomy and claim only that autonomy requires identification is to allow for several kinds of autonomy-undermining scenarios. On hierarchical theories, it seems that all kinds of connative and epistemic mental states, irrespective of their content, will maintain autonomy as long as the individual is able to identify with them. In other words, we seem to be able to insert whatever connative or epistemic mental states we choose on the first and second level. Autonomy always obtains if identification does. For instance, a willing addict will count as autonomous as long as she identifies with her first-order desire to take his drug at the second level:

FOMS = the desire to take the drug
SOMS = the desire to have the desire to take the drug

It is reasonable to regard addiction as a typical example of undermined autonomy, even under circumstances where the individual identifies with her craving. In addition, other kinds of autonomy-undermining scenarios are allowed. Consider, for example, the individual identifying with irrational impulses or delusional first-order mental states.[21]

It should be clear now, I hope, that neglecting the historical conditions of autonomy constitutes a serious problem for hierarchical theories. In this sense, hierarchical theories are problematic.

---

[21] Delusion is here understood as a psychiatric term.

Importantly, the historical conditions of autonomy must be established. This is required even where identification is replaced by acceptance. With regard to hierarchical theories, nothing is said about historical conditions.

How are we to understand the historical conditions of autonomy? Next, I will discuss two kinds of historical condition which, I suggest, have to be established in a theory of autonomy.

The first condition must reflect the relevant FOMSs, i.e. be responsive to the way in which the FOMSs were initially acquired. Let me explain. The first historical condition concerns the origin of FOMSs. This condition prevents us from claiming, for instance, that severe psychiatric symptoms, like paranoid psychoses, are compatible with autonomy. External forces that are normally conceived of as undermining the exercise of autonomy would also be permitted – which is implausible.

The second historical condition concerns, the status, or accuracy, of second-order reflection as such in the development of SOMSs directed on FOMSs. This condition is sensitive to the way in which the individual reacts and reflects upon her FOMSs, and the way in which the relevant SOMSs are formed through second-order reflection. This second condition excludes cases of impaired reasoning, where the development of SOMSs is dysfunctional. As was previously claimed, one's reflective capacity might be defective as the result of *inter alia* severe cognitive impairment. Consider addiction or dementia, for instance.

As I see it, if we are to clarify the sense in which a theory of autonomy requires historical conditions, these two conditions must be explicitly distinguished. The first focuses on the origin of FOMSs, while the second focuses on reflection as such, and the way in which the individual reflects upon and develops SOMSs.

To sum up, then, the history lying behind the individual's connative and epistemic states, of both first- and second-order kinds, must be taken into account in order to answer, in a plausible way, the question about what autonomy is and what it means to exercise it. Unfortunately, hierarchical theories that solely require identification permit several external forces that are normally regarded as incompatible with the exercise of autonomy. Consider, for instance, cases of brainwashing, indoctrination, hypnosis, lying or the withholding of information, where someone else, external to the individual agent, has inculcated in her certain beliefs or desires.

In work on autonomy and the historical problem one of the most common issues discussed is manipulation (Taylor 2005). The standard understanding of manipulation, as

a psychological term, is the use of various tools to cause another individual to behave or react in a certain manner while that individual is unaware of what is going on (Psykologilexikonet 2005).

Manipulation seriously undermines the exercise of autonomy. (However, as will be argued later in 5.5, even if manipulation can correctly be said to undermine the exercise of autonomy, it does not follow that autonomy, per se, is undermined.) In hierarchical theories of autonomy, manipulation does not seem to be a problem, since what is required is mere identification. Such theories neglect historical conditions. Hence, they do not satisfy the first condition suggested above, and autonomy persists as long as identification between the levels obtains. Neither is condition two satisfied here, since as long as identification takes place, the status of any accompanying reflection is irrelevant. It is not fruitful to defend a theory of autonomy that allows for too much.

Dworkin's later hierarchical variant meets the historical criteria. It is therefore less affected by the historical problem than his earlier views were. Dworkin's solution to the historical problem is procedural independence (which was presented in 2.1). He writes:

> Spelling out the conditions of procedural independence involves distinguishing those ways of influencing people's reflective and critical faculties, which subvert them from those which promote and improve them. (Dworkin 1988, p. 18)

Lack of procedural independence will hinder an individual from exercising her autonomy. It hinders via obstacles like hypnotic suggestion, manipulation, coercive persuasion, and subliminal influence (Dworkin 1988, p. 18). Where there is procedural independence, FOMSs become accepted, rejected, or revised by the individual herself. This seems correct, since what characterizes an autonomous individual cannot be understood without reference to her own mental activity and overall cognition. By introducing the requirement of procedural independence, Dworkin's theory satisfies the first condition above. The procedural independence requirement prevents us from claiming, for instance, that a manipulated individual can be said to exercise her autonomy

Dworkin claims that one must separate two significantly different modes of evaluation, i.e. those that undermine autonomy, or hinder the exercise of it, and those that do not (Dworkin 1988, p. 161). By distinguishing various kinds of evaluation, Dworkin's theory meets the second historical condition, which obliges us to take into consideration the status of reflection in the development of SOMSs directed on one's FOMSs.

A theory of autonomy that requires identification (or acceptance) while it at the same time neglecting historical conditions is problematic, since it results in an excessively narrow, as well as excessively broad, understanding of the concept of autonomy. The understanding is too narrow because it merely requires the capacity to identify with FOMSs. As was noted by Dworkin, the second-order capacity here involves more skills (rejection and revision). On the other hand, the understanding is too broad because it allows for obstacles that seriously hinder an individual from exercising her autonomy. Manipulation is one example. In the next section, I summarize the three objections to hierarchical theories.

2.3.1. Summary of the problems

What can be concluded about the infinite regress problem? As I see it, this problem has been over-emphasized in the autonomy debate. The analysis of the concept of autonomy as a metacognitive capacity of the individual prevents the infinite regress from occurring. In this sense, the regress problem becomes less serious. Further, it seems empirically correct to assume that the allegedly troublesome regress is not likely to arise in real human cognition.

The concept of autonomy is not to be understood in terms of degrees, but it is an all-or-nothing concept. Autonomy is a metacognitive capacity one either has or lacks. However, whether this capacity can be exercised or not depends on empirical circumstances.

In a theory of autonomy, historical conditions must be referred to. I have suggested that the origin of inserted FOMSs, as well as the status of reflection and the development of SOMSs, constitute such conditions. A theory that lacks, or pays insufficient attention to, historical conditions will permit external interferences, like manipulation and brainwashing even though this kind of interference is incompatible with the exercise of autonomy. To remove the problem of manipulation, historical conditions must be taken into account in any realistic account of autonomy.

2.4. Concluding discussion

Let me conclude by raising a question. If hierarchical theories are problematic because they do not take into consideration important aspects of what it means to be autonomous, and because they therefore permit too much, should they be rejected? As has been argued, Dworkin's later hierarchical theory is the most plausible and fully elaborated one. I hope that it has been explained why. First, it solves some of the problems discussed above. Second,

Dworkin's characterization of autonomy as a second-order capacity is in line with the initial assumption presented in 2.0, since it promotes an understanding of autonomy couched in terms of a metacognitive capacity. Third, Dworkin's theory approaches the concept of autonomy from a global perspective and centralizes the individual. Fourth, it downplays the role of identification as a requirement of autonomy. However, I have suggested that the concept of identification has to be abandoned from the debate altogether: instead it should be replaced by the concept of acceptance. Finally, Dworkin's theory satisfies the two historical conditions suggested above.

Let me now point to some weaknesses in Dworkin's theory. It is at least implicit in Dworkin's theory that developed SOMSs have higher status than FOMSs. Thus, it is doubtful whether an individual who is sometimes steered by FOMSs without reflecting upon them is autonomous. In this sense, FOMSs will not be sufficient for autonomy. However, it is questionable whether second-order reflection must always takes place in order for an individual to be autonomous.

As I see it, Dworkin's theory places too much emphasis on explicit reflection: it is too demanding to require that autonomy, from a global perspective, always involves explicit reflection. Dworkin does not succeed in giving an adequate account of the various levels of reflection potentially involved in the second-order capacity with which he is concerned, even if, in one passage, he claims that reflection need not be an articulated and conscious process (Dworkin 1988, p. 17). Unfortunately, he does not develop or clarify this claim in further detail. No characterization of the various levels of reflection is given. In this sense, I believe, the role of FOMSs is inevitably downplayed. Consider intuitions, urges, or rapid responses. Do they undermine autonomy if one does not develop SOMSs that are directed on them? As I see it, FOMSs can play a role in the absence of explicit second-order reflection.

Dworkin's initial reply to this objection would be that the second-order capacity is a genuine human capacity, and that this capacity, and the exercise of it, is what characterizes autonomy. He would continue by claiming that the first level is insufficient to determine whether an individual is autonomous.

> …we fail to capture something important about human agents if we make our distinctions solely at the first level. We need to distinguish not only between the person who is coerced and the person who acts, say, to obtain pleasure, but also between two agents who are coerced. One resents being motivated in this fashion, would not choose to enter situations in which threats are present. The other welcomes being motivated in this fashion, chooses (even pays) to be threatened. A similar contrast holds between two patients, one of whom is

> deceived by his doctor against his will and the other who has requested that his doctor lie to him if cancer is ever diagnosed. (Dworkin 1988, p. 19)

This approach can be queried. I would claim, at any rate, that whether or not an individual has to develop second-order attitudes to her FOMSs to be autonomous will depend on the context. For instance, second-order reflection need not take place in situations where one has to act rapidly (e.g., flight behaviour), although it is needed in circumstances where more complex tasks have to be managed (e.g., planning one's holidays). Consider, here, demanding varieties of problem-solving, or conflict resolution, where a particular FOMS gives rise to conflict in the cognitive system of the individual as a whole. In this situation, second-order reflection is needed.

In contrast, well-learned strategies, habits or procedural skills do not necessarily require to be backed up by reflection. FOMSs need not call for reflection. This view is in line with work by Berofsky (1995; 2005) in which it is claimed that FOMSs essentially work as natural and basic motivations and therefore need not constitute obstacles to autonomy. In my view, this is probably correct. To be constantly engaged in developing SOMSs directed on one's FOMSs would be disadvantageous.

However, one should not, of course, trivialize the second-order capacity to develop SOMSs. As I see it, the capacity to accept, reject, or revise FOMSs is sometimes necessary for autonomy. If the individual lacked the capacity for rejection, she would, for instance, be unable to resist irrational impulses or unable to select relevant information. Second-order reflection enhances the process of formulating ideas about FOMSs so that one can accept, reject, or revise them on reasonable grounds.

Even if context determines whether a SOMS toward one's FOMS has to be developed, one must have in mind that second-order reflection, as well as the status of FOMSs, might be impaired, thereby undermining autonomy. Severe cognitive impairment might, for example, result in a dysfunctional reasoning capacity. Consider the frontal lobe patients mentioned earlier, or cases of psychosis and delusion. We will come back to these issues in Chapter 4 and Chapter 5.

How can we reach a plausible understanding of second-order reflection which, at the same time, properly accommodates a global perspective on autonomy? In the next chapter, Ekstrom's coherentist approach to autonomy is presented and assessed. Ekstrom's theory places an emphasis on our capacity for evaluation which, like Dworkin's theory, is in keeping with the initial assumption about metacognition presented in 2.0. As I shall argue, evaluation is central in the understanding of autonomy as a metacognitive capacity.

Ekstrom states explicitly that evaluation need not take place consciously. Her theory also provides solutions to some of the problems that strike hierarchical theories; and it understands the concept of autonomy from a global perspective in a more fully elaborated way. What Ekstrom tries to do, in my view, is to develop and suggest an alternative theory to Dworkin's, as well as to other hierarchical theories.

Chapter 3

Autonomy and Coherence

3.0. Coherence Between Mental States

The coherence theory of autonomy (CTA) developed by Ekstrom (1993; 1999; 2005a; 2005b) will now be considered. In an effort to overcome some of the problems that beset hierarchical theories, CTA gives an alternative view of autonomy: it adds the requirement that the autonomous individual's mental states be *coherent*.

In contrast with some hierarchical theories that consider solely (say) a first-order desire and a second-order desire about that first-order desire, CTA requires internal consistency between FOMSs and the individual's overall system of preferences and acceptances, i.e. the individual's evaluation system. In other words, CTA emphasizes the individual's capacity to evaluate her FOMSs with respect to her already established preferences.[22] Given these elements, CTA is, as the analysis below will show, in harmony with the concept of autonomy as a metacognitive capacity. Nevertheless, and as in Dworkin's theory, the concept of metacognition is not used explicitly in CTA. However, the evaluative capacity that CTA deals with can be understood as a metacognitive capacity.

CTA suggests solutions to some of the problems discussed in Chapter 2. There are three reasons to why CTA gives a better understanding of autonomy than hierarchical theories do. Yet in some respects CTA does not differ from Dworkin's later versions (e.g. with respect to the global perspective on autonomy). The first reason is that CTA to some extent solves, or at least mitigates, the infinite regress problem. It does this in a somewhat similar way to the way that was argued for in Chapter 2. Second, CTA solves the problem of manipulation, since it requires historical conditions. Finally, CTA states explicitly that FOMSs need not be evaluated consciously in a way of which that individual is aware. However, and as in Dworkin's theory, this claim is not elaborated in further detail.

The next section introduces and describes CTA. Thereafter CTA is discussed in connection with Lehrer's epistemological account of coherence and evaluation.[23] I then

---

[22] See also Watson (1975) and consider the idea of an evaluative and motivational system. As I understand him, Watson emphasizes the role of evaluation in reasoning.

[23] Lehrer's use of an epistemic concept of coherence to characterize knowledge and reason inspired Ekstrom's coherentist analysis of autonomy.

explain how CTA deals with the problems presented in Chapter 2. However, I also point out some problems that arise specifically for CTA. Finally, CTA is evaluated with respect to its contribution to our understanding of the concept of autonomy as a metacognitive capacity.

3.1. Ekstrom's Coherentist Analysis

CTA, as it has been developed by Ekstrom, emphasizes the capacity of the individual to reflect upon her beliefs and desires. This capacity is needed in order to solve conflicts of first-order mental states, like beliefs and desires, and to achieve control.

Autonomy requires the capacity to develop coherent preferences that are sanctioned as one's own (Ekstrom 1993, p. 608). According to Ekstrom, an autonomous individual must have the capacity to authorize her first-order beliefs and desires as consistent with her self: "*S* is personally authorized at *t* in preferring *x* if and only if the preference for *x* coheres with the character of *S* at *t*" (Ekstrom 2005a, p. 63). She further writes: "A preference that is authorized for me, counts as truly mine, as one that I really want to have, since it coheres with the other things that I prefer and accept" (Ekstrom 1993, p. 615).

It will be seen at once that the concept of coherence in Ekstrom's theory is understood in terms of consistency between the elements in the character system. The term "character" refers to the stable and coherent subset of preferences and acceptances of the individual's whole system. The whole system is controlled by her character system. It is this subset, the integrated cognitive faculty of the individual's basic acceptances and preferences, that constitutes the core self (Ekstrom 2005a, p. 59).

Ekstrom understands an individual's acceptances and preferences as

> …those that she acquires and retains in her attempt to believe what is true and to desire what is good; that is, her acceptances and preferences. Now I wish to make the proposal that we take an agent's *true* or *most central self* to be a subset of these acceptances and preferences, namely, those that *cohere* together. One's preferences, I suggest, are authorized – or sanctioned as one's own – when they cohere with one's other preferences and acceptances. (Ekstrom 1993, p. 608)

As I understand the position here, to count as autonomous an individual must have the capacity to maintain internal consistency and control through authorization of her first-order desires and beliefs. Such a capacity "…helps us to control our behavior and so to control the direction of our lives" (Ekstrom 2005a, p. 49).

However, it is important to note that the autonomous individual will also be able to defeat, or neutralize, conflicting first-order beliefs and desires (Ekstrom 1993). Such a conflict may arise, for instance, in cases of competing or irrational desires or beliefs. Thus, maintenance of internal consistency involves the authorization and the defeating or neutralizing of first-order mental states. I take Ekstrom's term "defeating" to refer to the rejection of a first-order desire or belief. I take the term "neutralizing" to refer to the mitigation of the relevance of a competing first-order desire or belief.

It is important to see that CTA shares basic similarities with other philosophical theories of autonomy that emphasize evaluation and the self. One theory that is compatible with CTA has been developed by Noggle (2005). Below I go though some of the similarities between Noggle's theory and CTA.

While Noggle does not claim explicitly that he is a coherence theorist, he argues that the self is constituted by a subset of the overall psychology of the individual, and that this subset forms a skeleton for the rest of the individual's psychology (Noggle 2005, p. 100).[24] Internal forces that operate outside the subset may undermine autonomy if they are inconsistent with it. Such forces remain internal, since they are based on the psychology of the individual.

The skeleton Noggle has in mind seems to be similar to Ekstrom's character system. As in Ekstrom's account, the self is, according to Noggle, a stable system of beliefs in conjunction with a preference structure (Noggle 2005, p. 100). According to both theories, to be autonomous one must possess the capacity to control the system.

Even where the subset is stable, Noggle does not exclude the possibility of changing the core elements – e.g. in the light of new information – if this is internally motivated. According to Ekstrom, however, any change will require effort, since the elements in the subsystem are long-lasting and immune to change (Ekstrom 1993, p. 608). A drastic shift would result in systemic instability, if I understand Ekstrom right. We might, for instance, consider thought disturbances, or the experience of a disintegrated self in schizophrenia (e.g. see Campbell 2002; Proust 2006). Multiple personality disorder is another example. However, both Ekstrom and Noggle agree that peripheral elements, outside the subset, are easier to change in the light of new information than are elements of the core self.

[24] His theory has rather been described as doxastic (Taylor 2005).

Ekstrom's basic ideas originate from Lehrer's theory of evaluation, coherence and epistemic justification (Lehrer 1999b).[25] Ekstrom writes: "…Lehrer's coherence theory of epistemic justification provides a useful springboard for developing an account of coherence among states of preference" (Ekstrom 2005b, p. 144). She further claims that it is "…natural to look to the existing literature on coherence among mental items for help in understanding autonomous agency" (Ekstrom 2005b, ibid).

In my view, the proposal in Lehrer's theory that has had most influence on CTA is the notion that coherence between what the individual prefers and her evaluation system must be established if there is to be autonomy.[26] In Lehrer's analysis of knowledge, the comparative reasonableness of what the individual accepts (or prefers) and her evaluation system must be established. Further, evaluation has to be undefeated by error (Lehrer 1999b, p. 31). As in CTA, competitors (e.g. other desires and beliefs) might render an acceptance less reasonable to maintain. The only way for the competitors to be (as it were) beaten is by the evaluation system. However, there may also be competitors which suggest that an acceptance maintained by the individual is unworthy of trust. Consider, for instance, new information which confronts the individual and which might contradict an existing acceptance in her stable subsystem. This may then, through evaluation, become neutralized, or defeated (to use Ekstrom's terminology), by the system (Lehrer 1999b, p. 29). The individual is thus able to correct errors that confront her. This account is in agreement, I believe, with Ekstrom's account of coherence, and with its impact on the problem of autonomy.

Two important distinctions put forward in Lehrer's theory are also present in CTA: the distinction between desire and preference, and the distinction between belief and acceptance. A desire that is evaluated and authorized by the individual becomes a *preference*. Hence, there is a difference between a mere basic desire and a preference. A preference is an evaluated desire that coheres with, or becomes part of, the rest of the individual's character system. Similarly, a belief that is evaluated and authorized becomes an acceptance. Preferences and acceptances are evaluated/authorized desires and beliefs concerning what one desires as good and what one believes to be true, respectively. These elements cohere and are

---

[25] See also (Ekstrom 2005a, p. 61). Ekstrom also refers to the presentation of coherence, logical consistency and belief systems developed by Laurence Bonjour. No space will be devoted to the discussion of these epistemological topics here. Interested readers might consult Bonjour (1985) and Lehrer (1990, 1999b).

[26] Lehrer discusses his theory in connection with autonomy but seems to decline Ekstrom's view (Ekstrom 1999, p. 1062; Lehrer 1999a, p. 1071). I will not develop his arguments here.

what constitute the character system. Since mere desires and beliefs are not evaluated, they might not cohere with the character system (or the evaluation system in Lehrer's terminology). Ekstrom writes:

> A preference that is authorized for me counts as truly mine, as one that I really want to have, since it coheres with the other things I prefer and accept. Thus, to identify myself with some desire is to have an authorized preference for that particular desire to be the one that leads me all the way to action, when or if I act. To identify myself with some belief is to have an acceptance regarding the content of the belief that is coherent with my character system. And to identify myself with some course of action is to act on a desire for which I have an authorized preference. (Ekstrom 1993, p. 615)

According to Ekstrom, a desire that counts as a preference, P, must be effective in action, when or if one acts. It also has to be formed and evaluated in the search for what is good (intrinsically or instrumentally).[27] In other words, P must be formed upon the basis of reasoning about what the individual evaluates as generally good. Otherwise P is not authorized for good reasons. In this sense, the search for what is good must be understood in relation to some standard of goodness.[28]

Importantly, evaluation "…need not take place at the conscious level" (Ekstrom 1993, p. 603). Unfortunately, Ekstrom does not elaborate this claim. In addition, in a short passage Lehrer also claims that "Positive evaluation may occur without reflection when reflection would be otiose" (Lehrer 1999a, p. 4).

Both Ekstrom and Lehrer claim that evaluation need not be conscious. Lehrer's claim above supports the idea that the evaluation can function effectively without necessarily involving conscious awareness: it can operate in the background, on a lower level of reflection.[29] This claim will be examined in detail in Chapter 4, in connection with metacognition.

Next, I will discuss plausible solutions given by CTA to some of the problems discussed in Chapter 2.

---

[27] A quite similar process for beliefs is, I believe, reasonable.

[28] It can be sensibly argued that the individual must be able to authorize her preferences for good reasons. However, this position must also account for acceptances, as I see the matter.

[29] This might be why Lehrer uses "evaluation system" and "background system" interchangeably (see Lehrer 1999b, p. 1).

3.2. Evaluation of Ekstrom's Analysis

Below, CTA is evaluated. First, I will explain how CTA deals with the infinite regress problem that was presented and discussed in Chapter 2.

CTA suggests a plausible solution to the infinite regress problem by assuming the capacity to authorize, control and maintain internal consistency in the (evaluation) system. A stable system of this sort prevents an infinite regress from arising. More accurately, CTA solves the infinite regress problem by referring to authorized preferences. Ekstrom claims: "One's preferences, we might say, are *personally authorized* – or sanctioned as one's own – when they cohere with one's other preferences and acceptances" (Ekstrom 2005a, p. 58). Evaluation stops when a first-order desire or belief becomes authorized as a preference or an acceptance that is coherent with the system. Hence, an evaluative regress does not arise.

Let me describe an example. If no conflict between a first-order mental state and the overall cognitive system of the individual is detected, it seems unnecessary to require evaluation. However, imagine a case where there is a conflict between a spontaneous first-order mental state and the individual's overall cognitive system. In order to evaluate the first-order mental state, the individual might need to reflect upon it. For instance, a first-order desire to eat dinner is normally unproblematic and not in need of evaluation. Nevertheless, if the individual has decided to try to lose weight, the first-order desire to eat dinner might cause a conflict, and this might require evaluation of whether satisfying the first-order desire really is a good idea. Such evaluation might terminate in an acceptance or rejection of the first-order desire, and ultimately in a decision whether or not to eat the dinner. However, it seems unreasonable to suppose that the acceptance or rejection of the first-order desire, in normal cases, will in turn cause another conflict so that the individual must engage in a further process of evaluation. Evaluation in normal cases terminates on the second level. (If it does not itch, do not scratch!) Autonomy involves the capacity to develop plausible reasons in order to avoid an endless process of evaluation.

The ability to maintain consistency between one's mental states is essential for autonomy. For instance, a preference for a certain desire requires that desire to cohere with the system. On my understanding, the role of hierarchies becomes less important once we instead assume coherence between mental states – once we assume that FOMSs are consistent with the character system (i.e. the subset of preferences and acceptances). As a consequence of the coherence requirement, the individual does not evaluate a first-order desire or belief in isolation from her other preferences and acceptances. Evaluation of one's first-order desires or beliefs takes place in relation to her character system. In this sense CTA is in line with

Dworkin's global account of autonomy argued for in Chapter 2. CTA emphasizes the importance of considering evaluation in relation to the character system of the individual, not solely the congruence between a certain FOMS and a second-order attitude.

CTA offers us a plausible apparatus with which the problematic concept of identification can be avoided. According to Ekstrom (2005b, p. 152) identification, of the kind where the individual identifies with a FOMS, is better understood in terms of *authorization*. Authorization is a more precise concept than identification. It is similar to the concept of acceptance argued for in Chapter 2.

Ekstrom claims that procedural conditions are required for autonomy (Ekstrom 2005b, p. 152). Procedural conditions take into account the history and origin of one's preferences and acceptances. CTA therefore incorporates historical conditions. Clearly, CTA pays attention to the origin of one's preferences and acceptances, and the way in which new information, like that carried in first-order desires and beliefs, is treated with respect to these. For instance, for a preference, P, to be acquired via a process of evaluation, P must cohere with the rest of the individual's overall character system.

Ekstrom claims explicitly that preferences "…must not be coercively formed" (Ekstrom 2005a, p. 64). Coercive formation is here understood in terms of external interference – the kind of interference involved in brainwashing and manipulation. A coherent system that does not suffer from interference of this sort is presupposed by autonomy. First-order desires and beliefs are evaluated in relation to the stable preferences and acceptances that constitute the character system.

Moreover, CTA emphasizes the evaluation process as such: the individual must be able to evaluate against standards of goodness. This claim highlights the importance of the way in which the individual evaluates: it is not enough that she has, merely, the capacity to evaluate her first-order desires and beliefs. This, as we have seen, need not take place with the evaluator's conscious awareness.

While Dworkin does not distinguish explicitly between desires and preferences, CTA does. Moreover, Ekstrom explicitly discusses evaluated beliefs (acceptances), which are not motivations, as well as connative states, like desires. Thus, the analysis takes into account the capacity to evaluate upon both connative and epistemic first-order mental states.

In my view, CTA takes a further, and more insightful, step towards treating the concept of autonomy as a metacognitive capacity by emphasizing both connative and epistemic mental states. Unlike Dworkin's theory, then, CTA deals explicitly with the epistemic content of

first-order mental states. Dworkin's theory seems solely to concern the capacity to reflect upon one's first-order motivations, i.e. one's first-order connative states.

Let me finally point to some drawbacks of CTA. As was the case with Dworkin's theory, we are left with the need to clarify the function of lower-level reflection, i.e. reflection that is not conscious. Even if Ekstrom is explicit in claiming that evaluation need not be conscious, her position is not, in this regard, developed. In my view – and here I tend to side with Dworkin's theory – CTA puts exaggerated focus on conscious evaluation.

It is questionable whether the identity of the character system is as stable over time as CTA appears to assume. The elements in the character system are long-lasting, even if the individual is to some extent able to refashion it (Ekstrom 1993, p. 608). However, too much instability in the character system would result in a disintegrated self, as I understand it. This might be the case if the individual is not immune to changes that are outside her control. Consider for instance thought insertion in schizophrenia that was discussed in 1.2. Ekstrom claims that the system should be immune to change, since the authorized elements are our deepest attitudes (1993, pp. 607-608). Such attitudes are long-lasting and central to one's character. However, this claim presents quite a static view of the autonomous individual.

On one interpretation of CTA, the concept of autonomy can be regarded as relational with respect to one's preferences. On this interpretation, the concept of autonomy is understood from a local perspective. However, since preferences are constitutive of the character system, and since the latter is claimed to be stable over time, this conceptualization of autonomy is compatible with a global perspective. As I understand CTA, it primarily emphasizes the capacity to control one's character system. Evaluation is not regarded as something that is isolated from the individual's character as a whole. I conclude this chapter with a comparative discussion of Dworkin's theory and CTA.

3.3. Concluding discussion

If we want to analyze the concept of autonomy in terms of metacognition, which elements of Dworkin's theory and CTA should be retained? I will present five ideas here. These will be dealt with in the remaining chapters of Part 1.

(i) It is a central claim of the account put forward here that both Dworkin's theory and CTA are compatible with the concept of autonomy as a metacognitive capacity because both focus on the capacity for second-order reflection and evaluation.

(ii) Both Dworkin's and Ekstrom's analyses of autonomy as a second-order capacity are similar to theories about the ways in which mental-state evaluation works in metacognition, i.e. how evaluation is described with respect to empirical research in, for instance, cognitive psychology (Metcalfe & Shimamura 1994; Metcalfe & Terrace 2005; Proust 2007). I will deal with such findings in Chapter 4.

(iii) Both theories support the claim that evaluation need not take place at the level of conscious awareness.

(iv) The concept of autonomy is to be understood from a global perspective.

