RATIONAL AND IRRATIONAL AGENCY

by

PETER G. CAMPBELL

B.A., The University of British Columbia, 1983
M.A., The University of British Columbia, 1988

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

Department of Philosophy

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

September, 1994

© Peter G. Campbell, 1994
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

(Signature)

Department of **PHILOSOPHY**

The University of British Columbia
Vancouver, Canada

Date **DEC. 19, 1994**
ABSTRACT

Only with a comprehensive detailed theory of the practical processes which agents engage in prior to successful action can one get a picture of all those junctures at which the mechanism of rationality may be applied, and at which irrationality may therefore occur. Rationality, I argue, is the exercise of normatives, such as believable and desirable, whose function is to control the formation of the stages in practical processes by determining what content and which functions of practical states are allowed into the process. Believable is a functional concept, and for an agent to wield it requires that he possess beliefs or a theory he can justify about which states are goal-functional. Desirable is likewise a functional concept, and its exercise requires that agents possess justifiable beliefs or a theory about which goals are to be functional. When the desirability belief functions, it does so according to ideals of the theory. For example, it functions saliently where desires become intentions. So long as the normatives function in these ways the agent is rational. To so function is to satisfy the ideal for agency itself.

Chapter 2 presents a fine-grained model of the fundamental terms and relations necessary for practical reasoning and agency. In this model, the functions of belief, desire and intention are described in naturalized terms. On the basis of this account of the terms of agency, a taxonomy of the possible failures of rationally controlled practicality is presented in chapter 3.
Chapter 4 presents a comprehensive and detailed account of intention formation comprised of the functions of belief, desire and intention. Wherever one of those functions occurs in the process is a juncture at which rationality may be exercised, and therefore a point at which irrationality may occur.

In chapter 5 I describe some of the main ways that dysfunctional states may disrupt agency, creating irrationality. The measures agents may take to ameliorate or otherwise control such failures are discussed and distinguished according to the ideal of agency. Finally, and in these terms, I address the problem of akrasia, in particular the views of Davidson and Mele, and show that the room they make for strict akratic action involves a significant compromise of the ideals of agency, and therefore is not as “strict” as they and others have claimed.
CONTENTS

ABSTRACT ii
ACKNOWLEDGMENT vi

CHAPTER 1 THE IDEALS AND FAILURES OF AGENCY
1. INTRODUCTION ................................................................. 1
2. ASSUMPTIONS AND METHODOLOGY .................................... 3
   2.1 The Common-Sense Folk-Psychological Framework .................. 3
   2.2 What is Agency? .......................................................... 4
   2.3 The Intentional and Deliberative Conception of the Agent ........... 5
3. THE PRACTICAL AND THE RATIONAL .................................... 8
4. THE LOGICAL SPACE OF AGENCY ......................................... 14

CHAPTER 2 THE TERMS OF AGENCY
1. INTRODUCTION ................................................................. 19
2. THE FUNDAMENTAL TERMS OF AGENCY .................................. 23
   2.1 Causation and Its Terms ................................................ 23
   2.2 The Representing and the Represented: The Types of Particulars Required for Agency ................................................... 24
   2.3 Distinctions Among The Mental Terms of Agency ................. 26
3. THE STATES OF PRACTICAL RATIONAL AGENCY AND THEIR FUNCTIONS ...... 28
   3.1 The Structure and Function of the Active Attitudes: The Fundamental Practical State .................................................. 28
   3.2 The Rational Origins of Desires-To .................................. 32
   3.3 The Pre-Practical Active Attitude: Desire’s Ancestor .............. 34
4. THE PRACTICAL FUNCTIONS OF BELIEF .................................. 36
   4.1 The Belief Function in Content Concatenation ..................... 36
   4.2 Content Decatenation .................................................... 41
   4.3 The Practical Effects of Certain Beliefs ............................. 43
   4.4 The Structure of the Content of Belief and Desire ................ 48
   4.5 Content Concatenation and Practical Reason ....................... 51
   4.6 Full Decatenation of Content—An Exercise in Rationality ......... 53
5. DISTINCTIONS AMONG THE PRACTICAL STATES OF THE PROCESS OF AGENCY .... 56
   5.1 Desire-that vs. Desire-to .............................................. 57
   5.2 Desire-to vs. Intention .................................................. 58