(v) The idea of maintained cognitive integration and control in the system is to be fleshed out in a way that shows we are dealing a metacognitive capacity of the *individual*.

The analyses presented in Chapter 2 and Chapter 3 explained the sense in which autonomy is a metacognitive capacity of the individual. The ability to evaluate one's first-order mental states, and to develop second-order attitudes to them, is a metacognitive capacity that autonomous individuals possess. However, developing SOMSs does not necessarily require evaluation to be conducted consciously. This is stated in both the Dworkin theory and CTA. Both theories focus mainly on conscious evaluation, however.

To conclude, then, according to Dworkin reflection need not be a fully conscious and explicit process. Unfortunately, he does not clarify what this means. Similarly, Ekstrom informs us that evaluation need not be conscious; and Lehrer states that conscious evaluation would in some cases be superfluous. However, their accounts, and the assumptions that inform those accounts, do not make explicit enough the distinction between reflexivity that takes place at the level of conscious awareness and reflexivity that takes place unconsciously and (as it were) lower down. Exactly how are these levels to be understood? Below, this issue will be dealt with. Let us now take a closer look on metacognition.

Chapter 4

Autonomy and Metacognition

4.0. Metacognition, Evaluation, and Lower Level Reflexivity

This chapter offers a detailed analysis of the concept of autonomy as a metacognitive capacity. I argue that autonomy is a metacognitive capacity of the individual. In the course of the discussion I acknowledge that this view requires us to allow that metacognition does not necessarily require *conscious* evaluation of one's mental states.

An essential component of metacognition is procedural reflexivity. Such reflexivity is routinely and takes place at a lower (non-conscious) level – or so it has been argued. The standard understanding of reflexivity is taken from Proust's theory of metacognition (Proust 2007).[30] Reflexivity – whether it takes place at the lower level or at the level of conscious awareness – is understood as the capacity to evaluate and revise *one's own* cognitive states. Thus, according to Proust, reflexivity is never about the mental states of other people. As I understand it, reflexivity always concerns the self. Before presenting Proust's theory in more detail (in 4.1.), I want to examine the way in which the general features of metacognition are normally understood.

In the existing literature on metacognition, philosophical as well as empirical theories support the claim that metacognition involves two cognitive levels, i.e. the first- and second-order level. Below, attention is paid to Nelson's model of metacognition. This is a model that harmonizes with the theories examined in Chapter 2 and Chapter 3. Thereafter I analyze the claim that metacognition involves lower-level reflexivity.

Nelson distinguishes between two cognitive levels: the object-level and the meta-level. He writes:

> At the object-level are cognitions concerning external objects. At the first meta-level would be cognitions concerning cognitions of external objects. In theory, at the second meta-level would be cognitions concerning the first-level cognitions. (Nelson 1996, p. 105)

According to Nelson, "…information flowing from the object-level to the meta-level is called *monitoring* and informs the meta-level about what state the object-level is in" (Nelson 1996, p. 105). On the other hand, "…information flowing from the meta-level to the object-level is

---

[30] To be explained in more in detail in 4.1.

called *control* and informs the object-level about what to do next (perhaps including no change from whatever the object-level had been doing)" (Nelson 1996, p. 105). He continues: "In short, the meta-level accomplishes goals by communicating back and forth with the object-level" (Nelson 1996, p. 105-106).

I suggest that first- and second-order mental states, as they are treated in this thesis, can be understood metacognitively with reference to Nelson's model. That is, the distinction between the object-level and the meta-level might be instructive in helping us to understand the kind of second-order reflection upon one's first-order mental states described in Dworkin's theory and CTA.

Consider how the first-order mental state was described in Chapter 2: the state of having a connative or epistemic mental state directed on x. A first-order mental state is simply either a motivation to do, or not to do, x, or a belief. Hence, it does not have another mental state as its object. As Frankfurt states, as regards desires, a first-order desire is simply a desire "…to do or not to do one thing or another" (Frankfurt 1971, p. 7). Consider, for instance, expressions like "I desire to x". There is nothing more to it.

Recall, also, the second-order mental state as it was described in Chapter 2: the state of having a connative or epistemic mental state directed on another connative or epistemic mental state. A second-order mental state is a mental state about a first-order mental state. Hence, it has another mental state, i.e. a first-order mental state, as its object. In Dworkin's terminology, this is understood as a second-order attitude toward one's first-order motivation.

In Nelson's model of metacognition, first-order mental states, like desires and beliefs, can be interpreted as cognitions on the object-level. On the other hand, second-order mental states can be interpreted in terms of cognitions on the meta-level, since in Nelson's model they are understood as cognitions about object-level cognitions (Nelson 1996, p. 105).

Accepting Nelson's account pro tem, I believe it is plausible to assume that what happens on the meta-level need not take place at the level of conscious awareness. However, and in connection with Dworkin's theory and CTA, I certainly think it is common to interpret the second-order capacity to evaluate one's first-order mental states as a conscious process. Nevertheless, as has been argued, the evaluation of a first-order mental state need not take place at this level. Rather, it can operate in the background, without the individual having access to it. One's having of a mental state about another mental state, then, does not necessarily involve conscious awareness. This claim, as I see it, is compatible with Nelson's model. The model seems to require, for metacognition, that there are two levels: the object-

level and the meta-level. However, it does not seem to require that control on the meta-level necessarily involves conscious representations.

The development of second-order mental states need not be a conscious and explicit process. As Beauchamp (2005) observes, with respect to hierarchical theories, autonomy might also be present in cases where the individual not has reflected upon her desires on the second level. As I see it, while evaluation of which the agent is consciously aware might be required in certain contexts, it need not be required in others. Conscious evaluation must be understood in relation to cognitive effort; and different situations require cognitive effort to a greater or smaller extent.

However, and as was claimed above, the theories examined in Chapter 2 and Chapter 3 lack, or have neglected, an articulated kind of reflexivity that plays a central role in metacognition, i.e. lower level reflexivity. They merely mention, briefly, that second-order evaluation or reflection need not be conscious or explicit. Probably it is this that resulted in the emphasis on evaluation as a conscious process.

My opinion is that, in several cases, the second-order capacity involves evaluation at the level of conscious awareness. However, from a global perspective, an autonomous individual is not constantly engaged in conscious evaluation of her first-order mental states. It would be disadvantageous continually to consciously develop second-order mental states directed on one's first-order mental states. To require that would be too demanding. It would also be too narrow a view of autonomy understood in terms of metacognition.

In a plausible and adequately elaborated metacognitive account of the concept of autonomy the emphasis on second-order evaluation, as a conscious process, must be downplayed. In metacognition, evaluation of first-order mental states need not take place at the level of conscious awareness.

The claim that our control of first-order mental states (or cognitions of the first level, in Nelson's terminology) need not be conscious can be understood with reference to the way in which procedural and declarative knowledge works.

Procedural knowledge ("knowing how") is usually understood in terms of implicit, and automatic, physical as well as cognitive skills (Dienes and Perner 1999; Proust 2003, 2007; Passer & Smith 2008). Consider, for instance: bicycling, typing, memory retrieving, or conditional behaviour. Normally, the individual does not have explicit representations of such procedural skills.

Since procedural knowledge normally does not require explicit representation, such knowledge can be claimed to be cognitively impenetrable to the individual. The concept of

cognitive impenetrability is taken from Pylyshyn's work on visual perception (Pylyshyn 1999). It is a concept that might be instructive for present purposes. Pylyshyn writes that "…an important part of visual perception, corresponding to what some people have called early vision, is prohibited from accessing relevant expectations, knowledge, and utilities in determining the function it computes – in other words, it is cognitively impenetrable (Pylyshyn 1999, p. 341). Procedural knowledge seems to work in a way that is similar to the way Pylyshyn describes vision and cognitive impenetrability. For instance, in order to be functional, the evaluation process need not be represented at the level of conscious awareness.

The possession of procedural knowledge is advantageous. As was previously observed, it would be both superfluous and disadvantageous to have constant representations of one's cognitive processes. According to Dienes and Perner, "…the advantage of procedural knowledge is its efficiency. Procedures need not search a large database because the knowledge is contained in the procedures" (Dienes and Perner 1999, p. 744). As I see it, metacognition involves precisely these kinds of procedure. This is why it is plausible to hold that autonomy, understood in terms of metacognition, involves lower-level reflexivity. Again, first-order mental states need not necessarily call for evaluation at the level of conscious awareness.

Let us now turn to declarative knowledge. Declarative knowledge ("knowing what") involves evaluative and explicitly performed skills and memories that can be verbally expressed. You can declare to yourself that you have it. Thus, declarative, as opposed to procedural, knowledge is not tacit. While procedural knowledge relates to habits, skills, and conditioned reactions, declarative knowledge concerns hypothetical reasoning and evaluation at the level of conscious awareness (Dienes and Perner 1999). This latter kind of mental state, as I see it, is what is typically described, in Dworkin's theory and in CTA, as a second-order capacity. Plausibly, theories about procedural and declarative knowledge are in line with the concept of autonomy as a metacognitive capacity.

Let me summarize the discussion so far. Metacognition involves monitoring and control of one's first-order mental states. However, first-order mental states do not necessarily call for evaluation at the level of conscious awareness. Such evaluation can occur on a lower level that is cognitively impenetrable to the individual. Evaluation is still functional at this lower

level.[31] The discussion of procedural knowledge illustrates why. Importantly, procedural and declarative knowledge are metacognitive skills.

In the above discussion a plausible hypothesis about the concept of autonomy has been put forward – namely, that autonomy, as a metacognitive capacity, involves a procedural form of reflexivity. On this lower level, first-order mental states are monitored and controlled without surfacing at the level of conscious awareness. Nevertheless, and with respect to the autonomy debate, I believe that too much emphasis has been put on second-order evaluation as a process of which the individual is consciously aware.

The hypothesis presented above raises two important questions that have to be dealt with in more detail. First, what are the main, defining characteristics of metacognition as a capacity of the individual? Second, is metacognition necessarily representational? As has been argued already, metacognition involves a procedural form of reflexivity. Next, I will discuss Proust's theory of metacognition, which explicitly states that procedural reflexivity is an essential feature of metacognition.

4.1. Proust's Theory of Metacognition

The essential features of Proust's theory of metacognition are as follows (Proust 2007). Recall the definition of metacognition put forward in 2.0, i.e. as the capacity of the individual to think thoughts about her thoughts. As Proust understands metacognition, one of its basic features is the ability to evaluate, predict and retrodict one's mental dispositions, properties and states for cognitive adequacy (Proust 2007, p. 291). By "cognitive adequacy" is meant the "…correct evaluation of the resources needed in a reasoning task, given its importance" (Proust 2007, p. 282). Evaluation, on the other hand is understood as follows:

> *Evaluating* future states involves in addition appreciating the efficiency of a given course of action, which means comparing internal resources with objective demands for the task. A judgment of learning, or an evaluation of one's emotional level, for example, involve norms of adequacy: the goal of such judgments is to find an efficient or reliable way of coping with a set of requirements. (Proust 2007, p. 281)

[31] For a discussion of cognition and its connections with conscious awareness and sub-personal system, see Brinck (2003, 2007).

Evaluation might involve trying to appreciate, or remember, a source of information, trying to predict one's ability to reach a cognitive goal, trying to learn new material, or trying to make efficient plans in a new context (Proust 2007, p. 271).

In Proust's theory of metacognition evaluation of one's mental states plays an important role. However, according to Proust, metacognition is not necessarily metarepresentational: that is, it need not involve second-order representations (Proust 2007, p. 293).[32] Rather, metacognition and metarepresentations are functionally distinct. On my interpretation, Proust means that metacognition need not involve representations of mental states that appear at the level of conscious awareness.

Proust presents several arguments for the claim that metacognition and metarepresentation are functionally distinct. Let me rehearse three of these.

First, it is reasonable to assume that there are basic differences between metacognition and metarepresentation because metacognition, as was previously claimed, need not involve second-order representations. Thus, there are forms of metacognition that are not representational, according to Proust.

Second, metacognition is always about one's own cognitive states, while metarepresentations need not be. For instance, metarepresentations are present in other-attribution, where the individual metarepresents the mental states of other people. Nevertheless, such metarepresentation does involve one's own cognitive states. According to Proust, the essential characteristic of metacognition is reflexivity, i.e. the capacity to evaluate and revise *one's own* cognitive states. Since metarepresentation can be about other people's mental states, while metacognition cannot, metarepresentation is not necessarily metacognitive.

Third, procedural reflexivity is present in metacognition but not in metarepresentation (Proust 2003, 2007). The usual understanding of metacognition relies, according to Proust, on a metarepresentational view that requires a conscious subject. However, this reliance is unfortunate (Proust 2007, p. 293). This is for two reasons, according to Proust. First, the concept of metacognition need not be explained in terms of higher-order thought. Second, metacognition does not require a representational mechanism for producing conscious thoughts. Hence, an inferential structure enabling one to reason about one's own states is not needed in metacognition.

---

[32] As I understand it, Proust treats second-order reflection solely in terms of explicit and conscious representations of one's mental states.

As one can see, Proust's theory does not require metacognition to be conscious. As was observed in the arguments put forward in 4.0, there is functional lower-level reflexivity that does not require a conscious subject. According to Proust, there are forms of self-engagement that do not occur on the person-level. Metacognition is, in its most basic form, a reflexive function "…allowing an explicit form of self-representation to eventually emerge" (Proust, 2007 p. 292). Metacognition is in part unconscious and does not, therefore, presuppose consciousness. As I understand it, metacognition can be delivered by the subject's own procedural self-knowledge. For instance, evaluation does not require metarepresentation if the activity one is engaged in is that of jumping over a ditch. "You merely use your implicit non-conceptual, dynamic knowledge; you don't need to declare to yourself that you have it" (Proust 2007, p. 279). This claim is reasonably compatible with the concept of procedural knowledge presented in the last chapter.

From an empirical point of view, and bearing in mind the above discussion about the role of procedural reflexivity in metacognition, unconscious processes play a causal role in guiding behaviour (Dijksterhuis & Meurs 2006; Dijksterhuis & Olden 2006). Empirical data on the function of unconscious thought have been claimed to facilitate decision-making and problem solving; and it has been claimed that unconscious thought serves a practical function in our life-projects (Rorty 1991; Neisser 2006). Empirical research has given credence to the hypothesis that unconscious thought is, in some situations, more effective than evaluation at the level of conscious awareness (Dijksterhuis & Meurs 2006; Dijksterhuis & Olden 2006). The empirical findings in question here concern decision-making with respect to post-choice satisfaction and the role of unconscious impact in creativity.

The hypothesis that unconscious thought contributes to effective decision-making, choosing and impression formation has gave rise to the so-called "unconscious and conscious thought theory" (UCT). UCT is reasonably compatible with the hypothesis of lower-level reflexivity in metacognition put forward above. According to Dijksterhuis and Meurs, "…UCT maintains that conscious and unconscious thought have different characteristics, making them differentially applicable in different situations" (Dijksterhuis & Meurs 2006, p. 145).

As I see it, UCT supports the idea that autonomy, as a metacognitive capacity, involves lower-level reflexivity, since reasoning in order to be functional need not take place at the level of conscious awareness.

As has already been mentioned, according to Proust metacognition does not require metarepresentation. However, the concept of autonomy as a metacognitive capacity, as we

shall see, requires reflexivity that is both metarepresentational and procedural. Yet, I think Proust emphasizes something important about human cognition with respect to human understanding of procedural reflexivity.

To sum up, then, Proust concludes that metacognition does not presuppose consciousness, and that the only perspective of interest in metacognition is one's own. This contrasts with the situation as regards metarepresentations, where reasoning about the mental states of other people can take place. However, what is the primary function of metacognition? It is to this issue that we now turn.

4.2. Metacognition and Control

According to Proust, self-guidance is an essential function of metacognition (Proust 2007). On her view, it is metacognition that makes the constitution of self-identity possible. All levels of reflexivity presuppose a form of adaptive control for practical purposes. As Proust claims, without metacognition "…it is not clear how a self might represent itself 'from the inside' and develop (more or less) coherent preferences over time" (Proust 2007, p. 293).

As was claimed above, Proust's theory of metacognition and control does not presuppose consciousness. For instance, metacognition does not require the individual to be able to verbally report her mental states or attribute them to herself (at the level of conscious awareness). Rather, metacognition is partly to be understood in terms of the self-guidance that takes place at the procedural level. Thus, metacognition does not require a fully-fledged representation of oneself (Proust 2007, p. 292).

It has been suggested that metacognition can, in a fruitful way, be linked to various aspects of executive control (Nelson & Narens 1990; Shimamura 2000). Shimamura interprets the Nelson model presented above in 4.0 in terms of control, claiming that metacognition is the control or regulation of information processing. Knowledge of executive control, it is claimed, might inform research about metacognition (Shimamura 2000). According to Shimamura, executive control is a mechanism that enhances reinforcing task-relevant processing and the inhibition of task-irrelevant processing.

A link between metacognition and executive control might better define the components of metacognition. In his article, Shimamura discusses various metacognitive skills that can be understood, as I see the matter, in terms of executive control: error correction, conflict resolution, attention, emotional regulation, and the inhibitory control of irrelevant information.

Empirical research into executive control has drawn attention to the role of the frontal lobe regions of the brain. For instance, empirical findings report a relationship between metacognitive regulation and executive control in the frontal lobe (Shimamura 2000). It has also been reported that patients suffering from frontal lobe damage show impairments in metacognitive monitoring. According to Metcalfe (1993), these patients have serious difficulty monitoring their own mental states.

As I see it, there is good reason to expect that the essential feature of metacognition is control, and that control involves reflexivity at the procedural as well as the level of conscious awareness. Further, the suggestion that we should link metacognition to executive function seems promising for future research (yet, executive function is considered as the opposite to procedural processing).

The view of control taken in the above theories of metacognition can partly be understood by reference to a common philosophical conception of control that goes back to Plato. The ancient understanding of control emphasizes the individual's capacity to regulate and resolve internal conflicts (Quinn 1998). Lack of control stems from a conflict between different parts of the soul. According to Quinn, control (in the ancient view) requires a higher part of the self that regulates the lower part.

The basic idea of a lower part of the self that is regulated by a higher part, thereby affording control, is to some extent similar to modern understandings of metacognition. For example, metacognition normally presupposes two cognitive levels.[33] The view of control presented by Quinn above also share certain features with Dworkin's theory of second-order capacity, where the first-order mental states are evaluated at the second level. This second-order capacity can be interpreted as a process in which a higher part of the self regulates a lower part.

It is reasonable to suppose that, through the second-order capacity, the individual controls her mental states. If she lacks this capacity, she would plausibly also lack autonomy. Importantly, Dworkin's theory does not explicitly claim that the concept of autonomy has to be understood with respect to control or the control of one's mental states. However, the second-order capacity that is described in his theory can plausibly be interpreted in terms of control.

---

[33] Consider, for instance, Nelson's model (1996).

In part, CTA can also be understood with reference to the ancient idea of control since it emphasizes the (evaluation) capacity of the individual to maintain control and consistency in the character system.

In Lindley's philosophical theory of autonomy, the concept of autonomy is explicitly understood in terms of control. According to Lindley, autonomy requires control over one's life (Lindley 1986). In order to be autonomous it is required that the individual is, for instance, able to regulate irrational impulses and false beliefs (Lindley 1986, p. 49).

The philosophical theories about the concept of autonomy that have been dealt with so far can be interpreted in terms of control. An autonomous individual is able to control herself because she possesses a metacognitive capacity. To conclude, the concept of autonomy as a metacognitive capacity can plausibly be understood in terms of control.

Which aspects of control will be important in a philosophical theory of autonomy? Below I shall discuss two characteristic aspects of control that I regard as important and elucidatory. In Chapter 5 these aspects – *self-knowledge* and the *understanding of relevant information* – will be connected with undermined autonomy, the latter being interpreted in terms of metacognitive impairment.

Let me begin with self-knowledge. In order to control one's own mental processes, I suggest that self-knowledge, in a minimal sense, is required. To possess self-knowledge in a minimal sense does not mean that one has to know everything about oneself. It would be very hard to know everything about one's physical condition or genetic makeup. Minimal self-knowledge requires, in a rudimentary sense, experience of oneself as the same individual over time: in the past, in the present and in the future.

Minimal self-knowledge requires immunity to error through misidentification, i.e. that the individual cannot be mistaken about the experience of herself as the same individual.[34] It also requires the capacity to ascribe one's own mental states as one's own. This is what Campbell (2002) calls the "ownership of thoughts". To be the owner of one's thoughts, the individual must, for example, experience her preferences and beliefs as originating from her own mental activity, and must take an active role in them and relate them to herself.[35]

The notion that self-knowledge is an essential aspect of control is important. For instance, lacking the sense of oneself as a subject, or not being able to ascribe one's own mental processes as one's own, would make it hard to plan for future goals and to project oneself in

[34] For a discussion of this topic, see Brinck (1997) and Campbell (1999, 2003).

[35] Recall Dworkin's condition of procedural independence.

the future (as well as in the past), to have insight into one's condition, and to experience oneself as responsible for one's actions.

CTA indirectly stresses the importance of minimal self-knowledge in requiring the self to be coherent. For instance, in order for a desire to count as a preference, the desire must be evaluated as consistent with the rest of the individual's self, i.e. her character system. In my view lack of self-integration, and alienation of one's own mental processes, deprives autonomy. Lacking the capacity for procedural independence, as it is described by Dworkin, would probably also be in keeping with this claim. (We shall return to examples of undermined autonomy and lack of minimal self-knowledge in Chapter 5, where we shall examine lack of insight into illness and thought insertion in schizophrenia.)

The ability to understand relevant information in order to deal with it effectively is the second essential aspect of control I want to emphasize.

In order to have control the individual must be able to select, understand and evaluate information that confronts her in daily life; but importantly, understanding does not require the individual to have access to all the information available. Indeed it is questionable whether this is possible at all.

That complete information is not required for understanding has also been pointed out by Beauchamp. He argues that substantial understanding is sufficient for autonomous decision-making (Beauchamp & Childress 2001, p. 59).[36] Understanding need not be complete, and it seems hard to be fully informed (Beauchamp & Childress 2001, p. 88). Nevertheless, that does not mean that one necessarily lacks adequate information. What is adequate, or sufficient, is a grasp of the facts in conjunction with relevant beliefs (Beauchamp & Childress 2001, p. 89). To understand relevant information is not the same as having full understanding. Quite what information is relevant must be determined with respect to the present situation.

For instance, it would be too demanding to expect patients to have full knowledge concerning medical alternatives and their consequences. Perhaps the physician does not have complete knowledge in order to understand fully (say) every consequence of the medical alternatives that the patient has to choose among.

Consider, for example, a medical conversation between a patient and her physician. The patient is expected to be autonomous, so she must make a medical decision herself. It is

---

[36] As one can see, Beauchamp and Childress consider here what it means for an action to be autonomous, not what it means for an individual to be autonomous. However, his line of thought anyway fits the present discussion.

questionable whether it is possible to require that she must obtain complete information from her physician in order to make a plausible decision. As was previously claimed, the physician himself probably does not have access to all of the information that could be given. It would be implausible to require that one must grasp a total amount of information in order to have control. That would be beyond our cognitive capacities.

The cognitive effort needed in order to deal with information that confronts the individual depends on context. For instance, information that is novel to the individual, or a situation where a complex decision has to be made, might require evaluation at the level of conscious awareness. On the other hand, information that has to be processed rapidly, as might happen in a threatening situation, is monitored at the procedural level. Thus the understanding of relevant information, and one's dealing with it effectively, need not always be conscious processes.

To have control an individual must be able to select the relevant information, understand it in a substantial way, and to integrate it with respect to the situation and her already existing information. The ability to sort out and understand relevant information is required for planning, and for discerning what follows from one's beliefs and desires. Consider a medical situation. In such a situation, it is reasonable to assume that the individual must be able to understand the consequences of various medical alternatives she is offered, and that she must be able to relate these to her long-term interests (Hermerén 2006). The metacognitive capacity to understand, imagine and evaluate the possible scenarios and outcomes of one's available alternatives is crucial for control. However, for control to be maintained it is sufficient that the individual is able to sort out what information that is relevant and substantial.

To sum up, autonomy is a metacognitive capacity. The essential function of metacognition is control, and therefore the essential function of autonomy is control. Essential aspects of control are minimal self-knowledge and the understanding of relevant information. Importantly, control is maintained at the procedural level as well as the level of conscious awareness.

Proust's distinction between metacognition and metarepresentation will soon be dealt with. However, let me first summarize Proust's theory of metacognition and control as it has been presented so far.

4.2.1. Summary of Proust's view

According to Proust, it is not required that one needs to metarepresent cognition in order to control it. By now this should be clear enough. The supporting argument goes like this: metacognition does not normally involve second-order knowledge of oneself, i.e. a metarepresentational re-description of what happens at the object-level is not a precondition of control. Recall the example of jumping over a ditch given in 4.1. That example illustrates the phenomenon of an individual using her implicit and non-conceptual knowledge. In such cases, metacognition takes place at the procedural level, through the individual's procedural knowledge. In such a case, the individual need not declare to herself that she possesses the knowledge. According to Proust, metacognition require neither metarepresentation nor conceptual knowledge (as the ditch example shows). Thus metacognition is not to be understood – at least, in a primary sense – in terms of metarepresentation.

The following conclusions can now be drawn about Proust's theory. First, metacognition does not require metarepresentation; hence it is not inherently metarepresentational. Second, metarepresentation is not inherently metacognitive, because we can think about a thought without thinking with it. In sum, metacognition and metarepresentation are functionally distinct according to Proust. Below I will discuss these claims, arguing that the metacognitive capacity that is essential for autonomy has two components. Special attention will be put on the capacity to represent the mental states of other people. This, it will be argued, is an essential component of metacognition as concerns the concept of autonomy.

4.3. Metacognition and Metarepresentation

Proust's theory is consistent with the concept of autonomy as a metacognitive capacity since it accounts for autonomy in terms of procedural reflexivity that describes lower-level metacognitive functioning. As noted above, lower-level reflexivity is merely hinted at, and not explicitly mentioned in Dworkin's theory and CTA. The relevant claim is that second-order reflection or evaluation not necessarily is an articulated and conscious process. This claim, it was argued, is in need of development. However, the concept of procedural reflexivity offered an explicated understanding of lower-level reflection. Metarepresentation, as understood in Proust's theory, is consistent with the second-order capacity described in Dworkin's theory and in CTA. Moreover, Proust's theory of metacognition is described from a control-perspective that is compatible with the understanding of autonomy in terms of metacognition.

However, Proust claims that "…although metarepresentations can redescribe metacognitive contents, metarepresentation and metacognition are functionally distinct" (Proust 2007, p. 271). As was argued, metarepresentation is distinct from metacognition in the following key ways.

First, metacognition is always self-referential and reflexive while metarepresentation need not be. Metacognition is possible without metarepresentation.

Second, metacognition is normally procedural. As I understand Proust, procedural reflexivity is normally not explicit (i.e. apparent) to the individual; it is "cognitively impenetrable" (to use Pylyshyn's terminology). Still, such reflexivity is essential to metacognition. Explicitness can *eventually* emerge in metacognition if I understand Proust correctly.

Third, metacognition is possible without metarepresentation. Metarepresentation involves mental concepts; metacognition need not do so. According to Proust it is plausibly the case "…that metacognition does not require a mentalistic metarepresentational capacity". She continues: "…cognitive adequacy can be a goal for a cognitive system unable to metarepresent its own states as representations" (Proust 2007, p. 285).

Thus, metacognition is, according to Proust, primarily procedural and reflexive, whereas metarepresentation is neither procedural nor, necessarily, reflexive (i.e. because metarepresentation need not be about the self). Moreover, metacognition does not require a conscious subject who is able to represent her own mental states.

Is it necessary to separate metacognition and metarepresentation in the way Proust does? I think it is, since metarepresentation need not be self-reflexive. Nevertheless, I would claim that metarepresentation is an essential component of metacognition as it occurs in autonomy.

Proust's theory of metacognition understood with respect to the concept of autonomy, is, I believe, too minimal. While Dworkin and CTA tend to exaggerate the role of conscious evaluation in reasoning, Proust's theory of metacognition tends to exaggerate the role of procedural reflexivity, treating it as the sole characteristic of metacognition. Her theory also seems to neglect an important part of metacognition – namely, that metacognition partly is an intersubjectively grounded capacity. Next, it will be argued that the capacity for higher-order reasoning, i.e., metarepresentation, not only includes one's own thoughts. To reason about oneself requires that one can reason about other people. This intersubjective ability is essential for autonomy, or so it will be argued.

According to the present understanding of the concept of autonomy, the form of metacognition that is relevant for autonomy has two components. The first is procedural

reflexivity, which occurs at the lower level. Procedural reflexivity is self-reflexive, i.e. it is about the self. The second component is metarepresentation, which consists of inferential reflexivity and other-attributiveness. Inferential reflexivity takes place in evaluation at the level of conscious awareness. The exercise of declarative knowledge is a typical instance of inferential reflexivity. Other-attributiveness is the intersubjective component of metacognition and involves the ability to represent the emotional states, intentions and beliefs of other people (to be dealt with soon). To summarize, an adequate understanding of autonomy as a metacognitive capacity involves metarepresentation with respect to both inferential reflexivity and other-attributiveness.[37]

Importantly, on the present account, autonomy is *still to be understood in terms of control*, because control – as I shall soon argue – requires the capacity to represent other people's mental states and deal with them effectively. Since other-attributiveness can be understood in terms of control, it can be combined with the concept of autonomy as it has been analyzed so far.

It seems reasonable to assume that autonomy involves the ability to represent other people's mental states, and deal with others effectively, in order to maintain control. In my view one cannot develop a sense of self without the interaction with others. For instance, in order to control oneself and function in society, one must be able to cope with, and understand, other people to a reasonable extent. Therefore, the second part of metarepresentation, i.e. other-attributiveness, is essential for autonomy. An argument from intersubjectivity will explain why.

The capacity to grasp intersubjectivity makes it possible to form adequate beliefs about the world and oneself. It also elucidates the importance of being able to understand the mental states of other people – e.g. their intentions, emotions and beliefs. The view that intersubjectivity plays an important role for autonomy is discussed by Agich. He writes:

> …the world of experience does not occur in an "objective" space, but is essentially defined intersubjectively. This means that autonomy involves an essential connection with others. (Agich 2004, p. 297)

[37] The two components of metacognition make it possible to test for autonomy: that is, the components of metacognition make possible to determine whether an individual is autonomous. We will come back to this issue in Chapter 5.

Importantly, whether an individual is autonomous must be understood in connection with how well her own perceptions correspond with other's perceptions of the world and herself. To require intersubjectivity for autonomy is to take into account the content of one's beliefs with respect to established (or well-known) facts. Intersubjectivity requires that the beliefs held by the individual be to some extent anchored in what is generally claimed to be reasonable. The individual must be able to exercise control and check that her beliefs are, to a degree at any rate, reasonable given what other people regard as reasonable. Beliefs cannot solely be checked for consistency on subjective grounds.[38] As I see it, humans share a common world intersubjectively, and they seem to experience this world in a quite similar way (Davidson (1991, 1992; Brinck 2004). When the mechanisms that are crucial for interpreting external and internal inputs become impaired, autonomy, and thus control, seems to be out of reach.

An example of failure in representing other people's mental states is the sociopath – an individual who lacks empathy and has difficulty relating to other people and their intentions. Consider also the impulsive psychopath as described by Berofsky:

> Only a person who is capable of appreciating the distinctive satisfaction of long-term projects and extended commitments will be able to acquire a deeper understanding of the objects and persons around him. In the absence of this sustained interest, the impulsive fails to be able to utilize critical and emotional resources to test hypotheses, suspend judgement, explore relationships, and revise earlier conclusions in light of new experience. (Berofsky 1995, p. 65)

Berofsky further writes:

> Although the dangers of impulsiveness are evident − those who act in haste often repent in leisure − one particular impairment, the reduction of freedom and autonomy, is especially relevant. For the impulsive personality is less likely to understand the nature of the objects to which he directs his attention and is less likely, therefore, to have relevant knowledge. (Berofsky 1995, ibid)

It is reasonable to assume that other-attributiveness, as an essential metacognitive component, is an element of control. I agree with Proust that metacognition need not be metarepresentational and that metarepresentation need not be metacognitive. However, that metacognition partly is an intersubjectively grounded capacity is not considered in Proust's theory of metacognition. I suggest that the intersubjective capacity explained above is to be regarded as one of the essential components in understanding the concept of autonomy in terms of metacognition.

---

[38] We will return to subjective and external conditions of autonomy in Chapter 5.

Picking up threads from the above discussion, I want to return to the infinite regress problem that was discussed in Chapter 2. The metacognitive analysis of autonomy presented in the analysis above mitigates the infinite regress problem in certain respects that were not dealt with in Chapter 2. Below, the focus will turn to the functional role of lower-level reflexivity and its impact on the infinite regress problem.

The regress problem presented in the existing literature arises when an individual evaluates at the level of conscious awareness. The role of second-order evaluation has been questioned by Berofsky (2005). He tries to mitigate regress by questioning whether unconsidered desires and beliefs really create a barrier to autonomy (2005, pp. 64-65). He writes: "The unselfconscious use of the belief as an assumption in decision making is not an automatic barrier to the autonomy of the process should the agent accept the (true) belief because of its basis" (Berofsky 2005. p. 61).

According Berofsky, not all of our desires and beliefs can be critically assessed. For instance, some of our basic desires and beliefs were set before the time at which we became critically capable.[39] In Berofsky's words, such desires and beliefs "…just are" (Berofsky 2005, p. 64). However, the barrier to autonomy does not depend on whether the individual has reflected upon these states or not. Let us label mental states that have not been reflected upon by the individual "unconsidered". Unconsidered basic desires and beliefs can be functional, since they might be effective in action. Therefore, one need not critically assess all one's desires and beliefs. For instance, suppose the individual learns that P as a child and never needs to question it as she ages and matures. Her acting on this belief need not threaten autonomy on the assumption that the belief is reasonable.

If Berofsky is right to suggest that unconsidered desires and beliefs need not be barriers to autonomy, no regress arises, because no evaluation at the level of conscious awareness takes place. Berofsky's claim is in keeping with Proust view of procedural reflexivity. We do not need to evaluate all our first-order mental states for cognitive adequacy. Some of them are effectively dealt with at the procedural level.

The regress problem, as I understand it, concerns evaluation at the level of conscious awareness. However, and as was argued in Chapter 2, the infinite regress becomes irrelevant when we understand autonomy as a metacognitive capacity in the sense argued for here. It is halted by this capacity at both the procedural level and the level of conscious awareness.

---

[39] Quite when seems to be an open question.

It is time to conclude the above discussion. It has been suggested that metacognition has two components, and that autonomy is constituted by these: procedural reflexivity and metarepresentation. Metarepresentation, in turn, can be divided into inferential reflexivity and other-attributiveness. Thus, according to this view argued for here autonomy requires metarepresentative capacities. Metacognition is partly metarepresentational as concerns the concept of autonomy.

However, the role of conscious evaluation of first-order mental states has been over-emphasized in the autonomy debate, while correspondingly lower-level reflection has been neglected. A theory of metacognition that involves procedural reflexivity as one of its components is able to give an account of the function of the lower-level reflection that is mentioned, but not analyzed, in Dworkin's theory and CTA.

The next section will deal with the components of metacognition presented above. Metacognition will be examined in the light of the claim that autonomy is to be understood from a global perspective. Hence, we will return to the discussion of global and local perspectives on autonomy. As concerns other-attributiveness, it will be argued that autonomy is best viewed relationally with reference to empirical circumstances. Attention will also be drawn to the functional role of emotional mechanisms in metacognition.

4.4. Global autonomy, Independence and External factors

The concept of autonomy can be understood from either a global or a local perspective (Dworkin, 1988; Beauchamp & Childress 2001; Oshana, 2006). However, it is a common problem in the autonomy debate that the local and global perspectives tend to be conflated. It is sometimes hard to know whether it is the autonomous action or the autonomous individual that is being dealt with in a theory. This often creates a problem in analyses of autonomy (Oshana 2006). The autonomy debate is confused because the distinction between an autonomous action and an autonomous individual is blurred: the former refers to a local perspective while the latter refers to a global one.

As was claimed in Chapter 1, in this thesis the focus is on the global perspective: I am interested in the metacognitive capacity of *the individual*, not in certain actions or situations. Below I will deal with the difference between global and local autonomy; and I shall try to explain in what sense a global perspective is compatible with the present inquiry.

Both Dworkin's and Oshana's theories deal with the concept of autonomy from a global perspective. Both are therefore compatible with the present inquiry. According to these

authors, autonomy characterizes the individual as a whole, not particular acts. As Dworkin remarks: "…I am not trying to analyze the notion of autonomous acts, but what it means to be an autonomous person, to have a certain capacity and exercise it" (Dworkin 1988, pp. 19-20). According to Oshana, a global perspective will concentrate explicitly on the individual's personal and social life. Oshana writes:

> The difference between the local and global notions is evident in the fact that a person's degree of global autonomy is not fully determined by facts about how autonomous or nonautonomous the person is *vis à vis* particular choices. A person is autonomous in the global sense, the sense that is our concern here, only if she manages her life. (Oshana 2006, p. 2)

In the global perspective, autonomy is a metacognitive capacity that expands over various contexts. The concept of autonomy, understood from a global perspective, considers individuals as autonomous agents who act over a period of time and in doing so exercise their autonomy in various contexts. On my understanding, autonomy, understood from a global perspective is a metacognitive capacity that one either has or does not have, irrespective of context.

Beauchamp and Childress, on the other hand, understands the concept of autonomy from a local perspective (2001, p. 58). A local perspective typically considers the autonomous action in a certain situation, rather than enquiring into the autonomy of the individual. When the autonomy of the individual is considered from a local perspective, the focus is normally on the individual's *competence* to make a decision in certain situation.

From a local perspective, the concept of competence is usually understood as the "…ability to perform a task" (Beauchamp & Childress 2001, p. 70). Beauchamp states that "…the criteria of particular competencies vary from context to context because the criteria are relative to specific tasks" (ibid). In the local perspective competence is such that an individual might be competent in a certain task in one context, but not in another (Buchanan and Brock 1989; Charland 1998; Beauchamp & Childress 2001).

The local understanding of autonomy couched in terms of competence focuses on the action performed by the individual in a specific situation. Such a relativistic view is unfortunate for present purposes. As was claimed in Chapter 1, the reason for why the individual is in focus is that the analysis is meant to facilitate the implementation of certain health care guidelines and prescriptions that concern the autonomy of the individual, for instance her right to govern herself in the everyday life. Thus, the metacognitive capacity of

the individual is not to be assessed relative particular acts, but with respect to extended periods in the future: in various kinds of situations in society.

Does the global perspective require the individual to be constantly engaged in exercising her metacognitive capacity? The answer is no. The exercise of autonomy can be temporally undermined in certain empirical circumstances. Cases of emergency, and cases when the individual is in a state of shock, illustrate that. Such temporary losses are compatible with a global perspective. As I see it, the global perspective must allow that an individual can become temporally unable to exercise her autonomy, even if she is autonomous in a more general sense. A similar view has been put forward by Christman (1989), who claims that an individual can be autonomous in general while lacking autonomy in specific situations. In my view, an individual who has the metacognitive capacity can be hindered temporally in the exercise of it as a result of the empirical conditions that obtain. Consider manipulation, for instance. We shall return to this issue in Chapter 5.

Autonomy is undermined when metacognition does not function adequately in a global sense. It is eroded, that is to say, when metacognition does not operate in a number of contexts and over extended periods – when metacognition breaks down and undermines the individual's functioning over longer periods. For instance, global autonomy is undermined when the metacognitive capacity is impaired, generating too many bugs for the individual.

That empirical circumstances can impair the exercise of autonomy raises the question whether the individual must be independent of external influences in order to exercise her autonomy. I want now to address the question whether the concept of independence is compatible with the concept of autonomy, as the latter has been analyzed so far.

4.4.1. Independence

Autonomy is sometimes understood in terms of independence (Dworkin 1988, p. 6; Christman 1989, p. 3). Below it will be argued that this approach to autonomy can plausibly be understood only in a weak sense. One must therefore distinguish between two views of independence: strong and weak.

According to a strong view of independence, both external and internal influences can undermine the exercise of autonomy. By contrast, a weak view would not regard external and internal influences as necessarily undermining the exercise of autonomy. Indeed some of these influences facilitate autonomy. If we assume a weak view of independence, some

misunderstandings and confusions about autonomy interpreted in terms of independence can be avoided. Let us now take a closer look at what the strong view means.

4.4.2. Strong Independence

In the autonomy debate it is sometimes complained that theories of autonomy tend to isolate the individual from her social relations and personal interests. Theories of this kind can be understood as hyper-individualistic and atomistic.[40]

On a strong interpretation of independence, autonomy requires one to be independent of influences that are external as well as internal.[41] In this sense, autonomy is the opposite of heteronomy. The latter is an atomistic view which excludes the role of both external and internal influences in reasoning and decision-making. It is a pure rationality conception of autonomy (Lindley 1986, p. 17). Humans are autonomous, as they have the capacity to validate their actions independently of their own interests, and in the absence of controlling constraints. On the strong interpretation, autonomous individuals are able to take an impartial position in reasoning that does not involve emotional influence or external factors. Taylor takes this conception of autonomy to point to an individual whose "…will is entirely devoid of all personal interests" (Taylor 2005, p. 1). However, according to Lindley, "To be autonomous is to act on self-chosen principles" (Lindley 1986, p. 28).

That a strong view of independence delivers an implausible account of autonomy, as concerns individuals, is not difficult to understand. If this view were correct, autonomy would presuppose a pure kind of reasoning – a presupposition that seems hard to defend (Christman 1989, p. 10).

The strong view implies that humans are "cold" calculators. This is both misleading and confusing since emotions, according to this view, are not part of "reasoning". For instance,

---

[40] See, for instance, Lindley (1986), Oshana (2001), Christman (2003, 2004), Christman and Anderson (2005).

[41] This claim raises a question about substantial independence that is sometimes discussed in connection with the concept of autonomy (see Dworkin 1988; Christman 2003). To claim that an individual is independent in a substantial sense is to say that she is autonomous irrespective of certain values and lifestyles. Substantial independence theorists ask whether autonomy requires certain values that the individual has to relate to, and whether only limited sorts of lifestyle are compatible with autonomy. As this thesis does not treat autonomy as a moral concept, I will not go into this question. However, as I see it, the analysis of autonomy should be neutral vis-à-vis the various kinds of lifestyle and value. The aim of analyzing autonomy in relation to lifestyle is a quite different kind of project to that of trying to give an account of the concept of autonomy in terms of metacognition.

emotional influence would undermine the exercise of autonomy because emotions, according to the strong view, are not part of reasoning. Moreover, a strong view demands too much for a finite intelligence and is not in line with the way in which human cognition works. In fact, a strong view seems to result in some kind of hyper-intellectualism, given which few individuals, if any, would be ascribed autonomy. In addition, it conflicts with empirical data on the influence of emotions in reasoning.[42] It is questionable whether it would be fruitful to understand the concept of autonomy as the ability to set aside one's own personal preferences and values. A theory of autonomy cannot sensibly exclude the impact of an individual's personal history, desires and intentions.

To conclude, then, the strong view of independence fails to give a plausible account of autonomy. It is too atomistic, since it seems to exclude such factors as social relations and one's own preferences. Of course, it is possible that there are individuals who are independent in a strong sense. However, it seems misleading to insist that those individuals are typical and illuminating exercisers of autonomy. Humans are independent of neither external nor internal influences in a strict sense. Some external influences indeed facilitate the exercise of autonomy.

In the next section, and before I discuss the weak view of independence, I shall explain the functional role of emotion in metacognition. I wish to argue that metacognition involves emotional mechanisms.

4.4.3. Metacognition and Emotion

Emotional mechanisms influence reasoning, but not necessarily in an unfortunate way. Hume noted that emotions play a central role in reasoning, functioning as motivators (Hume 1739/2004). However, it has been disputed what role emotions play in cognition. A common view is that emotions disturb reasoning. Damasio explains:

> The "high-reason" view, which is none other than the commonsense view, assumes that when we are at our decision-making best, we are the pride and joy of Plato, Descartes and Kant. Formal logic will, by itself, get us the best available solution for any problem. An important aspect of the rationalist conception is that to obtain the best results, emotions must be kept *out*. (Damasio 1994, p. 171)

[42] This issue will be dealt with next.

The claim that emotions disturb reasoning has been criticized (Damasio 1994; Charland 1998; Ledoux 1998; Finucane et al. 2000; Slovic et al. 2007).[43] According to Damasio, the cool reasoning described in the quotation above is not practicable:

> At best, your decision will take an inordinately long time, far more than acceptable if you are to get anything else done that day. At worst, you may not even end up with a decision at all because you will get lost in the byways of your calculation. Why? Because it will not be easy to hold in memory the many ledgers of losses and gains that you need to consult for your comparisons. The representations of intermediate steps, which you have put on hold and now need to inspect in order to translate them in whatever symbolic form required to proceed with your logical inferences, are simply going to vanish from your memory slate. You will lose track. Attention and working memory have a limited capacity. In the end, if purely rational calculation is how your mind normally operates, you might choose incorrectly and live to regret error, or simply give up trying, in frustration. (Damasio 1994, p. 172)

On a strong view of independence, an individual is heteronomous if emotions influence her reasoning. However, from a biological perspective, emotions are functional. They are "…central aspects of biological regulations" and "…provide the bridge between rational and non-rational processes, between cortical and subcortical structures" (Damasio 1994, p. 128). Where an individual is rational, emotions, according to Damasio, play a crucial role. Indeed according to both Damasio (1994) and Charland (1999) emotions are indispensible in understanding, for example, goal-directed behaviour and rational responses to internal and external inputs. Both Damasio and Charland claim that emotions play a functional role in decision-making.

Empirical data have led philosophers to develop theories implying that emotion and cognition cannot be separated from each other, i.e. emotion and cognition are not necessarily competitive (Charland 1998, 1999). As Damasio writes: "The partnership between so-called cognitive processes and processes usually called 'emotional' should be apparent" (Damasio 1994, p. 175). Therefore, emotions are not necessarily non-cognitive. They rather, as cognitive mechanisms, perform certain functions (Charland 1998, p.71).

Charland gives several reasons for the view that emotions are cognitively functional. First, emotions are information-providing. Second, they are goal-directed. Third, they have a logical structure and are cognitively representational. For instance, they are reason-giving. Finally, they motivate action. According to Charland, emotions perform positive functions beyond the realm of decision-making, although it is primarily decision-making that is dealt with in his published work.

---

[43] I do not take any particular position on, for instance, Kantian interpretations of autonomy.

Charland maintains that "…emotions are one important source of value" (Charland 1999, p. 369). Emotions assist the individual in selecting the relevant options, and they facilitate the formation of preferences and values. Moreover, emotions make it possible to develop certain attitudes to one's own mental states. Charland claims that

> …emotions function as an information processing system designed to keep us appraised of our own internal states and conditions. This is the "feeling" or affective representational dimension of emotion. (Charland 1998, p.71)

And he argues that some of the cognitive capacities (e.g. the capacity for appreciation, to be dealt with in 4.4.5) that are normally claimed to be requirements of mental competence are based on emotional mechanisms.

Empirical findings in neuroscience support the idea that emotions play a functional role in human cognition and decision-making (Ledoux 1998). As regards the link between emotional regulation and metacognition, it has been claimed that the same brain activations are involved in cognitive regulation and emotional regulation (Shimamura 2000).

Research also shows that impaired prefrontal and emotional mechanisms result in difficulties in reasoning and making decisions (Damasio 1994). Damasio's data show that frequent failure in decision-making can be traced to damage in the prefrontal lobe regions and the limbic system of the brain. Patients suffering from prefrontal lobe damage are frequently unable to make decisions. The inability appears to result in defects in the later stages of the reasoning process (Damasio 1994).

Prefrontal lobe patients display a kind of pure reasoning behaviour. Damasio's clinical research on such patients indicates that when emotional mechanisms are impaired the individual's life, both personally and socially, become troublesome. Consider the case of Elliot, for example, who lost his capacity to reason and make decisions (Damasio 1994). Elliot's prefrontal and limbic deficits decreased his ability to reach a final decision in reasoning; and when he did reach a decision, he acted badly (Damasio 1994, p. 50). However, Elliot's reasoning capacity in the early stages of the reasoning process was not defective. According to Damasio, the problem lay instead in the final stages of reasoning, just before decisions were made. Importantly, while his emotional mechanisms were defective, Elliot exhibited normal higher-order neuropsychological functions. Damasio writes that "…the defect was accompanied by a reduction in emotional reactivity and feeling" (Damasio 1994, p. 51). He continues:

> …the cold-bloodedness of Elliot's reasoning prevented him from assigning different values to different options, and made his decision-making landscape hopelessly flat. It might also be that the same cold-bloodedness made his mental landscape too shifty and unsustained for the time required to make response selections, in other words, a subtle rather than basic defect in working memory which might alter the remainder of the reasoning process required for a decision to emerge. (Damasio 1994, p. 51)

In the strong view of independence (discussed in 4.4.2), the capacity to reason independently of one's emotional states and personal interests seems to be in conflict with what is known today about emotional mechanisms and their role in reasoning and decision-making. As Damasio claims, the cool strategy shown by patients like Elliott is to be understood in terms of prefrontal damage, not in terms of the ways in which humans normally behave in decision-making situations (Damasio 1994, p. 172).

Since autonomy is a metacognitive capacity, and since emotions are to some extent cognitive, an understanding of emotional mechanisms is required in the present analysis. That is, an analysis of autonomy in terms of metacognition must consider clinical research and empirical data on functional emotional mechanisms as well as dysfunctional ones. As was claimed, emotional mechanisms play an important role in reasoning and decision-making while dysfunctional ones deprive an individual's decision-making capacity. It seems plausible to hold that emotional mechanisms play a fundamental role in metacognition. For instance, without emotional mechanisms, in what sense can we say that an individual takes a certain attitude to her mental states?

The case of Elliot above illustrates the notion that emotions assist reasoning. Damasio's empirical data offer an improved understanding of the link between reasoning and emotion. The neurological damage described has consequences in the personal and social domains of an individual's life. Damasio's frontal lobe patients, I would claim, illustrate undermined autonomy from both a local and a global perspective. The research indicates that the frontal lobe and limbic regions of the brain are vital to the functioning of a controlled individual. With respect to metacognition, emotional mechanisms help the individual to maintain control.

Empirical findings on emotional mechanisms in cognition and prefrontal damage have led to the so-called Somatic Marker Hypothesis proposed by Damasio (1994). This hypothesis is relevant to metacognition. Below, I investigate it and ask in what ways emotions function in metacognition.

4.4.4. The Somatic Marker Hypothesis

The Somatic Marker Hypothesis (SMH) claims that emotions function as automatic responses that assist reasoning and the decision-making process. According to Damasio, "The terms reasoning and deciding usually imply that the decider has knowledge (a) about the situation which calls for a decision, (b) about different options of action (responses), and (c) about consequences of each of those options (outcomes) immediately at future epochs" (Damasio 1994, p. 166). Somatic markers help the individual to select, or make choices, among options. These automatic responses are what Damasio calls *somatic markers*. In general, the function of somatic markers is to facilitate and make effective the decision-making process. Moreover, somatic markers are claimed *probably* to increase the accuracy in decision-making (a claim to be explained below).

Somatic markers facilitate the decision-making processes because they reduce the number of possible options and help the individual to select relevant ones: "The automated signal protects you against future losses, without further ado, and then allows you *to choose from among fewer alternatives*" (Damasio 1994, p.173). In this sense, somatic markers assist reasoning about one's alternatives.

Conscious reasoning is not required in order for somatic markers to be functional. In reasoning and decision-making processes, somatic markers operate accurately on both the overt and covert level (Damasio 1994, p. 174). As automatic responses, they need not necessarily reach consciousness (Damasio 1994, p. 185). Yet importantly, such automated responses can be described as "gut feelings".

Damasio states that "While the hidden machinery underneath has been activated, our consciousness will never know it" (Damasio 1994, p. 185). He continues: "…the corresponding neural pattern can be made conscious and constitute a feeling. However, although many important choices involve feeling, a good number of our daily decisions apparently proceed without feelings" (ibid). Since somatic markers can be functional in the absence of conscious deliberation, SMH can be understood, in connection with the hypothesis of procedural reflexivity, as an essential component of metacognition as it was described earlier.

The claim that it is possible for somatic markers to be functional in the absence of conscious deliberation is in line with Dworkin's theory and CTA, which both (rather briefly) acknowledge that reasoning need not be a conscious process. Humans are equipped with a cognitive system that assists the individual in the evaluation of her beliefs and desires. Such assistance need not be accessible to consciousness.

According to Damasio, "Somatic markers probably increase the accuracy and efficiency of the decision process. Their absence reduces them" (Damasio 1994, p.173). The argument for this claim seems to be that, without somatic markers, reasoning and decision-making goes astray. Damasio writes, "When a negative somatic marker is juxtaposed to a particular future outcome the combination functions as an alarm bell. When a positive somatic marker is juxtaposed instead, it becomes a beacon of incentive" (ibid).

As I understand it, SMH is an attempt to clarify the sense in which emotions guide and facilitate reasoning and decision-making processes, together with an account, perhaps, of the way emotions might make decision-making more accurate. According to Damasio, SMH also facilitates decision-making processes in social situations and with respect to other people. This line of thought is interesting, especially in connection with the role of other-attributiveness (which, I have argued, is an essential component of metacognition). In 4.3 it was argued that other-attributiveness is needed for control, and hence autonomy. SMH can be understood in the light of this claim. Damasio writes that SMH is

> …compatible with the notion that effective personal and social behaviour requires individuals to form adequate "theories" of their own minds and of the minds of others. On the basis of those theories we can predict what theories others are forming about our own mind. The detail of accuracy of such predictions is, of course, essential as we approach a critical decision in a social situation. (Damasio 1994, p. 174)

As one can see, SMH concerns reasoning about one's own mental states, as well as the capacity to understand the mental states of other people in order to make predictions and decisions in social situations. Plausibly, such a metacognitive capacity is relevant to planning and cooperation. Moreover, and significantly, the capacity presented in the above passage involves, according to Damasio, emotional mechanisms.

Like the advocates of CTA, Damasio uses the concept of preference system, but in a biological sense. "Somatic markers are thus acquired by experience, under the control of an internal preference system and under the influence of an external set of circumstances which include not only entities and events with which the organism must interact, but also social conventions and ethical rules" (Damasio 1994, p. 179). Moreover, "The neural basis for the internal preference system consists of mostly innate regulatory dispositions, posed to ensure

survival of the organism" (Damasio 1994, ibid). It is important to note that somatic markers can be both innate and acquired.[44]

From an empirical point of view, I think SMH has something important to tell us about the functional role of emotion in metacognition. It can also help us to see why reasoning need not take place at the level of conscious awareness.

SMH also powerfully explains why a pure reasoning view, like the strong independence view presented in 4.4.2, is incorrect. With respect to reasoning capacity, a strong view would require pure rationality, which, in turn, requires emotional states, like desires, wishes and intentions, to be excluded. Nevertheless, empirical data about emotions and their role in metacognition might help explain why a strong view of independence is implausible. As Charland points out, practical reasoning would be blind without emotion (Charland 1998, p. 78).[45]

Somatic markers prevent superfluous reasoning. In a situation where a decision is reached, they help to sort the relevant options from irrelevant alternatives. The idea that a mechanism of some kind is required to sort out irrelevant options has an important bearing on metacognition. As was argued in Chapter 2, such a mechanism prevents the infinite regress from threatening.

Inability to select relevant options will undermine autonomy. In fact, in the global perspective it is hard to describe the sense in which an individual with this inability really directs and controls herself. Damasio's clinical data emphasize the difficulty of valuing options and comparing them with each other in order to make a decision. When the brain regions crucial for reasoning and emotional processing are impaired, evaluation goes awry.

According to Damasio, functional somatic markers require a non-defective brain and a normal culture. We have already seen that frontal lobe damage impairs the functioning of somatic markers. Sociopathic behaviour also falls into the general category of cases in which the brain is defective and emotional mechanisms become impaired. In such cases, somatic markers do not function accurately, and this leads to self-destructive behaviours (Damasio 1994, p. 178).

---

[44] Damasio distinguishes between primary and secondary emotions. Primary emotions are innate while secondary emotions are acquired from environmental learning processes like socialization and encounters with conventions.

[45] See also Finucane et al. (2000) and Slovic et al. (2007).

As was claimed earlier, external circumstances can prevent an individual from exercising her autonomy. Likewise, external circumstances can erode the functionality of somatic markers. As the cognitive system might be intact, it might not work well in a defective culture. Damasio presents some examples:

> The effect of a "sick culture" on a normal adult system of reasoning seems to be less dramatic than the effect of a focal of brain damage in that same normal adult system. Yet there are counter-examples. In Germany and the Soviet Union during the 1930s and 1940s, in China during the Cultural Revolution, and in Cambodia during the Pol Pot regime, to mention only the most obvious such cases, a sick culture prevailed upon a presumably normal machinery of reason, with disastrous consequences. (Damasio 1994, pp. 178-179)

Damasio's "normal culture" requirement is plausible. However, although it is quite easy to imagine defective cultures, it is nevertheless hard to decide, in a precise way, what a normal culture would be.