CHAPTER 3 THE FUNDAMENTAL AGENCY FAILURES: A NOSOLOGY OF AGENCY,
PART 1
1. INTRODUCTION .................................................................. 60
2. THE INFORMATIONAL DOMAIN OF AGENCY ............................. 63
3. THE CONTENT MALFUNCTIONS .......................................... 67
   3.1 Semantic Malfunctions ................................................... 67
   3.2 Syntax Malfunctions ....................................................... 70
4. THE DYSFUNCTIONS .......................................................... 72
   4.1 Recalcitrance ............................................................... 74
   4.2 Incontinence ................................................................. 76
   4.3 Garrulousness ............................................................... 78
   4.4 Reticence ..................................................................... 81
5. CONCLUSION .................................................................... 84
CHAPTER 4 INTENTION FORMATION: THE FUNDAMENTAL PROCESS OF AGENCY
1. INTRODUCTION
2. PRE-PRACTICAL REASONING: THE CAUSAL AND COGNITIVE ORIGINS OF DESIRE-TO
   2.1 From Proto-desire to Desire-that
   2.2 From Desire-that to Desire-to
3. PRACTICAL REASONING, FROM DESIRE-TO TO EFFICACIOUS INTENTION: THE
   CAUSAL AND COGNITIVE ORIGINS OF INTENTION
4. THE EFFECTS OF PRACTICAL REASONING
   4.1 The Practical Syllogism: Desire
   4.2 The Practical Syllogism: Intention
   4.3 Minimal Means
5. INTENDING
   5.1 An Intention's Relations with Desire and Belief
   5.2 Commitment
   5.3 Purity
   5.4 Relations with Planning and Other Intentions
   5.5 Sufficiency for Action
   5.6 Relations with Intentional Action and Responsibility
6. THEORETICAL SMOOTHNESS

CHAPTER 5 RATIONAL AND IRRATIONAL AGENCY
1. INTRODUCTION
2. THE CORE OF RATIONAL AGENCY
   2.1 Proto-desire to Desire-that
   2.2 Desire-that to Desire-to
   2.3 Desire-to to Intention
   2.4 Means-end Reasoning
3. RATIONALITY BEYOND THE CORE
   3.1 Thickening the Agent
4. STATE DYSFUNCTION IN THE PROCESS: A NOSOLOGY OF AGENCY, PART 2
   4.1 Recalcitrance
   4.2 Incontinence
   4.3 Garrulousness
   4.4 Reticence
5. IDEAL AND NON-IDEAL CONTROL, REPAIR AND AVOIDANCE OF NON-GOAL-FUNCTIONALITY
   5.1 Rationality and Control
   5.2 Ideal Repair of Dysfunctionality
      5.2.1 Emergent Dysfunctions
      5.2.2 Competing Desires: Incontinence and Compulsion
   5.3 Non-Ideal Repair of Dysfunctionality
6. THE IDEALS OF AGENCY AND THE PROBLEM OF AKRASIA
   6.1 Introduction
   6.2 The Possibility of Strict Akritic Action
   6.3 The Problem of Strict Akritic Action
   6.4 Davidson on Akrasia
   6.5 Against Davidson
   6.6 Against Mele
   6.7 Differences with Davidson and Mele over Desire and Motivation

BIBLIOGRAPHY
ACKNOWLEDGEMENT

I would like to thank Professor Coval for his dedication to a project that we have pursued together for several years and of which this thesis forms a part. His enthusiasm and encouragement made our work together a pleasure and its results much better than they otherwise could have been. Thanks also to Joe Naylor who contributed as only he can. Professors Patton, Jackson and Talmage gave the thesis a very careful reading and made several helpful comments for which I thank them. I would also like to acknowledge the financial support of the SSHRC Doctoral Fellowship Program and to thank the UBC Philosophy Department for their financial and moral support.
CHAPTER 1

THE IDEALS AND FAILURES OF AGENCY

1. INTRODUCTION

The subject of this essay is practical rationality and its failures, impracticality and irrationality. I will generate accounts of what may be ascribed to an agent when those processes of reasoning which, in agents sufficiently like us, are antecedent to, and explanatory of, their actions, are characterized by such general terms as 'practical,' 'impractical,' 'rational' and 'irrational.' Once the general concepts have been fixed, I will turn to the question of what is ascribed to agents when their agency processes are characterized by terms which pick out specific failures of agency, such as 'incontinence' and 'compulsion.' At the center of this subject is the process which, if there is to be action, as distinct from mere happenings, must have a causal and cognitive terminus in a state which is both cognitively and causally sufficient for action,\(^1\) a state, that is, which, when it is causal, has content sufficient to explain causation is terms of agency, and a state we have commonly named intention (or immediate or efficacious intention). Depending upon the

---

\(^1\)That intentions are causally and cognitively sufficient for action is argued for in *Agency In Action*, and will be more fully addressed below.
complexity of the agent,\(^2\) this process between and among states of agency may itself be more or less complex—may include such component processes as choice, deliberation, and planning, together with belief and desire formation and revision. Our question, then, will concern the conditions under which it is appropriate to characterize the components of agency as practical, rational, impractical and irrational.

There is considerable philosophical literature on agency and its failures. Most notably, perhaps, among the philosophical discussions of failures of agency, are discussions of such conceptually troublesome issues as the variously referred to problem of akrasia, weakness of will, or incontinence, and of compulsion, self-deception, and wishful thinking. It is not uncommon in the literature to find theorists dealing with the failures of agency individually, in a piecemeal fashion. My approach will be somewhat different: I will argue that one cannot get clear on the failures of agency without first getting clear on what counts as success. To that end, I will provide an account, in fundamental terms, of the cognitive and causal processes which produce an intention according to certain ideals we have for agency. With that account in hand, I will attempt to generate a systematic taxonomy of failures of agency. In such an account one would hope to find places for the common failures. But in addition, and perhaps surprisingly, the theory

\(^2\)Whether an action is preceded by these processes will depend in part on the complexity of the agent.
will indicate that there are more ways to fail to be a practical rational agent than perhaps we as yet have named in our fledgling theory of our own psychology. This attempt to systematize the failures of agency I call the “nosology of agency.”

2. ASSUMPTIONS AND METHODOLOGY

2.1 The Common-Sense Folk-Psychological Framework

The model of agency which is developed in chapters 2 and 4 will be described in the familiar terms of our common-sense folk-psychology. I intend this model of agency to be consistent with, but a considerable refinement of, that view of our psychology. The model employs causation as its fundamental relation which holds between mental states of a process which have the logical and functional properties of the folk-psychological items desire, belief and intention, which themselves are set within rational processes including intention formation, choice, deliberation and planning. In assuming this folk-psychological framework I intend not to beg any questions for or against any theory of mind, other than the fully eliminative theories. I take it that the explanatory power of folk-psychology is

---

As I will use the term, a nosology of agency is a systematic description and classification of “diseases” or failures of agency. The term is borrowed from David Stove, who, in The Plato Cult and Other Philosophical Follies, applied the term “nosology of thought” to the systematic and comprehensive theory of mistaken belief. Thanks to A.D. Irvine for the reference.

This account of intention formation is a modified version of the account to be found in chapters 2 and 3 of Agency In Action.
such that whatever the right theory of mind turns out to be, it will somehow accommodate the terms and relations of folk-psychology.®

2.2 What is Agency?

The subject of this essay is rationality and irrationality as they may affect agency. By ‘agency’ I will mean, roughly, that faculty or function of causal systems in virtue of which they may effect change in the world which is productive of, or is a means to the getting of, their goals. I will assume a causal theory of action and offer a causal theory of agency.

There may be some inclination to consider among the class of agents, not only we humans, but other more primitive animals as well, and more radically yet, perhaps some inclination to include plants and more primitive life-forms. The justification for extending the concept in this way will be, presumably, because all the above share the faculty, function or ability to change the world so that it accords with goals of the agent. Just how inclusive the concept agent is will naturally depend upon one’s model of agency, and whether we may properly speak of trees and amoeba, for example as having goals, acting in pursuit of those goals,

and as achieving their goals. I will not directly address this issue, although some of what will transpire here will have implications for even the most primitive agency systems: any system some states of which are goal-directed or teleological or purposeful, and which are causal of goal-directed behaviour, or action. Since my specific concerns are with practical reason and its failures, I limit the enquiry to those agents of which one may properly predicate 'rational' and 'irrational'. I will argue that this class includes only those agents for which belief plays an essential rôle in agency, that rôle to be articulated as we proceed.