In conclusion, the claim that emotional mechanisms facilitate reasoning and decision-making, overtly as well as covertly, in personal as well as social situations is relevant to the analysis of autonomy in terms of metacognition. Importantly, the covert functioning of somatic markers illustrates the function of procedural reflexivity in metacognition. SMH emphasizes the role of emotional mechanisms, and their functional role in reasoning and decision-making. As Damasio contends, somatic markers probably increase accuracy in cognition. In addition, empirical data on prefrontal lobe damage illuminate the intimate link between evaluation and emotion. The above discussion also explains the sense in which emotions fulfil natural functions, and why they do not necessarily undermine autonomy – at least, in the personal and social domain.

Let us now take a closer look at Charland's account of the functional role of emotion in *appreciation* – a capacity sometimes said to be one of the prerequisites of mental competence. It will be argued that appreciation, as an evaluative skill that takes place in metacognition, involves emotional mechanisms. This is why Charland's theory is relevant here.

4.4.5. Appreciation and Appraisal

From a legal point of view, the concepts of appreciation and appraisal are normally regarded as components of mental competence; and it has been maintained that mental competence relies partly on the individual's ability to evaluate various courses of action with regard for her own values (Buchanan & Brock 1989). Charland (1999, 2007) has developed an account

of the way in which the role of emotion in mental competence is to be understood. In his theory, the focus is on the concept of appreciation, which is claimed to involve emotional components.

Appreciation requires the capacity to have insight into one's own condition. Consider, for instance, insight into illness. An appreciation of one's own illness involves being able to acknowledge it, being aware of one's own situation, and being able to see in what sense the illness affects one's own life. With respect to mental competence, appreciation, as a cognitive concept, cannot exclude emotional mechanisms – at any rate, on Charland's approach. But appreciation is often understood from a purely cognitive perspective, a perspective in which emotional influences are neglected. For instance, in the Macarthur test, which is a standard commonly used to determine cognitive capacity, it never appears that emotions, in a functional sense, are involved in appreciation.[46] If emotions are considered, they are usually regarded as disturbing elements in reasoning (Charland 1999, p. 363). According to Charland, this old-fashioned understanding of emotions as irrational is to be rejected. Rather, emotions are important components in cognition. Charland writes:

> Whatever the case, a great number of emotion theorists today maintain some form of the thesis that emotion involves appraisal and that appraisal requires "cognitive" capacities of some sort. All of this invites the question why the cognitive capacities that underlie emotion should be unilaterally excluded from the cognitive capacities that underlie competence. (Charland 1999, p. 364)

According to Charland, the concept of appraisal plays an important role in appreciation. In other words, to appreciate involves appraisal. According to psychological theories of appraisal, appraisal is a process in which the individual manages the events or situations that confront her with the aim of identifying what is happening. Appraisal involves the experience, evaluation and interpretation of internal as well as external input; and as a result it allows the individual to relate to what is happening (Lazarus & Lazarus 1994).

As I understand Charland, theories of emotion explain, in effect, why cognitive theories of mental competence must be modified. Without appraisal, it would be hard to say in what sense an individual's appreciation of, say, her illness, really applies to herself. Charland states

---

[46] The Macarthur study distinguishes between four aspects of competency: understanding, appreciation, the ability to manipulate information rationally, and being able to communicate a choice (Appelbaum & Grisso 1995). For a discussion of the Macarthur study, see Charland (1998); Breden and Vollmann (2004); Tan et al. (2007).

that in order to appreciate, the individual "…must have the capacity to attribute personal significance to events and actions" (Charland 1999, p. 368). This capacity, as I understand it, involves emotional mechanisms. Charland further writes: "Without emotion, there are important respects in which individuals cannot evaluate what the choice *means* for them" (Charland 1999, p. 370).

The capacity for appraisal appears to involve minimal self-knowledge of the kind discussed in 4.2. For instance, the individual must be able to integrate relevant information – both internal and external – with herself, and she must be able to develop a certain attitude to it. The capacity for appraisal ensures the individual is able to manage the situation and take control over it. To return to themes in the above discussion, without functional emotions it would be difficult to adequately accept, prefer, reject, or revise one's first-order mental states. In Chapter 2, it was argued that these metacognitive skills are essential for autonomy. It is reasonable to conjecture that the employment of these skills will sometimes involve emotional mechanisms understood in terms of appraisal.

Let me summarize the above discussion. Emotional mechanisms perform important functions in connection with metacognition and mental competence. Theories of mental competence need to be modified, since appreciation, as one of its components, involves appraisal. Without emotion, cognition does not function advantageously. Empirical data support this claim (see 4.4.3).

The discussion above also brings out, I hope, the role of emotion in metacognition in autonomy. From a global perspective, metacognition, including emotion, helps us to manage and control our personal and social lives in a meaningful and valuable manner. As Damasio claims: "Whenever I call a decision advantageous, I refer to basic personal and social outcomes such as survival of the individual and its kin, the securing of shelter, the maintenance of physical and mental health, employment and financial solvency, and good standing in the social group" (Damasio 1994, p. 179). However, what Damasio's clinical research about prefrontal damage show is an inability to obtain these kinds of advantage. Consequently, in my view, emotional and prefrontal lobe damage is an obstacle to individual autonomy. Let us now return to the weak form independence that was briefly introduced in 4.4.1. Unlike the strong view, the weak view of independence accommodates the broad perspective on cognition discussed so far.

4.4.6. Weak Independence

Whether an individual is autonomous is to some extent determined by internal as well as external factors (Lindley 1986, p. 50). Autonomy is often claimed, directly or indirectly, to require the absence of controlling interference and intact reasoning capacities (Dworkin 1988; Lehrer 1999b; Beauchamp & Childress 2001; Taylor 2005). The first requirement here is often illustrated by manipulation, while the latter is illustrated by cases of neurological deficit.

In 4.4.2, it was argued that an understanding of autonomy couched in terms of independence in a strict and atomistic sense is untenable. To clarify, analyses of autonomy deploying the notion of strong independence are incorrect. Rather a weak form of independence allowing for emotional influence in reasoning, and allowing, to some extent, also, for external influences, is plausible.

External influences need not undermine the exercise of autonomy. It is when external influences determine an individual's behaviour that autonomy is violated (May 2005, p. 307). (Consider manipulation here.) However, some kinds of external influence, as well as internal, have to be allowed. While atomistic theories ignore, or downplay, the social context, a weak view of independence does not. If we adopt the weak view, some misunderstandings of the concept of autonomy, characterized in terms of independence, can be successfully jettisoned.

It is sometimes objected that theories of autonomy focus too strongly on the self-sufficient and independent individual (see Guinn 2002; May 2005). Instead, the concept of autonomy, it is argued, should be understood as relational and social.[47] External factors tend to be neglected in the autonomy debate (Christman 2003).

According to May (2005), autonomy presupposes a social setting, or context, with which the individual interacts. Autonomy cannot be understood solely in terms of the individual in isolation from her social relations. May writes:

> We are not self-sufficient, but this does not mean that we do not "rule" our own lives. In this way, the notion of autonomy can be developed as a practical notion for individuals living within the structure of a political and social system. (May 2005, p. 308)

As I understand May, autonomy is to some extent relational. External relations need not undermine the exercise of autonomy. For instance, loyalty to others, promising, relationships and judicial prohibitions do not seem to violate autonomy. At least, they do not as long as the individual accepts them through procedural independence (Dworkin 1988).

---

[47] For a recent discussion of the role of relational aspects of autonomy, see Oshana (2006).

Dworkin's requirement of procedural independence allows for a weaker form of independence given which the individual is able to accept, reject or revise (say) advice from other people or other kinds of external influence. Through procedural independence, these kinds of influence become the agent's own.

In addition, Agich argues against a strong view of independence. He writes:

> The liberal ideal of the autonomous agent as an independent rational decision maker must give way to seeing autonomy as involving such relationships with others. Some of these relations can thwart or even destroy the conditions of autonomy, paternalistic actions such as coercion, for example, but others support and enhance autonomy. (Agich 2004, p. 297)

Clearly, individuals are not independent of external factors in a strict sense. They are to some extent tied to social relations and cultural patterns (Christman 2001, 2004; Oshana 2006). Plausibly, weak independence of some kind simply has to be assumed in order even to begin to speak about autonomy, and its exercise, in a meaningful way.

To varying degrees, we are dependent on other people: we relate to, and understand, ourselves through them, while at the same time possessing procedural independence. These two claims are not contradictory. The autonomous individual can be dependent and at the same time retain a weaker form of independence. Let me flesh out this idea with the following example.

> Suppose that an individual X is uncertain to whether she should take the new job or not. Being uncertain, she calls her friend Y to get some advice, in order to facilitate her decision. When she calls Y, Y advises X that she should not take the job: it will require too much time travelling and probably threaten her family relations. X turns down the job.

This situation does not undermine X's autonomy, because X has the capacity to rationally evaluate Y's advice in relation to her own preferences and feelings. To be sure, if X acts upon Y's advice without evaluating what Y says, X will indeed have been influenced in a way that undermines the claim that she has exercised autonomy. However, on the assumption that X is able to reason about the advice from Y, X will be able, through procedural independence, to evaluate the advice she has received and make it her own. X was able, then, to make the decision on her own, even if it was Y who initially advised X to decline.

We often get advice from people with whom we come into contact. Advice from others may help us to clarify our decisions and plans. If external influences always undermined the exercise of autonomy, would marriage, promise-making or advice operate as autonomy-violating factors? Of course, it might be difficult to separate contributing and undermining

factors. There are borderline cases like persuasion. Persuasion seems to me to lie somewhere between advice and manipulation. It can undoubtedly be hard to determine whether or not an external influence violates the exercise of autonomy.

No man is an island. An individual exercises her autonomy in an external setting. To interpret the concept of autonomy as independence in the strict sense is to risk introducing confusion to the discussion of the concept of autonomy. If autonomy is to be understood in terms of independence at all, a weak interpretation of it is more plausible. Autonomy as a metacognitive capacity must be understood relationally. It is necessary to take into account the role of external factors, and to acknowledge that these factors can undermine, but also maintain, or facilitate, the exercise of autonomy. This idea is connected with the third component of metacognition argued for earlier. Recall the claim of Agich's presented in 4.3.

> …the world of experience does not occur in an "objective" space, but is essentially defined intersubjectively. This means that autonomy involves an essential connection with others. (Agich 2004, p. 297)

Autonomy is the metacognitive capacity to exercise control. This capacity is influenced by both internal and external factors. Without the metacognitive capacity to control internal and external inputs, autonomy can be undermined. Certainly, its exercise can be blocked by external factors, which in certain cases seems normal. Sometimes one has to adjust to external factors that are not possible to influence. Consider for instance governmental decisions. The remaining sections of Part 1 discuss internal and external factors that might undermine autonomy, or its exercise. Special attention will be paid to metacognitive impairment.

Chapter 5

Undermined Autonomy

5.0. Metacognitive Impairment

This chapter will consider the phenomenon of autonomy being undermined by impaired metacognition. It will also consider cases in which the exercise of autonomy is undermined by external factors. Through the metacognitive skills that have been discussed in this thesis, the individual is able to direct herself, and to control her internal as well as her external input. To be able to control internal and external inputs, the individual must not be suffering from metacognitive impairments. Nevertheless, metacognition must be understood with respect to the external setting and in particular external influences. For instance, to adapt to novel situations, to understand and respond to new information, and to interact with other people, requires that metacognition is to some extent intact.

External influences can undermine the exercise of autonomy: that is to say, empirical circumstances can eliminate the opportunity to exercise one's metacognitive capacity. This can happen irrespective of one's possession of an intact metacognitive capacity. Thus while metacognitive impairment undermines autonomy, external influences can also undermine the exercise of autonomy, and they can do so even in cases where metacognition is functional.[48]

It is often argued that a reduced, or impaired, reasoning capacity is an obstacle to autonomy (see Lindley 1986; Dworkin 1988; Edwards 1997; Beauchamp & Childress 2001); and certainly it is not difficult to imagine cases of undermined autonomy that involve some kind of metacognitive impairment. Imagine an individual who does not reflect at all. The brain-dead and chronically comatose represent uncontroversial cases of undermined autonomy that are easy to determine as such because of the metacognitive impairment they involve. However, other cases might be more difficult to determine. These cases are problematic because, in them, we find it hard to determine whether or not the individual is autonomous.[49] The present analysis may ease this difficulty.

Below, examples of undermined autonomy involving metacognitive impairment will be presented and discussed. In the examples, empirical data from neuroscience on metacognition,

---

[48] I will have more to say about this issue in 5.1 and 5.5.

[49] I think clinicians in the healthcare sector are often confronted with this kind of problem.

its basic underpinnings, and brain damage, will be used. Undermined autonomy caused by metacognitive impairment will be understood in a global sense: in other words, autonomy becomes undermined when the impairments generate too many (as it were) bugs for the individual to control herself in daily life. For instance, in impaired cognition, inaccurate evaluation of one's mental states might deceive one, in effect removing one's control. If inaccurate evaluation happens frequently, autonomy is obliterated.

Let be present an important note here. The use of empirical data to support philosophical theory may be problematic. First, in philosophy, unlike the empirical sciences, conceptual analysis is common. In empirical science, concepts are often operationalized for specific purposes set by the study at hand. Nevertheless, if we consider operationalized concepts alone we run the risk of neglecting important aspects of a concept and hence of failing to capture an adequate understanding of it. Another problem is that theoretical development and change happen more frequently in empirical sciences than in philosophy (Brinck 2005). Hence, empirical data that are used to illustrate and support a philosophical theory might become false and be discarded. At least, empirical theories, as opposed to philosophical theories, are more sensitive to the need for revision (which might happen faster than expected because of new, competing data).

These differences between philosophy and empirical sciences and the problems to which they give rise cannot be ignored, but a multidisciplinary work about the concept of autonomy remains fruitful for various reasons. First, a theoretical analysis of autonomy, metacognition and control has practical value. This kind of analysis is fruitfully applied to empirical sciences because it gives a more precise and detailed interpretation of what autonomy means. Philosophers deploy valuable theoretical knowledge and analytic tools in order to clarify concepts. In this thesis, theory is directed towards practice, i.e. theoretical considerations work as valuable guidelines in practice. Reciprocally, clinicians working in the relevant fields have valuable experience to share vis-à-vis the more practical dimensions of autonomy, metacognition, its impairment, and control. Their practical experience is productively considered in the course of theoretical discussion of the concept of autonomy. As I see it, mutual exchanges between philosophy and empirical sciences contribute valuable knowledge in both directions.

Before discussing metacognitive impairment, let me clarify the important distinction between *being autonomous* and *exercising autonomy*.

5.1. Being Autonomous and Exercising Autonomy

Being autonomous is not the same thing as exercising autonomy. This distinction is important. It reveals, among other things, that an individual who suffers from physical (bodily) impairment need not suffer from undermined autonomy. Irrespective of her physical condition, she can still exercise her autonomy so long as she possesses an intact metacognitive capacity. Moreover, and as we shall see, the distinction accounts for the fact that an individual who is autonomous (i.e. has metacognitive capacity) may still fail to exercise it in certain empirical circumstances.

The distinction between being autonomous and exercising autonomy has also been pointed out by Dworkin (1988) and Lindley (1986). According to Dworkin, "…autonomy is conceived of as a second-order capacity of persons" (Dworkin 1988, p. 20. He continues, "By exercising such a capacity, persons define their nature, give meaning and coherence to their lives, and take responsibility for the kind of person they are" (Dworkin 1988, ibid).

According to Lindley there is "…a confusion between being autonomous and exercising autonomy" (Lindley 1986, p. 69). About this distinction, Lindley writes that

> …it is important for those who take seriously the value of promoting autonomy among people. In trying to produce a conception of autonomy, one is primarily interested in 'being autonomous', with the question 'what is it to be autonomous?' On the other hand, in discussing possible policies in regard to respect for people's autonomy, the question of the exercise of autonomy is central. (Lindley 1988, p. 69)

In this thesis the concept of autonomy refers to a metacognitive capacity of the individual. To say that an individual is autonomous is to say that this individual possesses metacognitive capacity. The exercise of autonomy, on the other hand, is the realization, or instantiation, or effective exercise, of the metacognitive capacity – for instance, to reach a terminal point in reasoning, to evaluate information, to plan, and to make decisions.

The relation between being autonomous and exercising autonomy is asymmetrical. The exercise of autonomy requires one to be autonomous. An individual who lacks autonomy cannot exercise it. As Lindley notes, "…if a person is not autonomous, the question does not even arise of his making or being denied the opportunity to make, autonomous choices" (Lindley 1986, p. 69). However, and as was argued earlier, metacognitive autonomy, understood in a global perspective, does not require the individual to constantly exercise her autonomy if she is to be ascribed autonomy. It seems implausible that an individual is, or becomes, autonomous solely when she is exercising it. If that were the case, autonomy would

be impossible, and we would arrive at the problematic, local, and contextual understanding that was argued against in 4.4. Normally, it is expected that individuals are autonomous over time, not solely in temporary time-slices.

As I see it, one can be potentially autonomous without exercising one's autonomy. Consider, for instance, a prisoner who is rational, reasons properly under no illusions, but is unable to act. Lindley (1986, p. 69) asks in what sense the prisoner is more autonomous than someone who is less rational but has the ability to move herself around in her environment. For instance, it is plausible to suppose that an autonomous patient can decide that others will take care of her plans, decisions, and so on. In some respects, she ends up exercising her autonomy. However, she is still autonomous: the fact that she ceases to exercise her autonomy does not mean that she has lost her metacognitive capacity.

In the next section I ask: to what extent does one's physical condition matter to autonomy if the latter is understood in terms of metacognition?

5.2. Physical and Mental Action

It would appear that the exercising of autonomy does not necessarily involve physical action. It is questionable whether physical action of the kind involving bodily movement is required if an individual is to exercise her autonomy. Importantly, it is the metacognitive capacity of the individual, as opposed to her physical condition, that is essential to autonomy (and hence the exercise of it).[50] To clarify, to exercise one's autonomy does not require one to be able to perform physical actions. As was claimed in 5.1, the exercising of autonomy is rather to be understood in terms of mental acts, and in particular in terms of the fact that one is able to use one's metacognitive capacity. An individual who is unable to perform physical actions need not suffer from undermined autonomy (on the assumption that metacognition is intact). On the other hand, an individual who is able to perform physical actions but is suffering from impaired metacognition is more likely to suffer from undermined autonomy, and hence an inability to exercise it.

Below I will present two cases that illustrate the above lines of thought. The first case considers autonomy in relation to impaired metacognition; the second case considers a dysfunctional body which is such that the individual is unable to perform physical action. Importantly, the first case has already been presented in 1.2. For present purposes, I will use

[50] A similar line of thought can be found in Hermerén (2006).

this case again, and with the same aim – i.e. that of illustrating undermined autonomy in terms of metacognitive impairment.

(i) Mary

> Suppose a patient, Mary, has spent time in a hospital following a stroke. She is now discharged from the hospital, but because of her stroke several metacognitive skills have become reduced. Mary is now confused; her memory capacity is impaired and she suffers from chronic fatigue. In addition, her ability to evaluate information, to plan and to localize in time and space, has become weakened, and so also her sense of identity. Nevertheless, Mary's physical condition, her bodily functioning, is good. As she once could, and without difficulty, Mary can walk and move around using her body, but when she leaves the hospital she is unable to find her way home.

Is Mary autonomous? If so, in what sense? Plausibly, she suffers from undermined autonomy because of metacognitive impairment caused by the stroke. However, she is physically intact. What can now be said about the role of bodily function in autonomy? Mary is able to perform physical actions, to move and act, but without intact metacognition. Her reduced metacognitive capacity makes it impossible for her to find the way home. She no longer seems to be capable of controlling herself in daily life. She is hardly autonomous from a global perspective. Since Mary is not autonomous, she cannot be said to exercise autonomy.

This first case illustrates the way that metacognition is more essential in order to be autonomous than one's physical condition. It also illustrates the fact that one's being autonomous, and one's exercisings of autonomy, might be undermined when one suffers from metacognitive impairment, even if one's physical condition is intact. Thus, the exercise of one's autonomy is not to be confused with the use of one's body in physical action. Consider now the second case.

(ii) Matt

> Imagine an individual, Matt, who is bound to his wheelchair. Matt suffers from paralysis from his neck to his toes, and hence he cannot move his own body. Actually, Matt possesses effective reasoning capacity, and he can verbally express his preferences to relatives and to the healthcare personnel. Apart from verbally expressing things, Matt is incapable of performing physical and bodily actions.

Is Matt autonomous? Plausibly he is. Matt is able to verbally express his preferences. He also possesses the capacity to reason properly, as do most normal adults. Matt possesses intact metacognition, which makes him able to conduct a proper line of reasoning and hence leaves

him able to utter, or make known, his reasonable preferences. This is why it is plausible to say that Matt is autonomous and exercises autonomy. Even if his physical condition is seriously impaired and leaves him incapable of using his body, Matt is autonomous, because his metacognitive capacity is intact. This second case is meant to illustrate the fact that metacognition is more essential to autonomy than one's physical condition. It also illustrates the idea that an individual can be autonomous, and indeed exercising autonomy, even though that individual's physical condition is seriously impaired.

But in one respect the second case is problematic. Matt can be said to exercise his autonomy, at least in a meaningful sense, solely with respect to the possibility of having his preferences realized through the support of relatives and healthcare personnel. Some kind of communicative technique – verbally or through blinks – must be available in order for Matt to express his preferences, and for other people in the setting to be able to respond to them. (Consider cases where the individual is unable to verbally express her preferences, or is unable to use eye-blinks. In such situations there must be present some kind of technical aid in order for communication to take place.) With respect to Matt, and in similar cases, one must take into account the external setting in which these individuals exercise their autonomy, and the further fact that there is someone, or something, in that setting that is responding.

The external setting is also relevant in cases like Mary's. For instance, in performing interventions on patients with neurological impairment, one must have in mind that the external setting might be more or less easy to handle given the patients' metacognitive functioning. I want to stress here the important fact that metacognitive capacity must always be understood in relation to the external setting because the two elements are invariably in interplay.

I conclude that physical action is not to be mixed up with the exercise of autonomy. Metacognitive capacity, as opposed to physical capability, is what is essential to autonomy. Mental action is more essential to autonomy than physical action. The two cases above illustrate the fact that it is more likely that metacognitive impairment undermines autonomy and its exercise than it is that physical impairment does the same. Let me now further discuss metacognitive impairment and undermined autonomy.

5.3. Autonomy and Metacognitive Impairment

Several examples of undermined autonomy can be found in the literature. I will discuss undermined autonomy in terms of impaired metacognition, especially with respect to self-

knowledge and lack of insight into illness. Below I examine three examples of undermined autonomy: inaccurate self-assessment in Anton's Syndrome, dementia, and thought insertion in schizophrenia. I begin with Anton's Syndrome.

5.3.1. Inaccurate Self-Assessment: Anton's Syndrome

Accurate self-assessment as a requirement of autonomy has been suggested by Anderson and Lux (2004a, 2004b). They claim that inaccurate self-assessment of one's own capacity deprives one of autonomy. This claim is supported by empirical data on deficient knowledge of one's own condition that can be observed in patients suffering from Anton's Syndrome.

According to Anderson and Lux, three aspects have to be taken into consideration if we are to understand what autonomy is:

> …any defensible set of requirements for autonomy must include (or entail) the capacity for accurate self-assessment and that focusing directly on self-assessment has the threefold advantage of being (1) more neutral vis-à-vis competing theories, (2) more plausibly tied to the active reflexivity constitutive of autonomy, and (3) more directly supported by evidence from clinical neuroscience. (Anderson & Lux 2004b, p. 285)

Anderson and Lux argue that inaccurate self-assessment can be understood in terms of impaired cognition and failure of executive control. Executive control is understood as a "…set of neurocognitive capacities that are directly engaged when one plans, initiates, and carries out a goal-directed activity over time with appropriate self-monitoring and self-correction as one proceeds" (Anderson & Lux 2004b, p. 285). According to Anderson and Lux, the capacity for self-assessment is reduced as a result of brain damage, especially in the frontal lobe regions. This is in line with other empirical data, which also indicate that the frontal lobe is the neural basis of executive functions (Gazzaniga, Ivry & Mangun 1998; Anderson & Lux 2004b).[51]

Patients suffering from Anton's Syndrome suffer from impairment in executive control. These patients, it is claimed, are blind, but they deny that they really are blind. Consider the case described in the passage below.[52]

---

[51] See also the discussion of executive function in 4.2 and the discussion of prefrontal damage and emotion presented in 4.4.3.

[52] See also 1.2.

> A severe frontal injury contused the anterior portions of John's brain and at the same time shattered both of his orbits, severing his optic nerves and leaving him with no light perception at all. The resulting behavioral syndrome was quite striking in that John not only insisted verbally that he still had vision, but he also initiated behavior as if he did, trying to move about his room in the manner of a person with normal vision. As a result, he walked into walls and furniture, collided with objects in his path rather than avoiding them, and repeatedly placed himself in positions that were extremely precarious for a person who could not see. Despite his ability to initiate action in an apparently self-directed way, John's persistently mistaken assessment of his visual capacity with respect to his actions made it impossible for him to act as he intended. In this sense, many of his actions could not count as autonomous, not because he could not see—plenty of blind individuals are perfectly autonomous—but rather because his impaired self-assessment left him unable to make sense of what he was doing. At least with respect to those actions, he was deeply alienated from himself as an agent. (Anderson & Lux 2004b, p. 280)

According to Anderson and Lux, Anton's Syndrome patients lack a feedback mechanism of the kind that would enable them to learn from mistakes and experience. The patients are unaware of this. With respect to this data, a feedback-mechanism is claimed to be fundamental: it enables us to resolve internal cognitive tensions and to execute tasks appropriately (Anderson & Lux 2004b, p. 284). Because of their syndromes, sufferers from Anton's lack insight into their own capacity and suffer from a reduced ability to integrate knowledge of themselves. They lack a self-guiding character which, according to Anderson and Lux, is crucial for autonomy.

Although their suggestion comes from an empirical perspective, Anderson and Lux relate their data to the philosophical debate about the concept of autonomy. As I understand it, they try to bridge the gap between philosophy and empirical disciplines as regards autonomy, executive function and control. This, I believe, is fruitful. Empirical data on normal executive function, as well as about impaired executive function and inaccurate self-assessment, might provide us with an improved understanding of autonomy as a metacognitive capacity. This is precisely what Anderson and Lux try to do.

> In particular, there seem to be mutually reinforcing intuitions about, on the one hand, a cluster of capacities that constitute an intuitively plausible and conceptually coherent account of autonomy and, on the other hand, a package of neurologic capacities that has been termed "executive function". For, if it is plausible that the widely used neurologic concept of "executive function" is broadly isomorphic with capacities associated with autonomy, then the conceptual claims made thus far will find corroboration in observations about impaired executive function. (Anderson & Lux 2004b, p. 285)

Anderson and Lux emphasize something important about what it means to have control. They claim that the reduced capacity for self-assessment will "…impair guidance control" (Anderson & Lux 2004b, p. 288). According to them, some kind of external requirement must

be established in order to determine whether or not an individual is autonomous. Accurate self-assessment is such a requirement.

Since the patients discussed in the article by Anderson and Lux might be able to give reasons, while being at the same time unaware of their deficit, the subjective interpretation of one's condition, or capacity, does not guarantee autonomy. I think the case of John supports the claim that the role of beliefs must, in an important sense, be incorporated in an account of the concept of autonomy. For instance, a belief cannot be based solely on the subjective interpretation arrived at by the individual herself. It needs, to some extent, to cohere with the facts.

In their article Anderson and Lux explicitly defend the idea that subjective experience is insufficient to determine whether or not an individual has the capacity for accurate self-assessment. They write: "The accuracy of one's self-assessment is a function of whether it corresponds to the facts, not a function of what it is subjectively reasonable for one to believe about one's capabilities" (Anderson & Lux 2004b, p. 281-282). They also state that, "The subject's own 'sense' of capacity is not self-verifying" (Anderson & Lux 2005, p. 312). Their discussion of Anton's Syndrome illuminates this position. When the metacognitive capacity to manage internal as well as external information becomes impaired, control is also violated.

Lack of accurate self-assessment of one's own condition emphasizes the role of self-knowledge, understanding, and the importance of being able to imagine the consequences of one's actions. Moreover, the suggestion that accurate self-assessment is one a condition of being autonomous supports the more general project of explicating autonomy in terms of a metacognitive capacity.

Nevertheless, is the notion that accurate self-assessment of one's capacity is essential to autonomy too demanding? As I understand it, the capacity for accurate self-assessment is a normal, everyday metacognitive skill. Accurate self-assessment does not necessarily require reasoning at the level of conscious awareness in order to be effective:

> …human action is routinely reflexive in the most mundane cases. One's self-assessment does not have to be explicit or conscious for it to be effective in the action-guiding role it plays. On the contrary, it typically operates in the background. (Anderson & Lux 2004b, p. 289)

This claim can be compared with the procedural reflexivity component of metacognition presented in Chapter 4. Recall the example given in 4.1 of assessing one's own capacity to jump over a ditch. In such cases, you habitually use your implicit, non-conceptual and

dynamic knowledge, i.e. knowledge of your possession of which you need not declare to yourself. As I understand the matter, assessment of one's own capacity normally works in just this way. This capacity is impaired in the Anton's Syndrome patients because they cannot learn from mistakes and earlier experiences. In this respect, the patients can be claimed to suffer from impaired metacognition understood in terms procedural reflexivity.