2.3 The Intentional and Deliberative Conception of the Agent

An agent is a system with goals and the ability to cause those goals. Our topic, however, is narrower, because more complicated, than the successes and failures of agency simpliciter. We are concerned with the agency—that is, the goal-functional processes—of informational or cognitive systems. Furthermore, we are concerned with informational systems in which one important set of functions—the belief functions—determines the informational content, and thereby the functional rôle, of the other key states of its agency processes. That is, it would seem that any account of the agency of information processors such as agents like

---

6 Thus, in what follows, I will be understood to be using 'agent' in this, perhaps narrow, sense, unless otherwise indicated.

7 An account of the rôles which belief plays in agency is presented in chapters 2 and 4 below. See also, Agency In Action.
us would have to take into account the fact that, unlike primitive agents, we have the ability—certain states of ours have the function—to determine or control, to a greater or lesser extent, which active agency states, both cognitive and connative, we are in, and therefore to determine our behaviour, and, perhaps, to determine what sorts of agents we are. This view is sometimes referred to as “intentional realism,” the view that behaviour can be mediated by certain cognitive states, with the distinctive causal powers to play such rôles. It is this feature of ours which is expressed in the intuition that the rational agent is one in which Reason plays a decisive rôle in what the agent does and is inclined to do. This, in sharp distinction to the “primitive agent” who, in being without cognitive states which determine and control the content and function of his “active” or motivational states, is an agent who is a mere victim of the blind causal impress of non-cognitive nature, an agent whose active states are caused in him by the world, and thus an agent unable to cause himself to be the sort of agent he might like to be, since he entirely lacks that ability to affect his given motivational states. That we

---

have the ability to affect and effect the content and thereby the function of our agency states is one of those features which sets us apart from lesser agents: it renders us not merely passive with respect to the causal nexus we inhabit, and it is presupposed by any strong sense of choice, autonomy, deliberation, self-control, intentionality, responsibility, and punishment; presupposed, in fact, by all the concepts which comprise the logical space of rational agency.

I assume, then, a broadly Humean conception of the agent according to which we may distinguish two broad types of state in the ancestry of every action, distinguishable by their unique content and causal role, each of which is necessary for agency, and neither of which is alone sufficient. These are the states we have commonly given the names ‘desire’ and ‘belief’ and which will be discussed in detail below. This is to also assume an intentional and deliberative conception of the human agent according to which those states—desires—without which there

---


would be no action but which themselves are insufficient for rational action, may be "informed" by certain other states—beliefs—without which there would be no rational action; indeed, or so I will argue, without which there could be no rationality.

3. THE PRACTICAL AND THE RATIONAL

Intention formation—the process which, ideally, causes a state of the agent which is causally and cognitively sufficient for action—may be said to be practical, in one sense of the word, when it succeeds in producing such an ideal state, and impractical when it does not. In general, any constituent process of agency, whether it be intention formation itself, means-end reasoning, or belief revision, may be said to be practical when it functions so as to produce the goal it is in the service of, one which contributes to the overall practicality of the process which embeds it, and impractical when it does not. That part of intention formation distinguished as means-end reasoning—reasoning about how to cause some goal-event—is practical when it successfully identifies for the agent the event he may cause which has causal relations to his goal, and impractical when it misinforms the agent about how to cause his goal. In general, any process within agency, from the most complex, such as intention formation and planning, to the simplest, such

---

12 That beliefs "inform" desires is not an empty metaphor: what it means is addressed in chapter 2.
as the individual functions of desire and belief themselves, will be said to be “goal-functional” when it contributes to the overall “goal-getting” of the agent. On the contrary, any function which fails to perform its practical function will be said to be “non-goal-functional.”

In virtue of what, then, are the processes of agency said to be rational? It may seem as though there is nothing left of practical rationality to be uniquely rational if ‘practical’ means goal-functional. Under what conditions ought we to apply the predicate rational to a practical process? In the model of agency which is developed in chapters 2 and 4, the process of intention formation employs belief at every stage where information is added to, or subtracted from, the active or motivational states in the process. Thus the amenability of the motivational states to belief is presupposed by each step of intention formation. According to this model, we are agents who keep separate our epistemic and connative states, and therefore the ability of states of the first sort to inform states of the second is a precondition of goal-functionality. Without the ability of beliefs to affect and therefore control the motivational states, the agent could not be said even to have goals, since to have a goal is not merely to be in a state which is causal with respect to some state of affairs, but to be in a causal state—a desire—with respect to some state of affairs.

13 Just why we should have this functional bifurcation is an interesting evolutionary question. Perhaps in a world of constant change, a system which can respond selectively to its environment—a system which may act, that is—has an advantage over those systems which may merely re-act. See footnotes 5, 8, 10 and 11 above for references to the Humean two function theorists, and see Agency In Action, chapters 1 and 2.
to a state of affairs which one judges to be desirable. Furthermore, without information about the world, specifically about how to cause its goal, one's motivational states would be without information necessary for them to connect with the world so as to cause those goals. Finally, it is the ability of belief to determine what active states we are in, and how and when those active states themselves function, which gives substance to the concepts freedom and autonomy, among others.

Where rationality resides in agency is, I propose, in this informational dimension: rationality concerns the route that information takes from and to the world in agency. Thus the amenability of one's motivational states to belief is a presupposition of rational, and irrational, agency.

At the heart of agency itself is the concept of practicality, the essence of which is the getting of goals. Since our model of agency is causal, that getting of goals is the causing of goals. Thus a system qualifies as an agent if it can be said, non-metaphorically, to have goals, and to instantiate a causal process which is, or may be, causal or productive of those goals. A non-rational, primitive, agent would be one, the states of which lack those properties essential to rationality, such as cognitive or informational content, and one which lacks the agency function—belief—by which that content comes to be implicated in the cause and

---

14 See note 11 above on the deliberative conception of the agent.
15 Can we describe non-cognitive systems as having goals? If not, we cannot describe them as agents, if to be an agent requires having goals.
the explanation of its actions; implicated, that is, in the production of a state of the agent—an intention—which is both causally and cognitively sufficient for an action. If one were tempted to consider trees, for example, to be primitive, non-rational practical systems of this sort, it would be, presumably, because trees instantiate states which cause events which one is tempted to call tree goals. But if practicality is set within a system whose states instantiate properties such as cognitive content or information and which function in such processes as the construction of intentions and the selection of goals, then it is appropriate to apply the predicates rational and irrational according to whether the relations among those states and between the states and its goals, for example, conform to or violate some standards or ideals for systems so equipped. For an agent to be rational, as distinct from merely practical, presupposes that its agency states—its beliefs, desires and intentions—are amenable to belief. But simple amenability to belief is alone insufficient for an agency process to qualify as rational. We will say that an agency process or inference is rational just in case the implicated agency states were:

1. Amenable to belief, and
2. amended according to some belief or theory about how to be goal-functional, how to minimize non-goal-functionality; a belief or theory, that is, about how to be an agent.

Thus a practical inference will be subjectively rational when it produces a state which, by the agent’s own lights—i.e., according to its own theory of how to be goal-functional—will be goal-functional. A practical inference will be objectively
rational, then, when information is manipulated according to some objective standard or ideal for goal-functionality.

Practical irrationality occurs when the agency processes of cognitive or informational systems are not guided by relevant beliefs about how cognitive and causal states should function so as to be goal-functional. Thus a practical inference is irrational which fails to so pursue goal-functionality. Thus, on such a view, practical irrationality is a departure from those standards or ideals. Just what those standards or ideals are will be an open question and a matter for ongoing theorizing about the goal-functionality of agents. But despite the open-endedness of one's theory of practical rationality, there is a core which anchors all, perhaps competing, theories. That core is goal-functionality. This tells against those theorists who argue that rationality is a purely contingent matter.\(^{16}\) What conclusions and actions rationality may require of an agent will be a function of what content the agent contingently harbours; but no matter what content is given, rationality requires goal-functionality.

We may thus distinguish the following categories for a practical inference:

1. Practical and Rational: an inference which accords with the relevant theory about goal-functionality and which preserves goal-functionality.