Chadwick (2004) has put forward the following objection to the accurate self-assessment proposal made by Anderson and Lux. As the science of genetic testing grows, individuals also have the opportunity to know more about themselves – for instance, their characteristics and inabilities. As a consequence of these opportunities, an individual who does not want to know certain available facts about herself will have compromised autonomy.[53] In other words, if an individual has the opportunity to learn things about his genetic makeup (e.g. his inability to have children) but does not want such information, it is questionable whether we should conclude that this individual lacks accurate self-assessment and hence autonomy. Chadwick asks:

> Suppose there is information "available", through genetic testing, that would affect one's chances of having children, or of having children without a particular genetic disorder and, knowing that such information is available, a couple decides not to avail themselves of this. This is clearly an inaccurate self-assessment, but is their autonomy compromised? (Chadwick 2004, p. 299)

At first glance, Chadwick's argument seems reasonable. But on closer inspection, we might want to ask: does this objection really overturn the suggestion that accurate self-assessment is a requirement for autonomy as that suggestion is made by Anderson and Lux?

I agree with Chadwick that the right not to know is not autonomy undermining. However, in one respect I think Chadwick's criticism is misdirected. The examples of inaccurate self-assessment discussed by Anderson and Lux concern cases where information is available, but where the patient is incapable of understanding it. Anderson and Lux claim: "For John, as for many other patients with frontal lobe injuries, his executive function deficits rendered him unable to integrate a knowledge of any of his impairments, including his blindness, into his behavioral output, verbal or otherwise, at any level" (Anderson & Lux 2004b, p. 280).

Suppose the man in the couple described in the Chadwick excerpt above is given the information about his low chances of having children but anyway behaves as if he were

---

[53] Chadwick also discusses whether knowing facts about oneself can decrease autonomy. This may be so, according to her, in the sense that knowing facts, for instance, about a weakness of oneself, can be debilitating.

capable of having children. In such a case, I would claim, he lacks accurate self-assessment, because he cannot integrate the information given with the behavioural output unless he has other reasonable causes for not changing his behaviour. Such behavior would be delusional, and hence autonomy undermining, if I understand Anderson and Lux right.

One might consider here whether or not it is correct to decline genetic information. It might be argued that one should not pay too much attention to genetic tests, because other factors, in several cases, are as important as genetic ones. Rather it might be more rational to live healthily, to do the best as one can, and not to care about the tests. For instance, the genetic tests might in fact encourage people to act blindly, and make them less inclined to deliberate about themselves, since they believe that their traits, dispositions, and so on, are fully predetermined.

What Anderson and Lux emphasize in their article is the notion that autonomy is a certain complex capacity that can perhaps be traced to executive function in the frontal lobe regions of the brain. One of its components, *qua* complex capacity, it is suggested, is accurate self-assessment. Anderson and Lux would not maintain that the suggested requirement for autonomy must be understood with respect to the amount of available information about oneself that one has. That would be a too strict requirement, since clearly we cannot know everything about our physical and psychological makeup.

The reply Anderson and Lux make to Chadwick is that accurate self-assessment concerns "…having the ability to incorporate available facts about oneself into one's decision making about and execution of an intended action. It is this ability to integrate self-knowledge with self-initiated action that is critical to what we have called accurate self-assessment, not the possession of the knowledge itself" (Anderson & Lux 2004a, p. 310). They continue: "Nothing we say, for example, commits us to requiring that an agent must know anything at all about her genetic makeup to be autonomous, anymore than it requires her to know, say, what her lung capacity is before undertaking high altitude mountaineering" (Anderson & Lux 2004a, p. 310). Rather "…autonomy is diminished if, through denial, frontal cognitive disorder, or some other mechanism, she is unable to use the relevant information about her lungs that is in her possession" (Anderson & Lux 2004a, p. 310).

If accurate self-assessment is to be understood in terms of the way in which the individual manages information about herself, we cannot require all facts, including those facts not known by the individual, to be present in order for self-assessment to be accurate. Surely there are facts about me of which I am not aware, but this does not violate my autonomy.

Let me now point to a weakness of the suggestion put forward by Anderson and Lux. It was argued earlier that it is important to make explicit whether it is the action or the individual that is concerned in an analysis of autonomy. This is sometimes hard to determine. A problem with the suggestion Anderson and Lux make is that their proposal suffers from ambiguity. One might wonder whether their understanding of the concept of autonomy concerns the capacity of the individual or, alternatively, whether their aim is to identify what is required for an action to be autonomous. Sometimes they seem to emphasize autonomous action rather than a capacity possessed by the individual. These two views are used interchangeably. This is problematic, since they run the risk of mixing individuals with actions. Because of this ambiguity, their proposal can be interpreted as both task-relative and local, but at the same time as global. However, if their proposal is interpreted in line with the latter view, it harmonizes with the present inquiry, which deals with the autonomous individual.

I end this discussion of self-assessment by concluding that the accurate self-assessment account improves our understanding of the concept of autonomy as a metacognitive capacity. Anderson and Lux describe typical examples of undermined autonomy, and their empirical findings are interesting in relation to the analysis of the concept of autonomy as a metacognitive capacity. Consider, for instance, the claim that control is intimately linked to the prefrontal lobe and executive function, since impairment in these regions reduces one's capacity to self-assess one's own capacity.

Finally, if the ambiguity presented above is resolved, and we focus on the individual rather than the action, the proposal made by Anderson and Lux appears to be line with a global perspective on the concept of autonomy as a metacognitive capacity. It is reasonable to hold that accurate self-assessment of one's own capacity is a metacognitive skill that is instantiated in several contexts.

Nevertheless, from a global perspective, there will be individuals who are victims of their own inability to arrive at accurate self-assessment. That inability will, in turn, lead them in disadvantageous directions rather too often, even if they themselves act on their self-assessments as if they were accurate. These individuals seem to be unable to estimate accurately the consequences of their actions. However, such a skill is required if an individual is to be autonomous. We now turn to the second example of undermined autonomy.

5.3.2. Dementia

Dementia is a neurological disease that should be considered in any serious discussion of deprived autonomy. Below I will briefly discuss autonomy in relation to Alzheimer's Disease (AD), as this disorder erodes cognitive function (Guinn 2002; Berghmans, Dickenson & Meulen 2004).

Patients suffering from severe AD also endure undermined autonomy. AD is a brain disease that is believed to afflict 5–10% of those over 65 years old (Passer & Smith 2008, p. 275). AD impairs cognitive and intellectual functioning, and it does so progressively because it is a degenerative disease. It reduces the individual sufferer's ability to manage her daily life. Understood from a global perspective, this is the major reason AD is discussed here.

Typical symptoms in AD are memory problems, reduced time perception, loss of identity, poor insight into illness, and planning difficulties. Further, a patient's experience of herself as the same individual over time may deteriorate as the disease develops. In view of these symptoms, it can be claimed that AD afflicts autonomy from a global perspective.

In many cases it is hard to determine whether individuals suffering from forms of dementia like AD can safely be ascribed autonomy. This problem arises because the disease develops gradually. The question is *when* in the course of events the patient can no longer be ascribed autonomy. Since AD is a chronic, but degenerative disease, and develops progressively, the deterioration in cognitive function develops progressively over time. Thus the individual may be sufficiently autonomous to participate in biomedical research about, say, treatment, at the outset of the research, while at a later stage she will lack the cognitive capacity required to participate at all.

At what point in the development of the disease is it plausible to claim that the patient is no longer sufficiently autonomous to participate in the research? As I see it, the autonomy of the patient has become compromised when she lacks a sense of herself as an individual, when she no longer knows who she is, or where she is, or cannot understand, nor integrate, relevant information to herself in a controlled way. This is tantamount to saying that an individual with AD is no longer autonomous when her global metacognitive functioning is seriously impaired. I admit that this leaves a serious practical problem, in that it may be very difficult to determine exactly when such impairment has occurred.

As was previously claimed, a common symptom of AD is poor insight into the illness, i.e. difficulty understanding one's own impairment and its consequences in daily life. However, it is important to note that this kind of poor insight is common in other types of neurological and psychiatric disorder as well (Anderson & Lux 2004b; Robertsson, Nordström & Wijk

2007). Anderson and Lux claim: "Many forms of psychopathology are associated with diminished insight into illness, and clinical observations related to that fact, are not uncommon in psychiatric patients" (Anderson & Lux 2004b, p 280). Recall, for instance, Anton's Syndrome, the previous example of undermined autonomy.

With respect to AD and poor insight, it has been suggested that multidisciplinary knowledge is needed in order to integrate findings from various research areas where the lack of insight into one's illness is an important issue (Robertsson, Nordström & Wijk 2007).[54] Like the proposal put forward by Anderson and Lux, such findings can increase our understanding of poor insight into illness and its relationship to the brain.

The functions of the frontal lobe must be considered if we want to obtain a better understanding of the underlying causes of AD. Findings indicate that poor insight into illness is connected with fronto-temporal dementia, i.e. a frontal-lobe syndrome that can strike the AD patient (Robertsson, Nordström & Wijk 2007). In general, frontal-lobe regions seem to play a crucial role in the neurological causes of AD – at least, so far as poor insight into one's own illness is concerned.

To conclude, I have briefly discussed dementia and AD and the sense in which metacognitive impairment afflicts the autonomy of these patients. I have also noted that it might be hard to determine whether or not a patient who suffers from dementia is autonomous. It has been claimed several times in this thesis that neurological research might provide an improved understanding of autonomy in terms of a metacognitive capacity. Neurological research might help us to answer questions about metacognitive functioning with respect to AD and poor insight into illness. Let me finally discuss the third example of undermined autonomy: thought insertion in schizophrenia.

5.3.3. Thought Insertion

Schizophrenia is normally regarded as a thought disorder that involves unusual experiences and beliefs about reality, but also about oneself and others.[55] In schizophrenia, psychosis and delusions, like paranoia and auditory hallucinations, are common symptoms. In paranoid schizophrenia, the patient may falsely think that she is followed by other people, or have

[54] This article focuses on neuropsychological, psychological and socio-psychological approaches to AD and poor insight.

[55] A detailed characterization of the symptoms of schizophrenia can be found in Cullberg (2003) and Ottosson (2004).

unrealistic thoughts of conspiracy which in turn afflict her daily life. In auditory hallucinations, the patient's inner voices may substantially influence her daily life in a way not desired by the individual herself.

Schizophrenia is normally regarded as a persistent mental disorder. Plausibly the schizophrenic symptoms strike the individual's autonomy in a global way. Data indicate that patients are seldom cured (The National Board of Health and Welfare 1991). However, according to the American Psychiatric Association (2000) 25% of patients with schizophrenia recover from the disorder, while 10% remain permanently impaired. It is further claimed that 65% show temporary periods of normal functioning.[56] Of course, even if schizophrenia is claimed to be a disease that is hard to cure, it should not be mixed up with recovery. As I see it, schizophrenic patients can be said to recover from their symptoms over circumscribed periods. Nevertheless, that does not mean that the patients are cured of the disease.

Schizophrenic symptoms need to be discussed in connection with the metacognitive account of autonomy. It is important to note that these symptoms afflict daily life because of impaired cognitive functioning (The National Board of Health and Welfare 2003b). In some respects they therefore also undermine autonomy. This is why schizophrenia will be emphasized here. Moreover, patients with schizophrenia seem to suffer from a disintegrated self.

Below I will discuss a certain type of delusion that has been labelled "thought insertion". In thought insertion, the patient experiences her own mental activity as alien (Campbell 2002; Proust 2006). Thought insertion, it will be argued, violates an individual's autonomy, because it means that the patient lacks experience of a unitary self.

Plausibly, an individual's autonomy will be undermined when she experiences her own mental activity as alien. This is exactly what patients suffering from thought insertion have reported. In thought insertion, the patient has direct access to her own thoughts, but she experiences them as not generated by her own mental activity. Thus the patient feels alien in relation to her own mental processes: she does not self-ascribe them.[57] Campbell writes:

> The content of the experience seems to be exactly that token thoughts are being generated by some other person, and perhaps, with malice, inserted into the mind of the patient, so that the

[56] We will return to this issue in Part II.

[57] Detailed discussion of thought insertion and schizophrenia can be found in Campbell (1999, 2002) and Coliva (2002). For those interested in schizophrenia and delusion as such, see Campbell (2001)

> patient has direct introspective knowledge of a token thought which was generated by someone else. (Campbell 2002, p. 36)

While Campbell (2002) does not discuss the concept of autonomy, the claim put forward in this excerpt is relevant to the issue of autonomy, understood in terms of metacognition, because it is about one's evaluation of one's own mental states. As I see it, in thought insertion the metacognitive capacity to evaluate one's mental states is impaired. This is in line with the claim that patients suffering from schizophrenia have difficulties monitoring their own mental states (Proust 1999, 2006, 2007).

Above, in 4.2, minimal self-knowledge was discussed. It was argued that an individual who possess minimal self-knowledge is immune to error through misidentification, because she cannot be mistaken about her own mental activity.[58] If she makes a misidentification, she seems to be out of control. However, patients suffering from thought insertion do not seem to be immune to error through misidentification, since they take their own mental activity to be someone else's (Campbell 2002). They report that other people inside their own minds force them to act. The individual might, for instance, be aware of her alien and unwanted thoughts, but act on them nonetheless. In such cases, autonomy is undermined in a global sense: these cases are hardly instances of control. Thought insertion illustrates the phenomenon of the disintegrating self – a phenomenon, clearly, that makes it hard to control one's mental states accurately. What seems typically to be the case is that the patients lack the experience of a unitary self.

In connection with thought insertion, Campbell (2002) argues that introspective access to one's own thought is not sufficient to make one the owner of it. In order to be the owner of a thought one must also self-ascribe the generated thought as one's own. He writes:

> …the reason why the patient says that she is experiencing thoughts that are not her own is that she finds herself with introspective knowledge of thoughts of which she seems not to be the producer. The only way to explain why the patient makes this response is to say that there are two strands in the notion of the ownership of a thought: one relating to the possibility of introspection and the other relating to the person who produced the thought. (Campbell 2002, p. 38)

With respect to self-knowledge, the two strands required for ownership of thought are also needed for autonomy. It is not sufficient that one has introspective access to one's mental states. It is also required that the individual is able, through her metacognitive capacity, to

[58] For discussion of this topic, see Brinck (1997), Campbell (1999), and Coliva (2002).

ascribe them as her own. Here we should recall Dworkin's condition of procedural independence: "…procedural independence involves distinguishing those ways of influencing people's reflective and critical faculties which subvert them from those which promote and improve them" (Dworkin 1988, p. 18). Possession of procedural independence enables the individual to ascribe her own desires and beliefs as her own.

Another case that illustrates internal forces is that of so-called "madness crimes". These crimes have received considerable attention in Sweden over recent years. The causes of some of the crimes were understood in terms of the patient's psychiatric disorder, the idea being that the patients were "forced" by "inner voices" that led them to perform the crimes. These causes can plausibly be interpreted in terms of thought insertion; the patient acts in a way that is beyond her own control.

Now that I have presented a third example of undermined autonomy, I shall argue that external requirements must, to some extent, be established for autonomy: that is to say, autonomy cannot be determined on subjective grounds alone.

5.4. Autonomy and External Requirements

The discussion of metacognitive impairment confirms the need to establish external requirements in order to determine autonomy. Subjective requirements for autonomy are not sufficient, i.e. whether an individual is autonomous cannot be determined on subjective grounds alone. An individual who experiences herself as autonomous and controlled might, from an external point of view, and on reasonable grounds, not be regarded as autonomous. This is one of the reasons why external requirements for autonomy are needed.

It is not difficult to imagine cases where an individual takes herself to be autonomous, while most people would not concur. Consider, for instance, cases of delusions and dementia. To require solely that a subjective perception of the world and oneself is enough to establish autonomy would be disadvantageous. If there were no external requirements of autonomy, we would end up with an excessively permissive theory. A theory of autonomy lacking external requirements would not be fruitful.

The capacity for intersubjectivity might help us understand why we need to impose external requirements on autonomy. Intersubjectivity requires one to be able to form adequate beliefs about oneself, others, and the world. The notion that intersubjectivity plays an important role in autonomy, with respect to the way in which the individual interacts with the relevant external setting was argued for earlier.

Given that autonomy cannot be determined on subjective grounds, it will be important for the beliefs entertained by the individual to be at least largely anchored in further beliefs that are generally claimed to be reasonable. The individual must be able to control her beliefs and check that they are, to some extent, reasonable (e.g. ensure that her beliefs are mutually consistent).

As I see it, it is reasonable to insist that the capacity for intersubjectivity constitutes an external requirement of autonomy. Treated as an external requirement for autonomy, the capacity for intersubjectivity supports the second component of metarepresentation argued for in Chapter 4. It was there argued that autonomy must be based on the individual's capacity to represent other people's mental states and deal with them accurately.

I suggest that most of us are in general able to control accurately both the internal and external influences upon our everyday lives. (Of course, we sometimes fail to do that.) However, there seem to be individuals whose impaired metacognitive functioning frequently leads them in disadvantageous directions, even if, subjectively, they think that they possess control and think they evaluate their mental states accurately. However, when the mechanisms crucial for interpreting external as well as internal input become impaired, autonomy, and thus control, seems to be out of reach.

Next, I shall ask whether it might be possible to determine autonomy on the basis of the two components of metacognition argued for in Chapter 4. The suggestions presented below are tentative and in need of further development. However, they will, I hope, contribute to the project of developing practical guidelines that enable us to deal with autonomy, understood as a metacognitive capacity.

As was argued above, metacognition has two components: procedural reflexivity and metarepresentation. Metarepresentation can, in turn, be divided into inferential reflexivity and other-attributiveness. Let me begin with inferential reflexivity.

Whether or not an individual possesses inferential reflexivity might be determined by observing her reasoning capacity and evaluative skills. Consider, for instance, the reduced reasoning capacity of the frontal lobe patients described by Damasio, or delusions, like thought insertion, of the kind discussed by Campbell.

The second component of metarepresentation, other-attributiveness, might plausibly be understood in relation to the way in which the individual deals with her external environment, i.e. how she relates to it. This second component can perhaps be viewed in relation to the capacity for intersubjectivity and the ability to represent other people's mental states. Consider, for instance, sociopathic personality disorders.

Because inferential reflexivity and other-attributiveness are metarepresentational, it is possible to determine the status of these skills to some extent via phenomenological or verbal reports. In order to decide the status of inferential reflexivity and other-attributiveness, the individual's verbal reports on those of her reflective capacities that are exercised at the level of conscious awareness can be observed.

Procedural reflexivity, by contrast, which takes place on the lower level, might be determined by observing behavioural outputs, or responses. Consider, for instance, Anton's Syndrome. Patients suffering from Anton's Syndrome seem to lack procedural reflexivity. According to Anderson and Lux, they lack a feedback mechanism, because they cannot learn from their mistakes. This is why they lack a self-guiding character and have serious difficulty resolving internal cognitive tensions in the way that would be required if they were to execute a task appropriately.

Moreover, empirical data indicate that orbito-frontal patients experience regret, but do not adjust their strategies in the light of their experience of regret (Camille et al. 2004). Like individuals with Anton's Syndrome, these patients also exhibit difficulties in learning from experience. According to Shimamura (2000), frontal lobe patients seem to lack the capacity to inhibit task-irrelevant information. Data also indicate that orbito-frontal lesions lead to the disinhibition of emotional responses and, in turn, inappropriate social behaviour (Damasio 1994). Such lesions might result in emotional outbursts and risky decision-making behaviour (Shimamura 2000). As I see it, these kinds of behavioural output, generated by metacognitive impairment, might function as valuable guides to an individual's autonomy with respect to procedural reflexivity.

Two kinds of technique for determining autonomy have just been suggested: phenomenological/verbal reports and the observation of behavioural outputs. With respect to the former, I want to emphasize the role of communicative skills as output responses. These must be present if we wish to determine autonomy. What I want to stress, though, is that in order to determine autonomy, we must be dealing with an individual who, to some extent, exhibits abilities grounded in metacognition.

It will be hard to determine whether Matt, in the second case given above, is autonomous if there is now way for him to communicate his preferences. Suppose he lacked the ability to speak. What could we then say about Matt's autonomy? Probably not much, I think. If Matt cannot communicate his preferences, there is no way to know whether or not he possesses an intact metacognitive capacity. In normal cases, the primary tool of communication is verbal. It involves the individual's linguistic ability to express her preferences. If that is not possible,

however, as was claimed above in 5.2, some other kind of communication must be effective. Consider, for instance, the use of eye-blinks (in cases of severe physical handicap), or some kind of technical aid, like an eye-tracking machine. These could offer effective communication between, say, a patient and a physician.

I addition to the two kinds of technique discussed above, it might also be possible to determine autonomy, with respect to metacognitive functioning, through neuropsychological observation of the brain.[59] I will not develop this line of thought in detail here. As I see it, more knowledge about the brain, and about neurological impairment, is needed. Nevertheless, I believe we might learn more about normal metacognitive functioning by studying metacognitive impairment.[60] As Anderson and Lux note: "As is often the case, however, analyzing brain damage also reveals much about normal function" (Anderson & Lux 2004b, p. 291).

Moreover, the notion that metacognition should be linked to executive control and the frontal lobe, as discussed by Shimamura (2000), offers a promising alternative way of determining autonomy (see 4.2). From an empirical perspective, we might learn more about metacognitive autonomy by studying its relation to executive control in the frontal lobe regions of the brain.

However, and importantly, we do not know enough about the brain yet to be able to decide these issues merely by looking at the brain – except in very clear cases of brain injury. However, in cases other than those of significant brain injury we can determine in other ways as well. So neurophysiology is not an effective way of testing more subtle cases of possible metacognitive weakness.

Next, I will briefly consider the external circumstances that undermine the exercise of autonomy. Such circumstances can undermine the exercise of autonomy even if the individual possesses an intact metacognitive capacity.

---

[59] This issue is also discussed in Anderson and Lux (2004a).

[60] Similar views can be found in Agich and Mordini (1998) and in Agich (2004).

5.5. External Forces: Obstacles to Exercising Autonomy Irrespective of Intact Metacognition

In 5.3 examples of undermined autonomy arising from metacognitive impairment were discussed. Such impairment was described in terms of organic and neurological brain damage as well as in psychiatric terms. However, as has been noted earlier, an individual can be hindered in the exercise of her autonomy under certain empirical circumstances. Below I emphasize the role that external factors play in the exercise of autonomy.

While it has been suggested that autonomy might be determined by the pair of metacognitive components presented in this thesis, it is important to note also that the function of these components must be understood in relation to an external setting. If external factors were ignored, the resulting conception of autonomy would be too narrow and atomistic. As was argued in 4.4.6, a plausible account of autonomy must to some extent be understood as relational (e.g. by taking into account social ties to others). Further, the concept of autonomy as a metacognitive capacity must be understood in relation to empirical circumstances that might undermine the exercise of autonomy.

There are clear cases of empirical circumstances that undermine the exercise of autonomy. They do so irrespective of the individual's intact metacognitive capacity. For example, an individual might be unable to resist manipulation even if she does not suffer from metacognitive impairment. I think most humans, under certain circumstances, might become victims of the power of a manipulator. Nevertheless, it is plausible to claim that manipulation does not depend on metacognitive weaknesses of the sort described in this thesis. Threats, robbery, and coercion provide other examples of empirical circumstances that (temporally) undermine the exercise of autonomy.

While there are clear cases in which the exercise of autonomy is undermined, it is nevertheless hard to explicate, or characterize, in a precise manner, the empirical condition involved. This difficulty stems from the abundance of empirical circumstances that might strike the individual and compromise her exercise of autonomy. Since so many kinds of empirical circumstance undermine the exercise of autonomy, it is hard to capture and describe them in relation to one category. However, it is important to keep in mind the fact that intact metacognitive capacity does not guarantee the exercise autonomy.

Although external factors might undermine the exercise of autonomy, some of them do not, as has already been argued. Yet this distinction is problematic. Consider, for instance, the vague distinction between advice and persuasion. Advice is something we would not take to undermine the exercise of autonomy, but that attitude becomes questionable in cases of

persuasion. The next and final section of Part I integrates the metacognitive and relational understandings of the concept of autonomy.

5.6. Concluding Discussion: A Theory of Autonomy in Two Dimensions

An adequate theory of autonomy needs a two-dimensional perspective. However, common interpretations of the concept of autonomy tend to focus on the metacognitive/psychological approach, which emphasizes the reflective capacity of the individual, while downplaying the relational approach, which emphasizes social ties, the role of other people, and so on. While the relational approach understands autonomy in relation to external factors that operate outside the individual, the metacognitive/psychological put focus on internal factors that operate inside the individual. Seldom are both simultaneously considered in detail, but when they are they tend to become mixed up and distort the discussion. Normally, the focus is chiefly on one of the approaches, with this approach being rather clearly placed in front of the other. As I see it, theories of autonomy have not explained explicitly enough both the negative and positive influence of internal as well as external factors on metacognition.

Does the above analysis of the concept of autonomy in terms of metacognition force one to defend a metacognitive/psychological approach that neglects social/relational factors? The answer is no, because the analysis takes into account both approaches. A defence of a metacognitive approach that neglects the relational dimension would generate an excessively narrow and one-sided understanding of the concept of autonomy. For instance, the metacognitive skills of the individual require a setting in which to operate.

Murphy wonders: "Perhaps autonomy requires a subject to have both normal psychological capacities and the right sort of environment in which to exercise them" (Murphy 2004, p. 304). Interaction with the external setting takes place daily. Consider social relationships, cultural norms, laws and conventions. It would be unreasonable to claim that autonomy and its exercise are not to be understood with respect to external factors. The case of Matt discussed above illustrates the implausibility of neglecting external factors. In some cases other people, or some kind of technical aid that can provide support, must be available. Otherwise, Matt cannot be said – at least, in a meaningful sense – to exercise his autonomy. Thus, external factors play a fundamental role in Matt's exercising of his autonomy. In short, if no external support is present in the setting Matt inhabits, he plausibly retains the capacity for autonomy. What he cannot do is exercise it.

Below I summarize the conclusions drawn from the above analysis.

- As was initially argued, autonomy is an important but ambiguous concept that has to be explicated. Lack of a precise understanding of the concept of autonomy makes it hard to explain why autonomy is, in several respects, a goal. However, the analysis above offers a plausible understanding of the concept of autonomy, since it picks out a certain property of the individual.

- Autonomy is a metacognitive capacity of the individual that is understood from a global perspective. Its essential function is control. Metacognition has two components: procedural reflexivity and metarepresentation. In metacognitive terms, it is implausible to give an account of control without reference to these two components. (This was argued in Chapter 4.)

- An analysis of the concept of autonomy as a metacognitive capacity comprising both procedural reflexivity and metarepresentation hopefully makes it possible to determine whether or not an individual is autonomous. For instance, decisions about healthcare interventions should take into account the metacognitive capacity of the individual as well as the empirical circumstances with which her metacognitive capacity is in interplay. The components of metacognition, which were argued to constitute autonomy, must in turn be observed and evaluated with respect to the setting the individual inhabits. (This was argued for in Chapter 4 and Chapter 5.)

- An analysis of the concept of autonomy as a metacognitive capacity which explicitly acknowledges the role of external factors involved in the exercise of this capacity eases, or mitigates, the tension between the metacognitive/psychological and the relational approaches. As was argued in 5.1, exercising autonomy is not the same as being autonomous. Importantly, because autonomy is a metacognitive capacity, empirical circumstances in the external setting can impede its exercise. Thus, the exercise of autonomy has to be understood in relation to empirical circumstances.

- By considering both the metacognitive and relational dimension of autonomy we end up in a plausible and fruitful theory that can be applied in various areas. It is a theory that takes into account both the metacognitive capacity of the individual and the environment in which she exercises it. This is the why an analysis of autonomy

> requires a two-dimensional perspective. A two-dimensional analysis takes properly into account what makes possible, but also what undermines, autonomy and its exercise. The two-dimensional perspective might facilitate interaction and communication between the disciplines in which autonomy is an important issue.

The conclusions drawn will now be applied in the area of Swedish healthcare, especially in psychiatry – a field, as argued in the introduction, in which the concept of autonomy is both central and problematic. It is instructive to discuss the analysis put forward in Part I in psychiatry, since that will help to reveal whether it is possible to determine autonomy, and if so, in what sense.