\(^{16}\)The relativists argue that rationality is a purely contingent concept; that what counts as rational for you may not be rational for me, etc., especially if you and I are from different cultures, or the same culture but at different times, etc.. See R. Rorty, *Contingency, Irony and Solidarity*, (Cambridge, MA: Cambridge University Press), 48; John Dewey, “Changed Conceptions of Experience and Reason,” in *Reconstruction In Philosophy* (Henry Holt and Co., 1948), 77-102.
2. Impractical and Irrational: an inference which violates the relevant theory about goal-functionality and fails to preserve goal-functionality.

3. Rational and Impractical: an inference which accords with the theory but which fails to preserve goal-functionality.\(^{17}\)

4. Irrational and Practical: an inference which violates the theory about goal-functionality but which nevertheless preserves goal-functionality.\(^{18}\)

There is this sense, then, in which agency is "practical" when it is goal-functional, and impractical when it is not. But 'practical' has another, though not unrelated, meaning. An agent's chosen course of action is said to be impractical not only when that course is ineffective, and so fails to cause the goal, but also when, although it causes the goal, it does so inefficiently. This implies another set of goals about how the agent goes about satisfying his "actionable" goals; goals with respect to such considerations as efficiency, cost and return on investment of agency. Thus practicality has to do not only with the fact of goal-functionality, but also with how goal-functionality is achieved. From the agent's point of view, and other things being equal, it may be subjectively impractical in this latter sense for an agent to opt for a course of action which will get him his goal but at a higher cost than an alternative, more efficient, course he believes is open to him. Thus is brought within the ambit of our discussion of agency, practicality and rationality,

\(^{17}\) Note that rationality would require that such an occurrence occasion a revision of the theory of how to be goal-functional. This reveals the iterative nature of rationality which may demand adjustments to whatever it governs, whether that be intention formation itself, the theory of rational intention formation.

\(^{18}\) Again, in rationality, this calls for a revision of the theory.
those non-ideals of agency which include carelessness and recklessness. These will be addressed in chapter 5.

4. THE LOGICAL SPACE OF AGENCY

At the center of the concept of rational agency is the concept of a cognitive and causal state—an intention—which is practical when it causes what it represents; when, that is, it is goal-functional. There may be certain primitive agents in which such states are otherwise caused to occur. But for agents complex enough to be of interest to us here, intentions are produced by a process of intention formation in which belief is the preferred supplier of content or information. It is how belief functions in this capacity which is addressed, I have suggested, by our concept of rational practicality. When belief contributes to the production of a state which is causally and cognitively sufficient for action by supplying content under the guidance of a theory of how to be goal-functional, that process, we will say, is rational.

If one function of belief in rational practicality is to determine the informational content of an agent’s “active” or motivational states, then we may distinguish the following individual tasks involved in that enterprise, each of which will have a corresponding ideal in rationality. First, and perhaps simplest, is the supply of means to a previously existing goal. This is that part of practical reasoning we call means-end reasoning. The ideal for means-end reasoning would
be the supply of goal-functional content: i.e., true means beliefs. An agent in which reason extended only as far as the supply of means to previously existing goals, an agent, that is, where belief had only the function of the handmaid to desire,\(^{19}\) would be, perhaps, rational in only the thinnest sense.

But belief need not have its function absorbed solely in the supply of means to a previously existing goal. It would be an increase in rationality, and, perhaps, in autonomy, for an agent’s belief to have a hand in the selection of goals. So, rather than goals being self-selected as they are above, perhaps due to their individual causal strengths, an increase in (potential) rationality would be marked by belief having the additional “doorman’s” rôle\(^{20}\) of itself selecting, from among a set of desires, which of those desires would be goals of the agent—which desires, that is, would make their way through to intention—and which would remain as mere desires, despite their inherent strength.\(^{21}\) The ideal at work for goal-selection will again be the supply of goal-functional content, not now about how to cause one’s goals, but about what one’s goals, as opposed to one’s desires, are to be, a distinction marked by the presence of the content that an event or state of affairs is not merely desired but judged to be desirable. For goal-selection to be goal-functional is for belief to select from among the set of desires only those desires:

1. which are satisfiable by the agent, and

---

\(^{19}\) I have borrowed Hume’s metaphor.

\(^{20}\) This is to extend Hume’s metaphor.

\(^{21}\) This function of belief is described in detail in the coming chapters.
2. the satisfaction of which will not frustrate or preclude the satisfaction of desires which belief determines to be more desirable.