The suggestion that the present analysis is relevant to psychiatry is hardly controversial. For instance, it has been claimed that schizophrenia is a control disorder. Neuropsychological data about metacognition in patients who suffer from schizophrenia show a reduced capacity to monitor one's own cognitive states, and a reduced capacity to adjust control in response to such monitoring (Proust 1999, 2006). In addition, the synergy between philosophy and psychology is valuable in that it promotes an improved understanding of metacognition (Nelson 1996, p, 115). Personally, I also want to emphasize the valuable synergy between philosophy and psychiatry as concerns metacognitive functioning and impairment. We now turn to Part II, which is more practically oriented in its focus on healthcare and, especially, psychiatry.

PART II

Autonomy in Healthcare

Chapter 6

Autonomy: Capacity, Right, or Duty?

6.0. Introduction

In this, Part II of the thesis, I propose to apply the analysis of the concept of autonomy developed in Part I to a number of issues in healthcare. As was observed in Chapter 1, the concept of autonomy is important in healthcare.

Part II of the thesis is divided into three chapters. Chapter 6 (the present chapter) deals with the claim that the concept of autonomy is important in healthcare. At the same time, however, the concept is both ambiguous and problematic. For instance, in healthcare, the distinction between the concept of autonomy as a metacognitive capacity and as a right to exercise one's autonomy is seldom explicated. The reason, it will be argued, is that in the context of healthcare the concept of autonomy is normally understood as correlative with a right to be respected, while important questions about the nature of the individual's autonomy tend to be neglected.

The present chapter also deals with the question whether it is possible to respect autonomy in cases where an individual can no longer be claimed to be autonomous. Special emphasis will be placed here on the so-called "substituted judgement standard". I end the chapter by asking whether or not the patient's right to autonomy sometimes tends to be over-emphasized in healthcare. Is it plausible to expect that patients' are autonomous? For instance, what will happen with vulnerable groups in society, like the elderly, or individuals who suffer from severe psychiatric disorders, if these individuals can be reasonably said to have undermined autonomy but are anyway expected to be autonomous? Are they also expected to be autonomous? As will be argued, the expectation (and accompanying demands) of autonomy might be negative for the patient if she in fact lacks the capacity to make her own decisions and take care of herself. If the right to exercise autonomy is defended in a too broad sense, this might place unrealistic demands on the vulnerable groups in society, or so it will be argued.

Chapter 7 discusses the concept of autonomy in relation to Swedish psychiatry. Special emphasis will be placed on deinstitutionalization and the participation, in society, of individuals suffering from persistent mental disorders. Some comments about the link between the concept of autonomy as a metacognitive capacity and coercive care are also

presented. Finally, in Chapter 8 I put forward some suggestions about how to further deal with the concept of autonomy in relation to psychiatric issues and general healthcare issues as well. So far as the concept of autonomy in psychiatry is concerned, an important suggestion here is that the decision makers who direct healthcare for patients suffering from persistent mental disorders must take into account the patient's metacognitive capacity *and* the social environment in which it is exercised. I conclude that healthcare efforts stressing individual autonomy must understand the concept of autonomy in the two-dimensional manner presented and argued for in the end of Part I.

6.1. The Patient's Reinforced Position

It is now time to consider the patient's reinforced position in healthcare. It is sometimes claimed that the right to exercise one's autonomy is a vital interest of society's (Lindley 1986, p. 106). Autonomy, it is said, is in many respects one of the most central and important concepts in healthcare (Le Granse, Kinébanian & Josephsson 2006; Anderson 2008); and "…the principle of patient autonomy is highly influential in modern healthcare" (Sjöstrand & Helgesson 2008, p. 113). In this sense, autonomy is a key concept in healthcare. If it were not, it would not be defended to the extent that it in fact is. Anderson states that while the concept of autonomy might be overvalued, it is a concept that has been given much attention in healthcare discussions, in the courtrooms, and among the agencies providing social services (Anderson 2008, pp. 7-8).

In order to maintain the individual's own capacity, potential, and integrity, common directions in Swedish healthcare emphasize the patient's right to participate in, and decide upon, her own care. The protection of such rights benefits the patient's opportunity to determine her own life herself (e.g. see SFS 1982:763; SFS 1993:387; SFS 2001:453; MFR 2002). This is the reason why the right to exercise one's autonomy has become so important in healthcare. Similarly, in the Act concerning Support and Service for Persons with Certain Functional Impairments (SFS 1993:387), which covers individuals suffering from mental impairments, it is stated that equal living conditions and full participation in society are to be promoted.

The directions presented above emphasize the right to have one's autonomy respected by medical staff. In several respects (coercive care excluded) this right is fundamental in healthcare practice. In the clinical setting, it is commonly presupposed that the patient has the

right to decide upon, and influence issues arising from, her own care. For example, she has the right to choose among the medical options being offered.

The interpretation of the concept autonomy as a right is often associated with the principle of autonomy. This principle is built on respect for the individual's right to make choices based on beliefs and preferences held by the individual herself. This is a common understanding of the principle (Beauchamp & Childress 2001, p. 63). In general, the principle of autonomy proclaims a right not to be treated in certain ways. For instance, the patient's right to exercise her autonomy is threatened in cases of manipulation because she is not able, in a controlled way, to evaluate her goals, desires or beliefs. The exercise of autonomy is restricted by the manipulative treatment she has undergone. As Beauchamp and Childress say: "In health care, the key form of manipulation is informational manipulation, a deliberate act of managing information that nonpersuasively alters a person's understanding of a situation and thereby motivates him or her to do what the agent of influence intends" (Beauchamp & Childress 2001, p 95). However, the deprivation of autonomy in this kind of circumstance can also be understood as a measure taken to protect her and to maintain what is left of her reasoning capacity.

However, it is important to keep in mind the thought that the principle of autonomy need not necessarily have priority over all other ethical values and principles (Beauchamp & Childress 2001, p. 57).[61] If the individual is endangered, or may cause harm to others, the principle is overridden by competing principles.[62]

The patient's role in healthcare, as in many other areas in society, can be described in terms of a customer who is active and makes her own choices with respect to the rights she has (Nordgren 2003). This is why the position of the patient becomes, to a greater extent than before, reinforced. As I see it, the patient's reinforced position in healthcare emphasizes the importance of being autonomous and having the right to exercise autonomy. Very often the patient is expected to participate in planning, and in decisions that concern her own care (Henrik Levinsson 2006). Patients are, in several ways, expected to be active, self-governed and information-seeking (Hansson 2006). These expectations influence current healthcare policies. However, in my view such expectations require the patient to have the metacognitive

[61] There are, according to Beauchamp, three more central principles of professional ethics: nonmaleficence, beneficence and justice (Beauchamp & Childress 2001, p. 12).

[62] Beauchamp and Childress (2001, p. 65) consider suicidal and drug-dependent individuals and infants.

capacity required for autonomy: in other words, the possession of such rights requires the patient to be autonomous.

The patient's reinforced position, and the defence of her right to exercise autonomy, is sometimes overused (or overvalued) in healthcare, I think; and this in turn might lead to negative consequences for the individual. The thought that that the right to exercise one's autonomy may have become overvalued can also be found in Anderson (2008). In my view, one might reasonably raise the question whether the right to exercise autonomy is defended in too broad a sense in some parts of the western world. One might ask whether "autonomy" has become a vogue word in some respects.[63] In 6.4, it will be argued that defences of the right to exercise one's autonomy are sometimes presented too (as it were) blindly, with insufficient attention to the nature of autonomy; and that this can lead to unfortunate consequences for the individual. This is especially clear where the individual is expected to possess autonomy, but in fact does not.

If we choose to work with the analysis put forward in this thesis, the right to exercise one's autonomy will be understood in terms of the realization of one's metacognitive capacity. That is, to respect an individual's autonomy is to respect the individual's right to exercise her metacognitive capacity. To respect autonomy is to respect a person's right to control herself. Plainly, but importantly, the right to have one's autonomy respected requires that one is autonomous. However, this claim is, as far as I can see, seldom sufficiently dealt with. It is rather expected, or assumed without question, that the individual is autonomous. This is probably one of the reasons why the concept of autonomy can be said to be problematic in healthcare.

In the next section the distinction between autonomy as a right and autonomy as a capacity will be explicated. I will argue that the concept of autonomy understood as a right not is to be conflated with the concept of autonomy as a metacognitive capacity.

6.2. The Capacity-Right Distinction

In Part I it was argued that, when it is examined in the correct perspective, the concept of autonomy is a property of the individual. This property, it was claimed, is a metacognitive capacity. This perspective concerns the nature of autonomy (as a capacity) and not the right to exercise it. The latter aspect is secondary. Unless you know the nature of autonomy, how are

[63] Autonomy is not always a positive value. In some non-western parts of the world it has a negative connotation.

you to determine what rights might be associated with it? The concept of autonomy as a right is applied in several contexts. As I see it, in healthcare the concept of autonomy is ambiguous. This is probably because those involved in the discussion seldom ask what kind of capacity autonomy is. The discussion has become dislocated from the primary question, about the capacity as such. It primarily concerns ethical issues, and because of this, little attention is given to the capacity itself.

The dislocation, I believe, depends on a common understanding of autonomy as a moral concept that is tied to values such as dignity and integrity. Surely, these values are intimately related. However, the emphasis on ethical issues has, I think, obscured the need to clarify what autonomy is. The capacity is, in a detectable manner, implicitly presupposed but not sufficiently defined.

To summarize, the notion that an individual is autonomous and the notion that this individual has the right to exercise autonomy are not semantically equivalent. They do not mean the same thing.[64] To be autonomous, it has been argued in this thesis, is to be in possession of a metacognitive capacity, while the right to exercise autonomy is connected with the principle of autonomy. This principle is seen as action-guiding. It is a tool with which to confront ethical issues in healthcare settings. It tells us to respect the patient's right to exercise her autonomy.

Since the notion that someone is autonomous refers to the metacognitive capacity of an individual, autonomy is not necessarily a moral concept. It follows that issues about the concept of autonomy need not be about rights or ethical considerations. It is very important to make this distinction, especially in healthcare discussions where the meaning of autonomy is far from clear. Autonomy is often regarded as a problematic and ambiguous concept precisely because this distinction is disregarded.

As I see it, ensuring that this distinction is explicit will make it easier to understand what autonomy means and to deal with it effectively in practice – for instance, in relation to various healthcare questions about how to meet the needs of patients. Let me now present a case that illustrates the distinction between the two (capacity and right ascribing) understandings of autonomy.

Imagine an autonomous patient who desires no longer to participate with clinicians and answer questions that concern her own care. A similar case, and one that is perhaps easier for the reader to grasp and for present purposes, would be that of an autonomous patient who no

---

[64] This claim is also made by Beauchamp and Childress (2001).

longer wants to exercise her autonomy. For instance, she now declares her wish that her relatives and the healthcare staff will make all decisions with respect to the medical services on offer. In addition, the patient wants these people, those nearest to her in her surroundings, to make the basic plans in her life. Most importantly, as the patient is autonomous (her metacognition being intact), she has adequate reasons for wanting to no longer use her right to exercise autonomy.

In this case it would be implausible to claim that the patient ceases to be autonomous when she waives her right to exercise her autonomy. That the patient no longer desires to exercise her autonomy does not mean that she is not autonomous *per se*. She retains the metacognitive capacity even if she does not want to continue to exercise it.

In one respect, as the example illustrates, the patient still is autonomous even if she has resigned her right to exercise her autonomy. The case shows that the right to exercise one's autonomy can be resigned. Nevertheless, that does not mean that an individual no longer is autonomous. As a rule, deliberatively giving up one's right to exercise autonomy requires considerable control.

The distinction between autonomy as a right and as a metacognitive capacity must be made explicit. Just as it is important to be explicit about whether it is the action, the desire, or the individual that is being analyzed as autonomous, one really must be explicit about the capacity-right distinction, i.e. one has to be explicit about which of these two understandings is being considered in the analysis one is offering.

I sympathize with Sandman (2004), who claims that the concept of autonomy is far from transparent. I also agree with Hermerén (2006), who points out that autonomy has many different meanings. The difficulty of knowing how to deal with questions that arise about autonomy might partly be the result of sloppy thinking. The present analysis of the concept of autonomy as a metacognitive capacity might clarify some of the problems that appear in healthcare discussions. The capacity-right distinction is an important and valuable clarification to have in mind.

Is it possible to respect autonomy in cases where an individual has been, but is no longer, autonomous? In the next section this question will be dealt with.

6.3. The Substituted Judgement Standard

It was previously noted that the principle of autonomy is normally associated with a right. Below I will argue that for this right to be respected it is required that the individual has, or has had, the capacity for control.

In normal cases, respect for the right to exercise one's autonomy concerns individuals who are autonomous in virtue of possessing metacognitive capacity. Nevertheless, there are exceptions to this view. An individual who has been, but no longer is, autonomous to some extent can have her autonomy respected anyway. This possibility is commonly understood (e.g. see Beauchamp & Childress 2001; Broström 2007) in terms of surrogate decision-making, or what is generally referred to as the "substituted judgement standard" (SJS).[65] This standard is characterized by Beauchamp and Childress as follows:

> Substituted judgement begins with the premise that decisions about treatment properly belong to the incompetent or nonautonomous patient by virtue of rights of autonomy and privacy. The patient has the right to decide but is incompetent to exercise it, and it would be unfair to deprive an incompetent patient of decision-making rights merely because he or she is no longer (or has ever been) autonomous. (Beauchamp & Childress 2001, p. 99)

One's autonomy is respected, according to SJS, when other people, like relatives or the medical staff caring for one, make a decision the non-autonomous patient *would have* made when she was autonomous. Respect for an individual's autonomy in this sense is based on the preferences expressed by the individual when (at an earlier time) she was autonomous.

Consider an individual who has now become seriously ill with dementia. She was until recently autonomous, but now she can no longer be claimed to be autonomous, because the metacognitive impairment caused by the degenerative disease from which she is suffering is too advanced. Since she no longer possesses autonomy, she is unable to exercise autonomy. Nonetheless, by applying SJS it is to some extent possible to respect her autonomy vis-à-vis her earlier interests. As Beauchamp and Childress claim, respect for autonomy in the way SJS requires is possible, because what the individual *would have* desired or believed *if* she were autonomous can be respected.

Cases of substituted judgement rely in part on a fiction. "An incompetent person cannot literally be said to have the right to make medical decisions if that right can only be exercised

---

[65] See also President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioural Research (1983, p. 132).

by other competent persons" (Beauchamp & Childress 2001, p. 99). As I understand it, in cases of substituted judgement the autonomy of the individual is respected indirectly.

SJS is "a weak autonomy standard" (Beauchamp & Childress 2001, p. 99). Importantly, "…we should reject the standard of substituted judgement for never-competent patients. No basis exists for a judgement of autonomous choice if a person has never been autonomous" (Beauchamp & Childress 2001, p. 100). As one can see, SJS cannot account for individuals who have never been autonomous, since they have never expressed any preferences. In fact, they might never have expressed anything at all.[66]

Recall the patient with dementia mentioned above. Before the onset of dementia, and in its early stages, she was autonomous. The preferences she expressed then are, according to SJS, now to be respected. However, these preferences must, as I see it, be to some extent realistic and compatible with the situation the patient now inhabits. To respect autonomy indirectly in this sense might be problematic.

The condition of a patient with dementia deteriorates gradually over time. It is hard to draw a precise limit after which the individual no longer can be claimed to be autonomous. In borderline cases it might be hard for relatives or medical staff to determine which preferences ought to be taken into account, depending on when, in the past, they were uttered by the patient. Should the preferences uttered closest in time to the moment at which the patient became demented be considered more important, or more worthy of respect, than those uttered earlier, perhaps before the onset of the degenerative disease?

Moreover, who knows best about the patient's interests: the clinician, or relatives; and if the latter, which relatives? In order to determine this, it is important to consider both the efforts and interests of relatives, and of course the support they can offer, when the patient is no longer capable of exercising her autonomy. In addition, a surrogate decision might be hard to make, since the condition of the patient might vary from day to day; it might change with the context.

It is certainly important to consider substituted judgement in connection with an indirect usage of the principle of autonomy. To argue that an individual must *presently* be autonomous, or must *presently* exercise autonomy, in order to have the right to be treated as autonomous is not necessary. Plausibly, a weak variety of respect for autonomy can be applied in situations where the autonomy of the patient has been undermined. Consider, for instance, cases of severe dementia, or cognitive impairment resulting from a stroke.

---

[66] For an interesting discussion of this topic, see John Davies (2002).

It is time to sum up. To have the right to autonomy one must be, or have been at some point in the past, autonomous. If one has never been autonomous, one cannot have the right to autonomy. In cases where autonomy is undermined, the autonomy of the individual can be respected indirectly on the assumption that she once was autonomous and expressed certain preferences that can now be considered and dealt with. Unfortunately, there is no space here to discuss in further detail various aspects of substituted judgement. However, one must have in mind the potential practical problems its application might involve.

In the autonomy debate, it is important to be explicit about the possibility of an indirect usage of the autonomy principle. This issue, and the distinction between autonomy as a right and a capacity, help to clarify some complexities in our dealings with questions about autonomy in the area of healthcare. For instance, they clarify what it means to be autonomous, and in what sense autonomy can be respected. They also emphasize the way in which the capacity for autonomy hangs together with the right to have one's autonomy respected. I hope that the present discussion will contribute to a wider debate about the important issues concerning autonomy in healthcare.

I will now discuss some problems that arise when the principle of autonomy is applied in too broad a fashion, i.e. extended to groups of individuals who are expected to be autonomous but whose autonomy can be questioned.

6.4. The Autonomy Triumph and Limitations of the Autonomy Principle

The principle of autonomy is sometimes treated almost as something sacred. This can lead to a sense that it should not be questioned and need not be defended, and sometimes it is indeed blindly advocated. Blind defences of the principle of autonomy might be problematic when it comes to individuals whose autonomy can be questioned. It can be argued that disrespect for an individual's right to exercise autonomy is objectionable, to be sure. But equally, a blind defence of autonomy is problematic.

Below I will discuss limitations on respecting the right to exercise autonomy and ask whether the principle of autonomy has been over-applied. Is there a risk that the autonomy principle itself has encouraged over-application? As I see it, it is important to emphasize the problem of defending the principle of autonomy in cases where the patient's autonomy is undermined.

I want to make two major points below. The first concerns the empty right to exercise autonomy. This right comes into play when the individual lacks the capacity to exercise

autonomy. The second is that it could be harmful to have a right to exercise autonomy if the individual is incapable of doing so. Let me begin with the first point.

Individuals who have the right to exercise autonomy sometimes lack the corresponding capacity to do so. In such cases, I would claim, the right to exercise autonomy is empty. In fact, I believe individuals who lack the capacity required are sometimes expected to be autonomous as a result of blindly defending autonomy, since a blind defence will tend to obscure, or de-emphasize, the importance of the metacognitive capacity required. As was claimed earlier, in healthcare the role of the metacognitive capacity of the individual has not been taken sufficiently into account.

One might wonder whether "autonomy" has become a prestige word even in situations where the right to autonomy reasonably cannot be defended and applied: where the right is empty. Autonomy, as a right to be respected, may have become overused in healthcare, and this might be damaging to the individuals suffering from a reduced metacognitive capacity. Perhaps respect for autonomy, in a controversial sense, is applied too broadly in healthcare. This is my second point.

It is reasonable to assume that depriving an individual of the opportunity to exercise her autonomy might be morally violating. For instance, in the Health and Medical Service Act (SFS 1982:763) it is explicitly stated that the patient is to have the right to participate, and to decide upon her own care. This is one of the basic characteristics of good care. Nevertheless, we might also consider cases where it would be harmful (and morally violating) to expect an individual to exercise her autonomy – cases in which the individual is not, in point of fact, autonomous. Equally, neglecting the individual's incapacity to lead a life under her own control might also be morally violating. Below I explain why.

To expect or require individuals to be autonomous is not always reasonable. An individual with limited autonomy might not receive the healthcare she needs because she is expected to be autonomous and assumed to be able to live a controlled life at the global level. To have one's needs satisfied, and in order to make one's voice heard, one might have to be healthy enough. The individual has to be active and make her own decisions (Hansson 2006). The expectation that individuals suffering from cognitive impairments are autonomous can sometimes lead to quite strong demands (Levinsson 2006). Leaving some patients in their own hands to take care of themselves might be too demanding. To some patients, liability can be difficult to realize, or genuinely take on. The liability might concern the management of one's daily chores, or situations where important decisions have to be made, say, about one's

financial arrangements. There are patients who are in need of great support, but whose needs risk being neglected because of the steady focus on respect for autonomy.

It has been pointed out that impaired cognitive capacity often leads to serious difficulties in the management of daily activities (The National Board of Health and Welfare 2003b). To expect weak individuals to be autonomous and able to exercise autonomy, while they in fact have serious difficulties with that, is to risk neglecting their basic needs. Evidently, if one's basic needs cannot be provided for, this will affect the individual's life from a global perspective. Such consequences, stemming from a preoccupation with respecting the patient's right to exercise autonomy, do not seem to be in line with the goal of good care stated in the Health and Medical Service Act (see SFS 1982:763).

According to Beauchamp and Childress (2001, p. 61) some writers have criticized the triumph of autonomy on the grounds that, when misused, autonomy might turn into forced autonomy. For instance, patients sometimes do not wish to receive information about their condition, or do not want to participate in, influence, or decide upon, their care. According to Beauchamp and Childress, it is important to emphasize that autonomy is a right, not a duty: the patient has the right to decide whether she wishes to influence her care. However, it is not her duty to do so.

In healthcare there is a danger of mistaking the right to autonomy for a duty. Perhaps this is the case more often than is recognized in everyday healthcare practice. We might speculate that lack of time among medical staff and economic factors are factors hindering the patient's opportunity to obtain the care she needs. Rather more liability is put on the patient's own potential to be active and take care of herself.

Since not all individuals want to be involved, or participate, in their own care, preoccupation with the principle of autonomy in healthcare might force individuals to make decisions. For instance, the patient might prefer not to receive, or manage, information about her disease and medication even if she is autonomous and therefore has the right to influence her care. In a situation where a patient refuses her right to exercise autonomy, this desire is to be respected.

If autonomy is indeed one of the highest values to be promoted in the healthcare setting, such (as it were) anti-autonomous preferences are to be respected. If not, it would be hard to see why respecting autonomy positively would contribute to the patient's influence over her care. However, respect for a patient's right to exercise her autonomy requires that the patient does not suffer from, say, impaired metacognitive functioning. Where she does, the medical

staff will normally have greater insight into the disease than she, and consequently they should be able to make wiser decisions with respect to treatment and so on.

The somewhat absurd consequences of the blind defence of the need to respect the patient's autonomy and its exercise with respect to preferences are exemplified in Sandman (2004).

> A former nurse was operated on for a kidney problem and returned to the ward after surgery. Since she was familiar with the ward and its staff, they offered her to choose whatever pain-killers she preferred. Still affected by sedation, she could only think of the mild kind of pain-killers she normally bought without prescription at the drug store and asked for a couple of those. Respecting her autonomy, she was given two such mild pain-killers which did not relieve her pain in any significant degree. (Sandman 2004, p. 266)

These remarks are illuminating. Sandman's case involves a patient who is simply too confused by, say, sedation, or perhaps anxiety and worry, to make effective decisions. The case forces us to ask to what extent autonomy should be respected. For instance, are patient preferences expressed in sedated states to be respected? Is it plausible to regard an individual in a sedated state as someone with the ability to exercise autonomy? As I see it, it is implausible to expect a patient under the influence of sedation to be capable of exercising control.

Other examples that illustrate the line of thought presented by Sandman, but where no medical preparations are involved, and where the right to autonomy can be questioned, can be given. Consider, for instance, cases where the patient's irrational care preferences contradict the recommendations of medical staff. Are these preferences to be respected by the medical staff? Suppose the preferences of the patient would, if acted upon, be harmful to her. What would be the grounds for not respecting the patient's preferences? If members of the medical team begin to intervene or decide for the patient, then, even if the intervention would be advantageous and is based on reasonable grounds, they to some extent break the principle of autonomy. As was claimed earlier, it is plausible to suppose that in cases where the medical team has more insight into the patient's situation than the patient herself – e.g. due to metacognitive impairment – they will be able to make wiser decisions concerning the care of the patient.

It may sometimes be unreasonable to defend the view that the recommendations of medical staff always are positive. One should be aware of the problem of how the medical staff might influence a patient's choices. As has been claimed earlier in this thesis, external influences, like advice from other people, can contribute positively to the exercise of

autonomy. Such factors need not violate control. Nevertheless, quite *which* external factors positively or negatively influence the exercise of one's autonomy can be difficult to say.

Consider a situation in which medical staff try to persuade a patient to accept a medical alternative because it is claimed to be better than another alternative. Suppose the patient finally agrees. Does the patient exercise autonomy in such a situation, or has she been persuaded by the staff? As I see it, it may be hard to determine whether the patient in such a situation is persuaded, or manipulated, rather than exercising her autonomy. In cases like the above it might be hard to separate the patient's own preferences and those stated by the medical staff.

I have discussed some reasons why the autonomy principle can be questioned in healthcare. For instance, preoccupation with the patient's right to exercise autonomy may lead to over-emphasis on autonomy. In some respects, such use might become coercive and negative. As has been claimed, an individual can be autonomous without wanting to be active in decisions about her care. However, this, I think, is normally seen as something that it is problematic to respect, since the patient, in such situations, does not take an active role in her own care. However, the autonomous individual may have good reasons for preferring not to influence her own care: she might have sound reasons for not wanting to exercise her autonomy and instead for wishing to be under the control of medical staff.

The discussion above seems to have identified at least three different understandings of autonomy that are relevant to consider with respect to healthcare. Let me begin with the first one.

The first is an understanding of the concept of autonomy in terms of a *capacity*, a property of the individual it is to some extent possible to measure and verify from an external point of view. Nevertheless, this first understanding is not clearly defined in healthcare. The present analysis has put forward an understanding of this property in terms of a metacognitive capacity. This analysis will, I hope, help to clarify what kind of property autonomy is, and in what ways it can be undermined.

The analysis of the concept of autonomy as a metacognitive capacity leads to a deeper discussion of the way in which the variety of autonomy at stake in healthcare might be operationalized. In Chapter 5 it was claimed that the determination of whether or not an individual's autonomy is undermined has, plausibly, to rely on external requirements. These external requirements, it was suggested, can be understood in terms of the two components of metacognition argued for in the thesis. These components are procedural reflexivity and metarepresentation. The latter involves both inferential reflexivity and other-attributiveness.

As I see it, this suggestion perhaps makes it possible to operationalize autonomy (and undermined autonomy) and implement it in healthcare.

To summarize, the operationalization of autonomy (and undermined autonomy) in terms of metacognition, as it is understood in this thesis, is tentative. However, the suggested idea might be instructive in relation to questions about how to determine, or measure, autonomy (and undermined autonomy), and how to implement the above theoretical ideas in healthcare.

The second understanding concerns autonomy as a *regulative principle* which proclaims a right to be treated in certain ways. In my view, this is the most common understanding of autonomy in healthcare.

Third, there also seems to exist an understanding that can be interpreted in terms of *forced autonomy*: this treats individuals as though they are rational and in control of their senses and emotions. However, and as was claimed above, to regard autonomy as a duty rather than a right is problematic. The concept of autonomy as a duty seems to presuppose an ideal picture of the individual. For instance, it can be argued that autonomy is too often assumed, even in cases where the patient does not want to participate in questions about her care (and again in cases where autonomy cannot be exercised because of metacognitive impairment).

Let me now turn to discuss the role of the concept of autonomy and the principle of autonomy in Swedish psychiatry.

Chapter 7

Autonomy and Psychiatry

7.0. Vulnerability and Healthcare Provision in Swedish Psychiatry

In the present chapter I will discuss the concept of autonomy and the principle of autonomy in relation to Swedish psychiatry. A potential problem in psychiatry arises where the application of the principle might consider individuals whose autonomy is undermined because of reduced metacognitive functioning. To expect autonomy from an individual suffering from persistent mental disorder can be problematic for her daily life. Plausibly, it is problematic to expect something from an individual that she does not have.

People who suffer from severe mental disorder belong to a vulnerable group in society with regard to the capacity to lead one's own life. Consider, for instance, schizophrenia and bipolar disorder. A major problem that needs to be considered in connection with the right to exercise one's autonomy is that severe mental disorders often are persistent.

From a global perspective, respecting autonomy becomes problematic, because the disorders affect the individual's life as a whole, both personally and socially. Research claims that recovery is rare among patients who suffer from schizophrenia and that schizophrenia is a disorder that still is incurable (The National Board of Health and Welfare 1991, 2003b). It is also believed that schizophrenia has a worse prognosis than other psychoses (The National Board of Health and Welfare 2003b). However, other findings claim that one-half to two-thirds of the patients recover from their symptoms or at least reach considerable improvement (Harding et al. 1987).

It is important to see, however, that treatment of schizophrenia must be distinguished from rehabilitation. While the former aims to reduce symptoms, the latter aims to increase the functional capacity of the individual and improve her chance of participating in society (The National Board of Health and Welfare 2003b).