Perhaps the principle of rationality which captures the ideal for this belief function is “maximize goal-functionality,” or “minimize non-goal-functionality.”

We may imagine a further increase in the scope and range of the belief function in agency. Rather than being limited to selecting from among a set of given desires which desire will become a goal, it would mark a considerable increase in rationality—control by belief—an increase which takes us to autonomy, if belief had the further function of being able to create the content of desires. The ideal for this rôle of belief would be to produce a set of desires each member of which was satisfiable by the agent, compatible and consistent with other members of the set, and so on. Perhaps the fully rational agent, i.e., the fully autonomous agent, is the agent who harbours no content that was not supplied or ratified by belief—the “master” of all it surveys. The corresponding ideal governing this rôle of belief would perhaps be to supply or ratify all and only that content which would be productive of the highest-order goal of the agent, perhaps the goal of being the best agent possible, given the other, non-belief-amenable contingencies of himself and his circumstances.

---

22 Perhaps autonomy is the concept which concerns the ability of belief to determine content of an agent's agency states. The broader the scope of this belief function, the more autonomous the agent.
So we find a set of ideals for rational agency which are all anchored by the notion of goal-functionality. Thus we have the concepts agency, action, intentionality, practicality, rationality, and autonomy. In addition to the ideals, we have a set of concepts which express certain non-ideals of agency. Among the set of concepts we have distinguished for significant departures from the ideals of agency, there are the general negative concepts unintentionality, impracticality, irrationality, and non-autonomy or heteronomy. In addition to the general non-ideals of agency, there are those concepts which express particular species of non-ideal agency, each falling under their appropriate genera. Among the species of non-ideal agency concepts are the following: mistake, accident, inadvertence, incontinence, compulsion, impulsion, weakness of will, carelessness, recklessness, wantonness, coercion, and necessity.

Given the centrality of the role played by belief in agency, we may expect that, in addition to the specific non-ideals of agency or intention formation, there will be a set of concepts which express non-ideals of belief formation. Since belief formation will involve the same functions as intention formation, there will be non-ideals of belief formation corresponding to the non-ideals of intention formation set out above. Among the non-ideals of belief formation there are those we call wishful thinking and self-deception.

23 See Agency In Action, chapter 4, for accounts of action, mistake, accident and inadvertence.

24 Necessity is the legal exculpatory concept.
Chapter 1

My project in this essay is to show how the concepts for the ideals and the non-ideals of agency are related. I will begin to unfold those relations in chapter 3. But first, it ought to be clear that one could get no purchase on the nature of the non-ideals of agency without first getting clear on the ideals. To that end I present, in the next chapter, a portion of a theory of rational agency in which the basic and constitutive functions of rational agency will be seen at work. It is perhaps a test of an account of practical rationality how well it explicates the failures of practical rationality. Once the basic and constitutive functions of rational agency are isolated and before us, then, I will begin to develop from them a taxonomy for all the functional failures. With that taxonomy in hand I will proceed to set the basic agency-functions within a richer subject—an agent with beliefs about himself and the world he inhabits, and with a corresponding theory of how to be an agent so circumstanced. With that additional complexity, and armed with the taxonomy of agency failures, we will begin to see what explains how agents such as you and I may fail to be the agents we might otherwise be.

This project is programmatic. I develop a theory of practical rational agency and generate a taxonomy of its failures. I then apply these accounts to specific cases of ideal and non-ideal agency, including *akrasia*. 

CHAPTER 2

THE TERMS OF AGENCY

1. INTRODUCTION

Our classificatory scheme for agency consists of concepts which we have devised for the purpose of, among other things, explaining ourselves to one another, qua agents. In the previous section I argued that the agency concepts classify the phenomena of agency according to certain ideals we have for behaviour of agents such as ourselves. That set of useful concepts includes terms for what we have deemed standards or ideals for agents, as well as terms for the common departures from those ideals. At the center of this set of concepts is action—the ideal of behaviour which is just as the agent intended: that is, behaviour which is goal-functional. On the other hand, there is a set of non-ideals for behaviour which fails; for behaviour, that is, which is non-goal-functional. The action failures include the “unintentionalities” mistake and accident. A mistake is a case of an intention failing to produce the goal due to a false means belief. An accident is a
case in which an intention is prevented from causing the goal by some intervening event.¹