In the act healthcare delivery it might be hard to respect the right to exercise autonomy among patients who suffer from severe mental disorders. As was claimed, they might lack the required capacity needed to function in daily life. As Lindley (1986, p. 162) illuminates, schizophrenia is a serious destroyer of an individual's autonomy. The research indicates that patients suffering from schizophrenia lack insight to their cognitive impairments which "…might be due to a more global deficit in metacognitive abilities" (Kircher et al. 2007, p.

254). However, research on relapse indicates that "…psychoeducational family interventions reduce the relapse and rehospitalisation rates of schizophrenia patients" (Pitschel-Walz et al. 2001, p. 86) Data also indicate that psychoeducational interventions with medical treatment are essential in the treatment of schizophrenia. Family interventions and medical treatment in combination, according to the study, are superior to medical treatment alone. Other empirical data also indicate that psychotherapeutic or family interventions might be effective when combined with community care of patients suffering from schizophrenia (Östman 2000; Thornicroft & Susser 2001).

The fact that the concept and principle of autonomy are debated in psychiatry is not difficult to understand. A major problem – one that raises conceptual questions about what it means to be autonomous as well as moral questions about the right to be respected – is that it is controversial whether it is possible to respect autonomy in cases of metacognitive impairment. This raises further questions about the right to be treated in certain ways, and about the circumstances under which coercive care is justified. An important question arises about the alternative kinds of healthcare that can be implemented in practice where patients have undermined autonomy.[67]

Below I will consider various approaches to psychiatric care in Swedish psychiatry. Special emphasis will be put on deinstitutionalization and participation in society. Coercive care will also briefly be discussed. I end the thesis by suggesting how improvements in psychiatric care could be made by taking into account what has been argued for in this thesis: the two-dimensional view of autonomy that emphasizes both the metacognitive capacity of the individual and the external setting in which it is exercised.

7.1. Deinstitutionalization

From a historical perspective, and following several psychiatric reforms in Sweden, institutionalization today is, as I understand it, controversial in Swedish psychiatry; in several

[67] There is no consensus on questions about autonomy and severe mental disorder. My experiences, gained while participating at psychiatry conferences, is that in some forums respect for autonomy is highlighted, while in others the impossibility of being autonomous when suffering from severe mental disorder is admitted. As was argued earlier, the problem of understanding exactly how to deal with the concept of autonomy in psychiatry might depend on moral issues about autonomy becoming mixed up with issues concerning the nature of the capacity required if an individual is to be autonomous.

respects it is regarded as an old-fashioned and inhumane kind of care. From an historical point of view, institutionalization has been criticized (Markström 2003, p. 112).

Institutionalization is sometimes associated with societal control and concerns about safety – with the idea that individuals who suffer from severe mental disorders are more dangerous and more often commit violent crimes than other groups in society. Research on violent crime looking at severe mental disorder nevertheless points in several different directions, and it is hard to draw firm conclusions about whether these patients, statistically, commit more violent crimes than those in other groups. According to Kullgren (2003), there is no connection between deinstitutionalization and violent crimes in society. However, the probability that an individual will commit violent crime is higher when she suffering from psychosis than it is when she is not (SOU 2006:100). For a discussion that deals with data on violence and mental disorder in Sweden, see Kullgren (2003).[68] On the problem of predicting future dangerous behaviour, see Grisso and Appelbaum (1997, p. 446).

In order to maintain or strengthen the autonomy of individuals suffering from mental disorder different reforms in Swedish psychiatry have been revised with the aim of creating and integrating more humane environments than before.[69] For instance, the decision to close the mental institutions made around 1970 was one move in this direction (Markström 2003, p. 113). A primary aim of the closure was to provide the opportunity for the patients to live a "normal" life outside the walls of the mental hospital. The right to societal integration and participation like that enjoyed by other groups in society has therefore become a central issue in the move towards more humane and community-based care for patients who suffer from severe mental disorder.

Swedish psychiatric enquiries carried out in Sweden over the years suggest that patients suffering from mental disorder should have the same rights and duties as everybody else in society (Markström 2003, p. 155; Grönwall & Holgersson 2006 p. 50). Moreover, it has also been stated explicitly that the patient's own choices and priorities are to be the starting point of all care efforts.

---

[68] See also The National Board of Health and Welfare (2003a).

[69] For detailed discussion of the Swedish psychiatric reforms, see Markström (2003).

According to Östman (2000), the idea of deinstitutionalization emerged during World War II.

> The era of deinstitutionalization began during and after World War II when mental health professionals were out of the state institutions, where they had been largely based, doing diagnostic screening and dealing with war-related psychiatric disorders among combatants. Their mission was to discharge the patient from hospital as quickly as possible and return them to their combat zones. (Östman 2000, p. 3)

Östman further writes:

> There was also a growing movement among mental health advocates and social scientists that institutionalization was inappropriate for many patients, that commitment procedures were often arbitrary and inhumane, and that long-term custodial care could result in dependency and increased dysfunction rather than improvement of the mental disorder. (Östman 2000, ibid)

According to Östman, institutionalization was not seen as the most appropriate treatment in rehabilitation in order to return to a new life. As I see it, the emphasis here on integration and participation rests on the idea that environmental factors positively influence recovery and well-being among patients who suffer from severe mental disorder. The idea behind integration is to move away from the isolation, dependence, and passivity caused by institutionalization, and towards socialization and activity.

For many patients deinstitutionalization has surely been an advantageous and significant step towards a better and more meaningful life; but perhaps not for all. The note of reservation here is expressed uncompromisingly by Appelbaum (1997). He states that

> …for some patients, discharge from the state hospitals was a blessing. For all too many others, it was the ultimate curse. Far from a panacea, the policy created as many problems as it solved, perhaps more. (Appelbaum 1997, p. 548)

While Appelbaum's claim concerns deinstitutionalization in the United States, it is important to note that Swedish psychiatry and the varieties of healthcare provided within it do not differ significantly from those of other western countries (Markström 2003, p. 112). It therefore would not be controversial to reason in a similar way about Swedish psychiatry.

Drawbacks of deinstitutionalization in the United States have also been pointed out by Hermann (1997).

> Anyone spending time in major urban center in the United States must be shocked by the significant number of mentally ill persons living on the streets—the "bag people" who sleep in doorways, on steam grates, on subway stairs. These people represent a new lifestyle made possible in part by a policy of deinstitutionalization of the mentally ill, which has been motivated largely by economic considerations and rationalized as a matter of mental health law reform. (Hermann 1997, p. 462)

Hermann continues by stating that "The policy of deinstitutionalization and these reforms of commitment law ignore the reality of mental illness—many mentally ill persons lack the ability to make rational decisions about their treatment needs" (Hermann 1997, p. 462).

Cullberg (2003, p. 298) claims that various forces in Sweden over recent decades have combined in a drive to increase efficiency in the public sector, and that this probably will enhance the care and custody of stronger groups in society. However, as Cullberg notes, what will happen to the vulnerable groups in society, especially patients who suffer from persistent schizophrenia and lack social networks, one can only speculate about.

Patients suffering from severe mental disorders, like schizophrenia, are vulnerable because they have a reduced metacognitive capacity. However, as compared to for instance dementia, the reduced metacognitive capacity in schizophrenia is often shifting and intermittent. Yet, it is questionable whether it is plausible to expect these patients to be able to live a "normal" life like everybody else. In some respects, to expect autonomy from vulnerable patients might involve more suffering than amelioration and recovery. In the living of a "normal" life a certain amount of control is needed in order to survive and to maintain everyday functioning.

What are the consequences of the closure of the mental institutions? As was claimed, deinstitutionalization may not have wholly positive consequences for the patient. This has been pointed out by Cullberg.

> One can only speculate about how society is affected by the massive return of those who, over many years, have been dispatched into closed institutions. The release of a large amount of the anxiety and pain that earlier was locked away and forgotten can hardly pass without leaving traces. (Cullberg 2003, p. 298, my translation)[70]

There appear to be no clear answers to questions about the consequences of the psychiatric reforms. However, one may speculate that the closure of the hospitals and the developments of more community-based forms of care are problematic in some respects. I will now consider some of the drawbacks of community-based activities.

7.2. Societal Participation

Following the closure of the mental hospitals, deinstitutionalization and community-based care have become central topics in psychiatry.[71] At the same time, the implementation of such care can be demanding for patients suffering from a persistent mental disorder. Thus it can be hard for the patient to respond to new and urban environments, which often are stressful and demanding. Research indicate that urbanization tends to increase the risk of persistent depression or psychosis (Sundqvist, Frank & Sundqvist 2004). According to Sundqvist, Frank and Sundqvist, such data must be considered by those who manage treatment and prevention of illness in individuals suffering from mental disorders.

> Our findings suggest that the level of urbanisation is associated with psychosis and depression in both women and men. For clinicians in urban areas who are involved in both treatment and prevention of disease, it is of great importance to consider possible pathways in the development of psychiatric morbidity. These pathways might include lack of social support, stressful life events and familial liability. Moreover, when planning the distribution of health care resources, it is important to consider the level of urbanisation in order to improve services for people who are at high risk of developing psychiatric morbidity. (Sundqvist, Frank & Sundqvist 2004, p. 297)

---

[70] In Swedish: Hur samhället påverkas av ett massivt återvändande av dem som man under många år förpassat in på stängda institutioner kan man bara spekulera över. Att släppa ut en stor mängd av den ångest och smärta som man tidigare låst in och fått glömma av kan knappast komma att gå spårlöst förbi (Cullberg 2003, s. 298).

[71] See, for instance, Burns and Firn (2005).

"Normal" life in society requires the individual to be active and able to find care and medical resources on her own, in order to fulfil her needs. It has also been stated that the specific kind of environment is crucial in the planning of healthcare efforts in psychiatry. Life outside the institution, in some cases, can be too demanding for a patient with a severe mental disorder.

It has been reported that psychiatric patients do not seek the care to which they have a right – or, at least, that they seek it, or use it, less than other groups in society do. This is problematic, on the assumption that the resources do in fact exist. According to a report carried out by the Swedish Government Official Reports (SOU), "Lack of initiative, impaired insight into illness, problems of estimating consequences, and communicative difficulties, are common problems involved in some kinds of mental state or as phenomena in mental impairment" (SOU 2006:100, my translation).[72] A reasonable interpretation of the data here would be that, even if care resources are available, the patient may have difficulties, stemming from her disorder, actively finding the care to which she has a right.

It is sometimes claimed that patients suffering from mental disorder do not receive the care of which they are in need. However, it is also claimed these patients "…underuse mental healthcare services" (Alonso et al. 2000).[73] According to Cullberg, patients, because of their mental disorder, are in need of others who can make their voices heard. As I understand the matter, a "normal" life outside of the hospital requires one to be able to make one's voice heard; and vulnerable groups, such as patients suffering from schizophrenia, have difficulty doing that. Research also indicates that social interaction is important for recovery from schizophrenia (The National of Health and Welfare 1991). As I see it, support offered by other people is central in community-based care. As Cullberg writes:

> It seems that the alternatives lie between living unseen in an institution or equally unseen on the pavement. Yet we do know that an initial deliberative effort would also lead to the living of a good life for many in these groups. (Cullberg 2003, p. 298, my translation)[74]

---

[72] In Swedish: Initiativlöshet, bristande sjukdomsinsikt, svårigheter att se konsekvenser eller svårigheter att kommunicera är vanliga problem som ingår i vissa psykiska tillstånd eller är fenomen i det psykiska funktionshindret (SOU 2006:100).

[73] For discussion of this topic see, for instance, Bijl and Ravelli (2000) and Alonso et al. (2007).

[74] In Swedish: Det är som om alternativen skulle ligga mellan att leva osedd på institution eller lika osedd på trottoaren. Ändå vet vi idag att med en medveten satsning från början skulle också stora delar av också dessa grupper kunna leva ett gott liv (Cullberg 2003, s. 298).

Medication is often claimed to be a first response used to bring about "normal" life for the patients suffering from persistent mental disorders (The National Board of Health and Welfare 2003b; Ottosson 2004). However, it is not uncommon to combine medication with other kinds of support, like psychotherapy or some form of social arrangement (The National Board of Health and Welfare 2003b; Ottosson 2004). For many patients with schizophrenia medication is needed in order function at all but also to achieve therapeutic results.

While several medical drugs reduce symptoms, they have side effects. Although side effects today are often minimized, it must still be borne in mind that some drugs have a tendency to reduce metacognitive functioning. For instance, laziness, passivity and emotional flattening can negatively influence the individual's capacity to cope with society.[75] Importantly, one's ability to exercise autonomy might be hampered by medication and its side effects. Even if the doctor claims that medication is necessary if the patient is to function better in society, at the same time the drugs used might impede the patient's everyday functioning and, hence, be an obstacle to the patient's efforts to exercise autonomy.

Let me summarize the line of thought presented above. Medication is claimed to be necessary if the patient with persistent mental disorder is to function in everyday life. However, we must also remember that the effect of the medication might be to reduce the patient's metacognitive capacity, and so might negatively affect her exercise of autonomy. In this sense, medication will not necessarily raise the probability of autonomy's being exercised. Both medical treatment and symptoms of the relevant disorder might decrease metacognitive functioning at a global level, and hence affect the autonomy of the individual.

From a societal perspective, the expectation that patients who belong to vulnerable groups will be autonomous is problematic. Research has shown that the risk of mortality and morbidity are overrepresented among patients who suffer from persistent mental disorders (Borgå et al. 1992; SOU 2006:100; The National Board of Health and Welfare 2003b). One can of course speculate about why this is the case. One speculation is that the individual is expected to be able to take care of herself when she in fact cannot. Life outside the hospital's walls affords more opportunity to maintain control and more opportunities to make one's own decisions; but it also involves facing more demands. For example, the patient is supposed to take care of her home, her financial arrangements, her medication, her hygiene, and her social networks.

---

[75] See also Lindley (1986, p. 156).

Without scientific support or evidence, we can still speculate sensibly as to whether, homelessness, suicide, loneliness, and isolation among individuals who suffer from persistent mental disorder are possible consequences of insufficient resources as well as the expectation of autonomy – an expectation that allows for opportunities but also places heavier requirements on the patient. Confirmation that isolation is common among some patients suffering from schizophrenia can be found in material published by The National Board of Health and Welfare (2003b).

Let me present a quotation from a survey carried out by The National Board of Health and Welfare that emphasizes the line of thought presented here.[76]

> Cognition is a summary designation concerning our capacity to receive and interpret impressions in order to handle them rightly, leading to purposeful actions. Examples of such fundamental functions are attention, endurance, memory, language and aspects of intelligence. The cognitive functions are therefore fundamental in order to operate in daily life. Impaired cognitive capacity is very common in schizophrenia. Several patients have such serious impairments that the capacity to manage the fundamental demands of daily life is compromised. (The National Board of Health and Welfare 2003b, my translation)[77]

Consider a patient with schizophrenia suffering from paranoia, delusion or apathy; or a patient with dementia who, in certain periods, does not know who she is. In an important respect, metacognitive functioning seems to be presupposed by the living of a "normal" life in society. The quotation above motivates further discussion and development of the concept of autonomy in psychiatry; it also requires us to look again at the kinds of healthcare routinely offered to patients who suffer from persistent mental disorder. Next, I will briefly deal with coercive care in relation to the concept of autonomy as a metacognitive capacity.

---

[76] This quotation was also presented in 1.3.

[77] In Swedish: Kognition är en sammanfattande beteckning på vår förmåga att ta in och tolka intryck för att sedan behandla dem på rätt sätt så att de leder till ändamålsenliga handlingar. Exempel på sådana grundläggande funktioner är uppmärksamhet, uthållighet, minne, språk och aspekter på intelligens. De kognitiva funktionerna är därför grundläggande för att det agliga livet skall fungera. Försämrad kognitiv funktionsförmåga är mycket vanligt vid schizofreni. Flera patienter har så allvarliga försämringar att det försvårar förmågan att klara de grundläggande krav som ställs i det dagliga livet.

7.3. Coercive care

Coercive care, as it is normally understood in psychiatry, has been explained thus by Breggin (1997):

> By coercion is meant any action, or threat of action, which compels the patient to behave in a manner inconsistent with his own wishes. The compelling aspect can be direct physical or chemical restraint, or it can be indirect threatened recriminations of indirect "force of authority" which convinces the patient that no other legal or medical alternative is available to him. (Breggin 1997, pp. 424-425)

As John Stuart Mill once claimed, society sometimes has to persuade, or interfere, when the individual has false beliefs or destructive intentions (Bromwich & Kateb 2003). Coercive care, as dealt with below, constitutes an exception to the principle of autonomy. As Tännsjö puts it:

> There are situations where the use of coercive care is appropriate. Society sometimes has to allow that medical or other kinds of treatment be *forced* upon patients who desperately need to be treated, but who refuse to undergo the needed treatment voluntarily. And, although this may be a shocking thing to say, society ought to sanction that sometimes patients and clients be in various different ways *manipulated* to receive the treatment they desperately need but refuse to accept voluntarily. (Tännsjö 1999, p. 1)

An important question in both psychiatry and law concerns the circumstances under which coercive care is legitimate. According to Tännsjö, psychiatric coercive care is appropriate when the patient can be claimed to be incompetent:

> When people who suffer from mental illness become incapable, because of their illness, of reaching a decision about their need for treatment, when they need treatment for their own sake (unless they are treated their health is put in jeopardy), then these people ought to be coercively admitted to a psychiatric ward and treated for their mental illness. (Tännsjö 1999, p. 103)

Moreover, in relation to the patient, coercive care must be understood with respect to "…his or her *capacity* to make decisions… [It is] not a matter of the content of these decisions themselves" (Tännsjö 1999, p. 11).

According to The Compulsory Mental Care Act (SFS 1991:1128) coercive care presupposes that the patient is suffering from a severe mental disorder and, further, that the patient is in need of patient care because of her psychiatric state. In addition, as already stated, coercive care is legitimate when the patient refuses, or cannot be claimed to consent to, the

care in question. If coercion is to be legitimate, it must also be considered whether the patient might run the risk of harming herself or others.

As can be seen from The Compulsory Mental Care Act (SFS 1991:1128), coercive care is appropriate when the patient who suffers from a severe mental disorder refuses care. In such situations, it is claimed, the authorities are better at deciding what is best for the individual than she is herself. Importantly, and from a long-term perspective, involuntary hospitalization need not violate the right to have one's autonomy respected. In one respect, an individual's autonomy can be indirectly respected by claiming that coercive care of the non-autonomous patient aims to restore autonomy in the future. In this sense, coercive care is seen as a benevolent intervention that is in the patient's best interests.

It is often claimed that coercive care is used as little as possible, but critics counter that coercion may affect too many patients suffering from severe mental disorder (Lindley 1986). For instance, the doctors who decide whether a patient is to receive coercive care might regard the patient as non-autonomous with respect to a certain task and thus decide that the patient is non-autonomous in general. To clarify this, the point of the objection is that there is a risk of generalizing an individual's capacity from a few competence-specific tasks. Rather, whether an individual is autonomous is to be understood from a global perspective (recall that there exist two common understandings of autonomy, the local and global conceptions).

Tännsjö (1999, p. 3) claims that criteria for coercive must be spelled out in clear and unambiguous terms. However, in the mental health literature, it is sometimes claimed that coercive care is poorly defined (O'Brien & Golding 2003). The analysis of the concept of autonomy in this thesis might bring clarity to the question under what circumstances coercive care is appropriate. I suggest that the underlying criteria determining undermined autonomy should agree with the criteria of coercive care. On my view, coercive care is compatible with the global perspective on the concept of autonomy emphasized in this thesis.

As I see it, the question under what circumstances coercive care is appropriate must be understood with respect to undermined autonomy. Whether coercive care is appropriate or not depends on the individual's metacognitive capacity. Coercive care is to be judged in relation to a correct view of the metacognitive capacity for autonomy. Decisions about coercive care must be motivated through, or rely upon, judgements about the individual's capacity for autonomy understood from a global perspective. Coercive care is appropriate when one's metacognitive capacity, in a global sense, is seriously reduced.

The global conception of autonomy as a metacognitive capacity of the individual is in agreement with the aims of the Swedish psychiatry reforms, which highlight participation in

"normal" life of the kind enjoyed by other groups in society. Reasonably, this requires that we understand what it means to be autonomous from a global perspective. As I see it, community-based healthcare initiatives rely on ideas of enhancing and maintaining global autonomy. If I am right, psychiatric reforms point to a global understanding of the concept of autonomy that considers large parts of, or periods in, an individual's life. As was previously claimed, if global autonomy is undermined, coercive care should be considered.

Let me briefly comment on the global and local conception of autonomy, and on their relation to decisions in forensic psychiatry. On what grounds should efforts be made when an individual commits, say, a violent crime? Is the individual to be provided with healthcare or punishment? With regard to the kind of intervention that is appropriate, one might ask whether the metacognitive capacity of the individual is to be considered solely with respect to the present situation, where the crime was committed, or whether the metacognitive capacity of the individual is to be considered from a global perspective. I leave this as an open question. However, I believe this question needs to be examined in more detail in connection with autonomy, healthcare interventions, forensic psychiatry, and the law.

It is time to sum up this discussion of coercive care. A patient's need for coercive care should be judged by criteria that concern the metacognitive capacity for global autonomy. This requires the concept of autonomy to be clearly understood. In the present thesis it has been argued that the concept of autonomy is to be understood in terms of a metacognitive capacity of the individual. This capacity has two components: procedural reflexivity and metarepresentation. These components, it was suggested, might work as external criteria of autonomy, i.e. as criteria determining whether or not an individual is autonomous. Only when such criteria have been considered can the patient's right to autonomy be assessed. This, in turn, will perhaps make it possible to understand the criteria for coercive care in improved, unambiguous and clarified terms. The hope is that the criteria for autonomy will permit decisions to be made about whether the patient should be submitted to coercive care in the individual case. Importantly, implementation of coercive care must reflect the understanding of autonomy both as a right and as a metacognitive capacity giving the individual control.

Chapter 8

Future Considerations

8.0. Suggestions and Improvements

I end the thesis by making some suggestions and outlining important some issues that might help to improve research into, and discussion of, the concept of autonomy in both philosophy and medicine (under the latter heading I include healthcare and psychiatric issues). It is my hope that the suggestions will function as advice, or virtual guidelines, for decision makers in psychiatry as well as other parts of the medical field in which various kinds of healthcare activity are evaluated.

The Importance of the Two-dimensional Perspective

Responses to the issues raised by autonomy tend to focus on just one of the two dimensions of autonomy while neglecting the other. This problem was presented in Part I. However, when developing healthcare provision for patients who suffer from severe mental disorder, the internal, as well as the external, dimension of autonomy has to be taken into account.[78]

If the relation between the metacognitive capacity of the individual and the social environment she inhabits becomes neglected, deinstitutionalization and community-based treatment may lead to inhumane rather than effective care. This problem must be considered in relation to the sometimes excessive importance accorded to autonomy. Several patients today suffer because they are left alone with responsibility in their own hands. Are these patients receiving the care they need? It is doubtful that they are.

More research into the metacognitive underpinnings of autonomy is needed, in both philosophy and the empirical sciences. In connection with research into metacognition it is important to emphasize that metacognitive functioning must be understood in relation to the external setting in which it is exercised. For instance, without intact metacognition it can be very hard for an individual to live an autonomous life in society. To clarify, in order to understand autonomy and its exercise one must consider *both* the metacognitive capacity of the individual (e.g. her ability to manage her own mental states and to plan in relation to future goals) *and* external factors (including those of the societal, cultural and legal kind). A

---

[78] This idea has also been raised by Guinn (2002) in connection with patients suffering from dementia.

patient who suffers from a persistent mental disorder might have difficulty relating either to her mental activity or to her social environment, or to both. Importantly, internal and external factors are in interplay: hence they influence each other. Neither can be excluded if we wish to understand autonomy in terms of metacognition.

Different Mental Disorders Require Different Kinds of Care

The role and possibility of autonomy must be considered in relation to the type of mental disorder. More knowledge of the nature of the mental disorders is needed, but also knowledge of the way in which particular disorders function in specific external environments. Different types of mental disorder invite and require different responses – just as happens, as is very well known, with physical or bodily ailments. Consider, for instance, how a physical handicap might become more or less easy to handle depending on the external setting.

In psychiatric discussions there is a risk, I think, of lumping together patients who suffer from severe mental disorders in one group. Equally, patients suffering from schizophrenia may be assumed to constitute a homogenous group, although in point of fact schizophrenia comes in several different types with different symptoms (Ottosson 2004). Like non-mental disorders, mental disorders have different aetiologies and symptoms. Therefore, healthcare provision must be flexible enough to reflect the specific disorder and its symptoms.

The above claim is problematic, however, for this reason: descriptions and symptoms of mental disorders sometimes overlap. As I see it, this is because we lack complete knowledge of the nature of the mental disorders. However, my aim here is to emphasize the thought that individuals suffering from severe and persistent mental disorders will not normally be in need of the same healthcare.

Let me exemplify the line of thought I am trying to get at here. About the aetiology of schizophrenia, research shows differences between the two sexes. For instance, empirical data suggest later onset of the disease in females (Strömgren 1987). Moreover, "…combined drug and social therapy seems to work much better in females than in males" (Strömgren 1987, p. 5).

It is important to consider the symptoms of the disorder when one is trying to determine what healthcare arrangements are required. The development of better diagnoses may lead to more effective support, but not without improvements in our understanding of the aetiology of the disorder. However, handbooks of diagnostic systems like DSM IV, as I understand them, work primarily with symptoms, not the origin of the disorder. Moreover, it would be crucial to consider how impaired metacognitive function interacts with the environment the patient

inhabits. A single individual might function differently in different environments so far as her metacognitive impairment is concerned.

To summarize, then, autonomy is to some extent relational. Internal, as well as external factors make the exercise of autonomy possible. However, they might also constitute barriers. As has previously been claimed, it is important to identify environments which maintain metacognitive functioning and those which do not. In the search for such identification, the role of metacognition in relation to an individual's functioning in external environments has to be considered. The metacognitive impairment the individual is suffering from can be more or less compatible with the environments in which she lives her life. As was mentioned, research indicates that urban environments may increase the risk of a descent into depression or psychosis (Sundqvist, Frank & Sundqvist 2004).

Patients who suffer from severe mental disorder also tend to become lonely when left in their homes and isolated from society (SOU 2006:100). If they are to live a "normal" life, healthcare built upon the idea of the active individual might not be advantageous for individuals who belong to vulnerable groups. For instance, why do patients who suffer from severe mental disorder seek care less than other groups in society do? Is it because they have difficulty exercising their autonomy? Is the range of care ineptly designed in relation to their capacity to seek help? Alternatively, might the claim that these patients seek care less be false? As I see it, these questions need to be answered.

On the other hand, research indicates that social programmes and family interventions, when combined with medical treatment, mitigate the symptoms schizophrenia, have a positive effect on recovery, and deter relapse.[79] However, relations between the factors involved in the healthcare regime can become problematic, since they might change over time. This can impact negatively (as well as positively) on an individual's autonomy. The obvious fact that environmental conditions might change has, I think, to be highlighted in connection with the provision of psychiatric care.

Global and Local Perspectives as Complementary

The different global and local conceptions of autonomy are problematic for various reasons. In medicine and bioethics, autonomy is commonly understood as a local feature focusing on decision-making competence in a particular situation – e.g. a medical situation in which a decision needs to be made about treatment. Nevertheless, in view of fact that one of the goals

[79] Moreover, personal proxies have begun to be incorporated in community-based healthcare on a frequent basis.

of the Swedish psychiatric reforms is to promote autonomy at large and the opportunity of sufferers to live "normal" lives in society, a local view of autonomy becomes problematic. This is a major reason why a global perspective has been put forward in this dissertation. The psychiatric reforms are compatible with a global understanding of autonomy. In other words, the global understanding of autonomy is consistent with the aims of the psychiatric reforms.

The relevance and value of local and global conceptions of autonomy varies depending on the kind of enquiry. They function differently depending on what kind of purpose one has and the kind of problem being dealt with. However, the global and local conceptions of autonomy can plausibly be seen as complementary. It is a problem with the global-local distinction that a patient may lack autonomy in a global sense but possess it locally. In what sense can the autonomy of the individual be promoted in such cases, given that the psychiatric reforms emphasize the right to live a "normal" life? On the other hand, a problem with the global perspective of autonomy is that it is hard to determine to what extent, exactly, an individual must be autonomous. As was claimed in Part I, a global view must allow for temporary losses of autonomy, like cases of emergency and states of shock. Probably, developments in methods of determining autonomy will mitigate this problem. To answer the question of how to determine autonomy we shall, I believe, require multidisciplinary teamwork involving competence from the humanities, medicine and the social sciences. The discussion of external requirements of autonomy in Part I explains why, I hope.

In psychiatry there are, I believe, at least two contradictory views about how to deal with the issues raised by autonomy and severe mental disorder. The first deal with the problem of recovery in schizophrenia and that these patients, as a result of cognitive impairment, have serious difficulty managing the basic demands that confront them in daily life. On this view, it is plausible to claim that the autonomy of the patient is undermined, and that the right to exercise autonomy is something that it would be hard to promote.

On the other hand, since psychoses, for example, are in many cases not chronic, it would be unwise here to claim that the patient necessarily lacks autonomy on the global level. The claim ought to be merely that she lacks it when she is in a psychotic state.

The two views previously presented become especially problematic when the local conception of autonomy is mixed up with the global conception. As I see it, it is important to be explicit about whether, in the ongoing discussion, we are considering local or global autonomy. One might further ask how this problem influences psychiatric care provision, and when it would be relevant to consider autonomy from a local perspective, and when from a global. However, as previously explained, because the global perspective on autonomy is

consistent with the aims of the Swedish psychiatric reforms, it is that perspective that has been emphasized here.

Importantly, persistent mental disorder does not necessarily imply a chronically disordered state. An individual's metacognitive capacity to exercise control can vary from day to day. This, of course, has an impact both globally and locally. One must acknowledge that it might be hard to generalize about a patient's capacity over time. This difficulty must be considered in the development of psychiatric healthcare efforts.

Autonomy, Law, and Coercive Care

It is important to consider the relation between coercion and autonomy. Does coercive care rely on the idea that the patient lacks, or has reduced, autonomy? If that is the case, how is the concept of autonomy understood? Might coercive care rely on an incorrect or insufficiently complete, understanding of the concept of autonomy? Here it would, I think, be fruitful to strengthen cooperation between law, philosophy and medicine.

The link between conceptual issues about autonomy and the criteria of coercive care should, I believe, be strengthened. This strengthened link might, perhaps, lead to fruitful developments on questions about coercive healthcare, and on the issue of how, precisely, the concept of autonomy is to be understood in the legal context.

8.1. Concluding Discussion

I want to end the thesis by acknowledging that much, in the concept of autonomy as it is deployed in healthcare, is unclear. To foster clarity I have argued that autonomy is not primarily a right to be respected. In healthcare and psychiatry, the concept of autonomy concerns the individual's metacognitive capacity for control. If we want to defend the view of the patient's reinforced position, this capacity must be clarified and made explicit. The above analysis and discussion have explained why: the right to have one's autonomy respected must rely on the individual's metacognitive capacity for control.

The possibility that one is autonomous and the possibility that one can exercise it must be separated and understood two-dimensionally. In a favourable environment, the patient who suffers from reduced metacognitive capacity might well be able to exercise her autonomy effectively. In other environments, she might not. This is why the metacognitive capacity of the individual must always be understood in relation to the social environment in which it is

exercised. Further, the social environment must be evaluated in relation to the metacognitive functioning of the individual.

If decisions concerning the care of patients are to be based upon an assessment of autonomy, both the internal and external dimensions have to be considered. The following two questions are important when the autonomy of the individual is to be determined, and in decision-making situations where healthcare provision is being evaluated.

(i) Does the individual possess the necessary metacognitive capacities (described in Part I)?

(ii) Will the metacognitive capacities of the individual function in the social context(s) she inhabits?

If autonomy is to be sought and fostered in healthcare and psychiatry, the emphasis must be on the individual's metacognitive capacity for control in relation to external factors. The latter factors might facilitate the exercise of autonomy, but importantly they might also undermine its exercise.

In some cases it might be possible to defend the imposed use of support measures while at the same time claiming to maintain autonomy. Consider for instance community treatment and housing support in psychiatry. As has been argued, external factors such as advice from other people do not undermine the exercise of autonomy.

However, sometimes support is more important to the patient than respect for autonomy. That is, in situations where autonomy is hard to promote or maintain, the provision of adequate support is more important than blindly defending the principle of autonomy. The present analysis has deepened our understanding of how to approach and deal with this claim.

Where global autonomy is undermined the implementation of varieties of healthcare that do not emphasize the right to exercise autonomy becomes more reasonable. Nevertheless, in healthcare the line between the concept of autonomy as a right (or sometimes as a duty) and the neighbouring concept of autonomy as a metacognitive capacity is indeed thin.

References

Agich, George J. (2004). Seeking the Everyday Meaning of Autonomy in Neurologic Disorders. *Philosophy, Psychiatry, & Psychology*, vol. 11:4, pp. 295-298.

Agich, George & Mordini, Emilio (1998). Autonomy and the ethics of neurosurgery. *Italian Journal of Psychiatry and Behavioral Sciences*. vol. 2, pp. 47–55.

Alonso, Jordi, Codony, Miquel, Kovess, Viviane, Angermeyer, Matthias C., Katz, Steven J., Haro, Josep. M., De Girolamo, Giovanni, De Graaf, Ron, Demyttenaere, Koen, Vilagut, Gemma, Almansa, Josue, Lépine Jean Pierre & Traolach, Brugha S. (2007). Population level of unmet need for mental healthcare in Europe. *British Journal of Psychiatry,* vol.190:4, pp. 299-306.

American Psychiatric Association (2000). *The diagnostic and statistical manual of mental disorders: DSM IV*. 4th. ed. Washington DC: American Psychiatric Association.

Anderson, Joel (2008). Disputing Autonomy: Second-Order Desires and the Dynamics of Ascribing Autonomy. *Sats - Nordic Journal of Philosophy*, vol. 9, pp. 7-26.

Anderson, Joel & Lux Warren (2004a). Accurate Self-Assessment, Autonomous Ignorance, and the Appreciation of Disability. *Philosophy, Psychiatry, and Psychology*, vol. 11:4, pp. 309-312.

Anderson, Joel & Lux Warren (2004b) Knowing Your Own Strength: Accurate Self-Assessment as a Requirement for Personal Autonomy. *Philosophy, Psychiatry, and Psychology*, vol. 11:4, pp. 279-294.

Appelbaum, Paul (1997). Crazy in the streets. In Edwards, Rem B. (ed.). *Psychiatry and ethics: insanity, rational autonomy, and mental health care*. New York: Prometheus books.

Appelbaum, Paul & Grisso Thomas (1995). The MacArthur Treatment Competence Study, I: Mental illness and competence to consent to treatment. *Law and Human Behavior*, vol. 19:2, pp. 105-25.

Beauchamp, Tom L. & Childress, James F. (2001). *Principles of biomedical ethics*. 5th. ed. New York: Oxford University Press.

Beauchamp, Tom L. (2005). Who Deserves Autonomy, and Whose Autonomy Deserves Respect? In Taylor, James S. (ed.). *Personal Autonomy: New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy*. Cambridge: Cambridge University Press. P. 310-329).

Berghmans, Ron, Dickenson, Donna & Meulen, Ruud Ter (2004). Mental Capacity: In Search of Alternative Perspectives. *Health Care Analysis*, vol. 12:4, pp. 251-263.

Berofsky, Bernard (1995). *Liberation from self: a theory of personal autonomy*. Cambridge: Cambridge University Press.

Berofsky, Bernard (2005). Autonomy Without Free Will. In Taylor, James S (ed.). *Personal Autonomy: New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy*. Cambridge: Cambridge University Press. P. 58-86.

Bijl, Rob & Ravelli Anneloes (2000). Psychiatric morbidity, service use, and need for care in the general population: results of the Netherlands Mental Health Survey and Incidence Study. *American Journal of Public Health*, vol. 90:4, pp. 602–608.

Bonjour, Laurence (1985). *The structure of empirical knowledge*. Cambridge, MA: Harvard University Press.

Borgå Per, Widerlöv Birgitta, Stefansson Claes-Göran & Culberg, Johan (1992). Social conditions in a total population with long-term functional psychosis in three different areas of Stockholm County. *Acta Psychiatrica Scandinavia*, vol. 85:6, pp. 465-73.

Breden, Torsten M & Vollman Jochen (2004). The Cognitive Based Approach of Capacity Assessment in Psychiatry: A Philosophical Critique of the MacCAT-T. *Health Care Analysis*, vol. 12:4, pp. 273-283.

Breggin, Peter R. (1997). Coercion of Voluntary Patients in an Open Hospital. In Edwards, Rem B. (ed.). *Ethics of Psychiatry: insanity, rational autonomy, and mental health care*. New York: Prometheus books. P. 423-436.

Brinck, Ingar (1997). *The Indexical 'I'. The First Person in Thought and Language*. Diss., Lund university. Dordrecht: Kluwer Academic Publishers.

Brinck, Ingar (2003). The objects of attention: Causes and targets. *Behavioral and brain sciences*, vol. 26:3, pp. 287-288.

Brinck, Ingar (2004). Joint attention, triangulation and radical interpretation: A problem and its solution. *Dialectica*, vol. 58:2, pp. 179-205.

Brinck, Ingar (2005). Review. John Campbell: Reference and Consciousness. *Theoria*, vol. 71:3, pp. 266-276.

Brinck, Ingar (2007). Situated cognition, dynamic systems, and art. On artistic creativity and aesthetic experience. *JanusHead*, vol. 9:2, pp. 407-431.

Bromwich, David & Kateb, George (eds.) (2003). *On liberty: John Stuart Mill*. New Haven: Yale University Press.

Broström, Linus (2007). *The substituted judgment standard: studies on the ethics of surrogate decision making*. Diss., Lund University. Lund: Faculty of Medicine.

Buchanan, Allen E. & Brock, Dan W. (1989). *Deciding for others: The ethics of surrogate decision making*. Cambridge: Cambridge University Press.

Burns, Tom & Firn, Mike (2005). *Samhällsbaserad psykiatrisk vård: en handbok för praktiker*. Lund: Studentlitteratur AB.

Buss, Sarah (2002). Personal Autonomy. (Elektronisk) *Stanford Encyclopedia of Philosophy*. Available: < http://plato.stanford.edu/entries/personal-autonomy/ > (2004-05-19)

Camille, Nathalie, Coricelli, Giorgio, Sallet, Jerome, Pradat-Diehl, Pascale, Duhamel, Jean-René & Sirigu Angela (2004). The Involvement of the Orbitofrontal Cortex in the Experience of Regret. Science, vol. 304:5674, pp. 1167-1170.

Campbell, John (1999). Schizophrenia, the space of reasons, and thinking as a motor process. *Monist*, vol. 82:4, pp. 609-626.

Campbell, John (2001). Rationality, Meaning, and the Analysis of Delusion, *Philosophy, Psychiatry, and Psychology*, vol. 8:2-3, pp. 89-100.

Campbell, John (2002). The Ownership of Thoughts. *Philosophy, Psychiatry, and Psychology*, vol. 9:1, pp. 35-39.

Carnap, Rudolf (1950). *Logical foundations of probability*. Chicago: The University of Chicago Press.

Chadwick, Ruth (2004). The Right Not to Know: A Challenge for Accurate Self-Assessment. *Philosophy, Psychiatry, and Psychology*, vol. 11:4, pp. 299-301.

Charland, Louis (1998). Is Mr. Spock Mentally Competent? Competence to Consent and Emotion. *Philosophy, Psychiatry, and Psychology*, vol. 5:1, pp. 67-81.

Charland, Louis (1999). Appreciation and Emotion: Theoretical Reflections on the MacArthur Treatment Competence Study. *Kennedy Institute of Ethics Journal*, vol. 8:4, pp. 359-376.

Charland, Louis (2007). Anorexia and the MacCAT-T Test for Mental Competence: Validity, Value, and Emotion. *Philosophy, Psychiatry, and Psychology*, vol. 13:4, pp. 283-287.

Christman, John (1988). Constructing the Inner Citadel: Recent Work on the Concept of Autonomy. *Ethics*, vol. 99:1, pp. 109-124.

Christman, John (1989). *The inner citadel: essays on individual autonomy*. New York: Oxford University Press.

Christman, John (1991). Autonomy and Personal History. *Canadian Journal of Philosophy*, vol. 21, pp. 1-24.

Christman, John (2001). Liberalism, autonomy and self-transformation. *Social Theory and Practice*, vol. 27:2, pp.185-207.

Christman, John (2003). Autonomy in moral and political philosophy. (Elektronisk) *Stanford Encyclopedia of Philosophy*. Available: < http://plato.stanford.edu/entries/autonomy-moral/ > (2004-06-10)

Christman, John (2004). Relational Autonomy, Liberal Individualism, and the Social Constitution of Selves. *Philosophical Studies*, vol. 117:1-2, pp. 143-164.

Christman, John (2005). Procedural Autonomy and Liberal Legitimacy. In Taylor, James S. (ed.). *Personal Autonomy: New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy*. Cambridge: Cambridge University Press. P. 277-298.

Christman, John & Anderson, Joel (eds.) (2005). *Autonomy and the challenges to liberalism*. Cambridge: Cambridge University Press.

Coliva, Annalisa (2002). Thought Insertion and Immunity to Error Through Misidentification. *Philosophy, Psychiatry, and Psychology*, vol. 9:1, pp. 27-34.

Cullberg, Johan (2003). *Dynamisk Psykiatri*. Stockholm: Natur och Kultur.

Damasio, Antonio R. (1994). *Descartes' error : emotion, reason, and the human brain*. New York: Penguin Books.

Davidson, Donald (1991). Three Varieties of Knowledge. In Griffiths, Phillips A. (ed.). *A. J. Ayer: Memorial Essays*. Cambridge: Cambridge University Press. P. 153-166.

Davidson, Donald (1992). The Second Person. In *Subjective, Intersubjective, Objective*. Oxford: Oxford University Press, 2001. P. 107-122.

Davies, John (2002). The Concept of Precedent Autonomy. *Bioethics*, vol. 16:2, pp. 114-133.

Dienes, Zoltan & Perner, Josef (1999). A theory of implicit and explicit knowledge. *Behavioral and brain sciences*, vol. 22:5, pp. 735-808.

Dijksterhuis, Ap & Meurs, Teun (2006). Where creativity resides: The generative power of unconscious thought. *Consciousness and Cognition*, vol. 15:1, pp.135–146.

Dijksterhuis, Ap & van Olden, Zeger (2006). On the benefits of thinking unconsciously: Unconscious thought can increase post-choice satisfaction. *Journal of Experimental Social Psychology*, vol. 42:5, pp. 627-631.

Dworkin, Gerald (1970). Acting Freely. *Nous*, vol. 4:4, pp. 367-383.

Dworkin, Gerald (1976). Autonomy and Behavior Control. *Hastings Center Report*, vol. 6:1, pp. 23-28.

Dworkin, Gerald (1988). *The theory and practice of autonomy.* Cambridge: Cambridge University Press.

Edwards, Rem B. (1997). Mental Health as Rational Autonomy. In Edwards, Rem B. (ed.). *Ethics of Psychiatry: insanity, rational autonomy, and mental health care*. New York: Prometheus books.

Ekstrom, Laura (1993). A coherence theory of autonomy. *Philosophy and Phenomenological Research*, vol. 53:3, pp.599-616.

Ekstrom, Laura (1999) Review. Keystone Preferences and Autonomy. *Philosophy and Phenomenological Research*, vol. 59:4, pp.1057-1063.

Ekstrom, Laura (2005a). Alienation, Autonomy, and the Self. *Midwest Studies in Philosophy*, vol. 29:1, pp. 45-67.

Ekstrom, Laura (2005b). Autonomy and Personal Integration. In Taylor, James S. (ed.). *Personal Autonomy: New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy*. Cambridge: Cambridge University Press. P. 143-161.

Feinberg, Joel (1989). Autonomy. In Christman, John (ed.). *The Inner Citadel: Essays on Individual Autonomy*. New York: Oxford University Press. P. 27-53.

Finucane, Melissa, Alhakami, Ali, Slovic, Paul & Johnson, Stephen (2000). The affect heuristic in judgments of risks and benefits. *Journal of Behavioral Decision Making*, vol. 13:1 pp. 1-17.

Frankfurt, Harry (1971). Freedom of the Will and the Concept of a Person. *The Journal of Philosophy*, vol. 68:1, pp. 5-20.
Frankfurt, Harry G. (1988). *The importance of what we care about: philosophical essays.* Cambridge: Cambridge University Press.

Frankfurt, Harry (1999). The Faintest Passion. In Frankfurt Harry (ed.). *Necessity volition and love*. Cambridge: Cambridge University Press. P. 95-107.

Frankfurt, Harry G. (2004). *The reasons of love*. Cambridge: Cambridge University Press.

Gazzaniga, Michael S., Ivry, Richard B. & Mangun, George R. (1998). *Cognitive neuroscience: the biology of the mind*. New York: W. W. Norton.

Grisso, Thomas & Appelbaum, Paul S. (1997). Is It Unethical to Offer Predictions of Future Violence? In Edwards, Rem B. (ed.). *Ethics of Psychiatry: insanity, rational autonomy, and mental health care*. New York: Prometheus books. P. 446-461.

Grönwall, Lars & Holgersson, Leif (2006). *Psykiatrin, tvånget och lagen; en lagkommentar i historisk belysning.* Stockholm: Norstedts juridik.

Guinn, David (2002). Mental Competence, Caregivers, and the Process of Consent: Research Involving Alzheimer's Patients or Others with Decreasing Mental Capacity. *Cambridge quarterly of healthcare ethics*, vol. 11:3, pp. 230-245.

Hansson, Kristofer (2006). Etiska utmaningar i en föränderlig hälso- och sjukvård. In Hansson, Kristofer (ed.). *Etiska utmaningar: i hälso- och sjukvården*. Lund: Studentlitteratur AB. P. 11-26.

Harding, Courtenay M., Brooks, George W., Ashikaga, Takamaru, Strauss, John S. & Breier Alan (1987). The Vermont Longitudinal Study of Persons With Severe Mental Illness, I: Methodology, Study Sample, and Overall Status 32 Years Later. *American Journal of Psychiatry*, vol. 144:6, pp. 727-735.

Haworth, Lawrence (1986). *Autonomy: an essay in philosophical psychology and ethics.* New Haven: Yale University Press.

Hermann, Donald H. J.(1997). A Critique of Revisions in Procedural, Substantive, and Dispositional Criteria in Involuntary Civil Commitment. In Edwards, Rem B. (ed.). *Ethics of Psychiatry: insanity, rational autonomy, and mental health care*. New York: Prometheus books. P. 462-483.

Hermerén, Göran (2006). Sjukvårdsetik i tider av förändring. In Hansson, Kristoffer (ed.). *Etiska utmaningar: i hälso- och sjukvården.* Lund: Studentlitteratur AB. P.161-191.

Hill, Thomas E. (1991). *Autonomy and self-respect*. Cambridge: Cambridge University Press.

Hume, David (1739/2004). *A Treatise of Human Nature*. New York: Dover Publications Inc.

Juth, Niklas (2005). *Genetic information: Values and Rights. The morality of presymptomatic Genetic Testing*. Diss., Göteborg university. Göteborg: Acta Philosophica Gothoburgensia.

Kircher, T.J. Tilo, Koch, Kathrin, Stottmeister, Frank & Durst, Volker (2007). Metacognition and Reflexivity in Patients with Schizophrenia. *Psychopathology*, vol. 40:4, pp. 254-260.

Kullgren, Gunnar (2003). *Våldsbrott och psykisk sjukdom*. Stockholm: Socialstyrelsen.

Lazarus, Richard S. & Lazarus, Bernice N. (1994). *Passion and reason: making sense of our emotions*. Oxford: Oxford University Press.

LeDoux, Joseph (1998). *The Emotional Brain: the mysterious underpinnings of emotional life*. London: Weidenfeld & Nicolson.

Le Granse, Mieke, Kinébanian, Astrid & Josephsson, Staffan (2006). Promoting autonomy of the client with persistent mental illness: a challenge for occupational therapists from The Netherlands, Germany and Belgium. *Occupational therapy international*, vol. 13:3, pp. 142-159.

Lehrer, Keith (1990). *Metamind*. Oxford: Clarendon.

Lehrer, Keith (1999a). Review: Replies. *Philosophy and Phenomenological Research*, vol 59:4, pp. 1065-1074.

Lehrer, Keith (1999b). *Self-trust: a study of reason, knowledge and autonomy*. New York: Clarendon Press.

Levinsson, Henrik (2006). Den självbestämmande individen. In Hansson, Kristofer (ed.). *Etiska utmaningar: i hälso- och sjukvården.* Lund: Studentlitteratur AB. P. 105-124

Lindley, Richard (1986). *Autonomy*. Basingstoke: Macmillan.

Markström, Urban (2003). *Den svenska psykiatrireformen: bland brukare, eldsjälar och byråkrater*. Diss., Umeå university. Umeå: Boréa förlag.

May, Thomas (2005). The Concept of Autonomy in Bioethics: An Unwarranted Fall from Grace. In Taylor, James S. (ed.). *Personal Autonomy: New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy*. Cambridge: Cambridge University Press. P. 209-309.

Mele, Alfred R. (1995). *Autonomous agents: from self-control to autonomy*. New York: Oxford University Press.

Metcalfe, Janet & Kober, Hedy (2005). Self-reflective consciousness
and the projectable self. In Terrace, Herbert & Metcalfe, Janet (eds.). *The Missing Link in Cognition: Origins of Self-Reflective Consciousness*. New York: Oxford University Press. P. 57-83.

Metcalfe, Janet & Shimamura, Arthur P. (eds.) (1994). *Metacognition: knowing about knowing*. Cambridge, MA: MIT Press.

Meyers, Diana (1987). Personal Autonomy and the Paradox of Feminine Socialization. *The Journal of Philosophy*, vol. 84:11, pp. 619-628.
Murphy, Dominic (2004). Autonomy, Experience, and Therapy. *Philosophy, Psychiatry, & Psychology*, vol. 11:4, pp. 303-307.

MFR (2002). *Riktlinjer för etisk värdering av medicinsk humanforskning: forskningsetisk policy och organisation i Sverige*. Stockholm: Medicinska forskningsrådet. (Rapport/Medicinska forskningsrådet: 2. reviderad version)

Ministry of Health and Social Affairs (2006). *Ambition och ansvar. Nationell strategi för utveckling av samhällets insatser till personer med psykiska sjukdomar och funktionshinder*. Stockholm: Fritzes. (Statens offentliga utredningar 2006:100)

Neisser, Joseph (2006). Unconscious Subjectivity. *Psyche: An interdisciplinary Journal of Research on Consciousness*, vol. 12:3.

Nelson, Thomas (1996). Consciousness and metacognition. *American psychologist*, vol. 51:2, pp. 102-116.

Nelson, Thomas & Narens, Louis (1990). Metamemory: A theoretical framework and new findings. In Bower, Gordon (ed.). *The psychology of learning and motivation*, vol. 26, pp. 125–140. New York: Academic Press.

Noggle, Robert (2005). Autonomy and the Paradox of Self-creation: Infinite Regresses, Finite Selves, and the Limits of Authenticity. In Taylor, James S. (ed.). *Personal Autonomy: New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy*. Cambridge: Cambridge University Press. P. 87-108.

Nordgren, Lars (2003). *Från patient till kund: intåget av marknadstänkande i sjukvården och förskjutningen av patientens position*. Diss., Lund university. Lund: Lund Business Press.

O'Brien, Anthony & Golding, Clinton (2003). Coercion in mental healthcare: the principle of least coercive care. *Journal of Psychiatric & Mental Health Nursing*, vol. 10:2, pp.167-173.

Oshana, Marina (2001). The Autonomy Bogeyman. *Journal of Value Inquiry*, vol. 35:2, pp. 209-226.

Oshana, Marina (2006). *Personal autonomy in society*. Aldershot: Ashgate.

Ottosson, Jan-Otto (2004). *Psykiatri*. Stockholm: Liber.

Passer, Michael W. & Smith, Ronald E. (2008). *Psychology: the science of mind and behaviour*. 4th ed. Boston : McGraw-Hill Higher Education.

Pitschel-Walz, Gabi, Leucht, Stefan, Bäuml, Josef, Kissling, Werner & Engel, Rolf R. (2001). The Effect of Family Interventions on Relapse and Rehospitalization in Schizophrenia: A Meta-analysis. *Schizophrenia Bulletin*, vol. 27:1, pp. 73-92.

President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research (1983). *Deciding to forego life-sustaining treatment.* Washington DC: U.S. Government Printing Office.

Proust, Joëlle (1999). Self-model and schizophrenia. *Consciousness and Cognition*, vol. 8:3, pp. 378-384.

Proust, Joëlle (2003). Does metacognition necessarily involve metarepresentation? *Behavioral and brain sciences*, vol 26:3, pp. 352-352.

Proust Joëlle (2006). Agency in schizophrenia from a control theory viewpoint. In Sebanz, Natalie & Prinz, Wolfgang (eds.). *Disorders of volition*. Cambridge: MIT Press. P. 87-118.

Proust, Joëlle (2007). Metacognition and metarepresentation: is a self-directed theory of mind a precondition metacognition? *Synthese*, vol. 159:2, pp. 271-295.

Psykologilexikonet (2005). Stockholm: Natur och kultur.

Pylyshyn, Zenon (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, vol. 22:3, pp. 341-365.

Quinn, Philip L. (1998). In *Routledge Encyclopedia of Philosophy*. Vol. 8, pp. 611-613.

Robertsson, Barbro, Nordström, Monica & Wijk, Helle (2007). Investigating poor insight in Alzheimer"s disease: A survey of research approaches. *Dementia*, vol. 6:1, pp. 45-61.

Rorty, Richard (1991). *Essays on Heidegger and Others: Philosophical Papers vol. 2*. New York: Cambridge University Press.

Sandman, Lars (2004). On the autonomy turf. Assessing the value of autonomy to patients. *Medicine, Healthcare & Philosophy*, vol. 7:3, pp. 261-268.

SFS (1982). *Hälso- och sjukvårdslag.* 1982:763. (The Health and Medical Service Act). Stockholm: Svensk författningssamling.

SFS (1991). *Lag om psykiatrisk tvångsvård*. 1991:1128. (The Compulsory Mental Care Act). Stockholm: Svensk författningssamling.

SFS (1993). *Lag om stöd och service till vissa funktionshindrade* 1993:387. (Act concerning Support and Service for Persons with Certain Functional Impairments). Stockholm: Svensk författningssamling.

SFS (2001). *Socialtjänstlag*. 2001:453. (Medical Services Act). Stockholm: Svensk författningssamling.

Sjöstrand, Manne & Helgesson, Gert (2008). Coercive Treatment and Autonomy in Psychiatry. *Bioethics*, vol. 22:2, pp. 113-120.

Shimamura, Arthur P. (2000). Toward a cognitive neuroscience of metacognition. *Consciousness and Cognition*, vol. 9:2, pp. 313–323.

Slovic, Paul, Finucane, Melissa, Peters, Ellen & McGregor, Donald (2007). The affect heuristic. *European Journal of Operational Research*, vol. 177:3, pp. 1333-1352.

Smith, David J., Shields, Wendy E. & Washburn, David A. (2003). The comparative psychology of uncertainty monitoring and metacognition. Behavioral and brain sciences, vol. 26:3, pp. 317-339.

Strömgren, Erik (1987). Changes in the incidence of schizophrenia? *The British Journal of Psychiatry*, vol. 150, pp. 1-7.

Sundqvist, Kristina, Frank, Gölin & Sundqvist, Jan (2004). Urbanisation and incidence of psychosis and depression: follow-up study of 4.4 million women and men in Sweden. *British Journal of Psychiatry*, vol. 184:4, pp. 293-298.

Tan, Jacinta, Stewart, Anne, Fitzpatrick, Ray & Hope, Tony (2007). Competence to Make Treatment Decisions in Anorexia Nervosa: Thinking Processes and Values. *Philosophy, Psychiatry, & Psychology*, vol. 13:4, pp. 267-282.

Taylor, James S. (ed.) (2005). *Personal Autonomy: New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy*. Cambridge: Cambridge University Press.

The National Board of Health and Welfare (1991). *Tvång - autonomi. Etik i psykiatri*. Stockholm: Socialstyrelsen. (Rapport/Socialstyrelsen:19)

The National Board of Health Care and Welfare (2003a). *Utredning av händelserna i Åkeshov och Gamla stan och dess möjliga samband med brister i bemötande och behandling inom den psykiatriska vården och socialtjänstens verksamhet*. Stockholm: Socialstyrelsen. (Rapport/ Dnr 00-5240/2003)

The National Board of Health and Welfare (2003b). *Vård och stöd till patienter med schizofreni: en kunskapsöversikt*. Stockholm: Socialstyrelsen.

Thornicroft, Graham & Susser, Ezra (2001). Evidence-based psychotherapeutic interventions in the community care of schizophrenia. *British Journal of Psychiatry*, vol. 178, pp. 2-4.

Tännsjö, Torbjörn (1999). *Coercive care: the ethics of choice in health and medicine.* London: Routledge.

Tännsjö, Torbjörn (2008). *Vårdetik*. Stockholm: Thales.

Waller, Bruce (1998). *The Natural Selection of Autonomy*. New York: State University of New York Press.

Watson, Gary (1975). Free Agency. *Journal of Philosophy*, vol. 72:8, pp. 205-220.

Östman, Margareta (2000). *Family burden and participation in care. A study of relatives to voluntarily and compulsorily admitted patients*. Diss., Lund university.