Among the mental agency concepts there is intention which is a state which, ceteris paribus, is causal and explanatory of action. If intention is the ideal, then we must distinguish intentions from those states which are otherwise similar, but which fail to produce what is intended. For this rôle we have nominated misintention.² There is an ideal for the state which plays the rôle of supplying goal-functional information to the agency process, a rôle here played by belief, and there are non-ideals for such states, including doubts, and wonderings, and those states we mention when we speak of someone as "not really believing" that such and so is the case.³ There is an ideal for the process which produces an intention—practical rationality—and there are those terms which indicate that the process has deviated from an ideal, terms which include incontinence, impulsiveness, and compulsion. Intention and intention formation are the subject of chapter 4, and their non-ideals are the subject of chapter 5. Finally, we have a term which expresses an ideal for the source of the content of one's agency states.

¹See Agency In Action, chapters 1 and 4, for a full account of action and non-ideal action. An inadvertence is a non-ideal case in which something, typically untoward, is caused in addition to what was intended.

²See Agency In Action.

³To say that someone "lacks the conviction of his beliefs" is to say that what he should believe, perhaps on the basis of his other beliefs, he does not; that what should function as a belief for him does not. Contrast this with someone who "lacks the courage of his convictions." This is someone who, despite having the conviction or belief, does not behave accordingly.
Autonomy expresses the ideal of the agent who is himself the source of his desires, in contrast with heteronomy which indicates some external source of content. Coercion is a case in which content and motivation have their source in some other agent.

The ideals and non-ideals of agency mark distinctions in the processes and states of agency and the effects which agency has in the world. Just what distinctions are marked we are in the processing of addressing. An analysis of the non-ideals presupposes an account of the ideals, and so clarification of the non-ideals must wait until we have an account of the ideals.

Rational agency, on this view, amounts to the processes of agency functioning according to an ideal we have for agency systems like us. Irrational agency, or practical irrationality, will amount to various departures from those ideals. Among the ways in which agency can go astray, there will be non-standard states and non-standard processes, and corresponding to these will be terms for common departures from our ideals which adjust the ideal states and processes of agency. Thus we find a term like ‘compulsion,’ which adjusts a certain state integral to a certain ideal of agency and therefore indicates that that state is non-ideal in a specific way. And we also find process adjuster terms like ‘compulsively’ which indicate that a compulsion has affected some process of agency.

Compulsion, incontinence and impulsiveness, are functional failures of specific sorts. Each is a distortion or aberration of a state or process we judge to
be ideal for agency. 'Incontinence,' 'compulsiveness,' 'coercion' and 'necessity' each indicates a unique way in which relevant processes of agency has deviated from an ideal for that process. Each of the agency deviations thus has a signature, as it were, which we have marked with a term of non-standardness. So an account of the standard functions will serve as the basis for accounts of the failures. This enquiry will therefore be organized around the fundamentals and ideals of the processes of agency. The more articulated an account of the functions integral to agency one has, the better equipped one will be to clarify the nature and relations of the agency concepts and therefore also the better off one will be in understanding problematic behaviour such as incontinence and self-deception. With that assumed, I turn first to an account of the fundamental terms of the processes which are antecedent to action in agency systems such as ourselves. Once the fundamentals of agency are before us, we will address the ways in which agency may depart from those ideals. The refinements of the model of agency, which begins its development in this chapter, and is completed in chapter 4, will be exploited in the later analysis of the non-ideals of agency.

In this chapter I describe a portion of practical reasoning so as to isolate the fundamentals of agency. These fundamental terms and relations can be held to a standard or ideal—goal-functionality—and will be described as meeting that ideal. Once described, their non-ideal, non-goal-functional, forms will be addressed.
Agency is a cognitive and causal process. The culmination of agency is, ideally, action. The state penultimate to action is a state which is the terminus of the cognitive and causal process antecedent to action, and is a state which, if it is to play its rôle in agency, must be sufficient, both causally and cognitively, for action. The states of agency fall within types in virtue of their content and the causal relations into which they may enter. The attitudinal states central to agency are desire, belief and intention, with secondary rôles performed by likes, doubts, hopes and wishes. The agency states themselves may be considered points within the agency processes, and so the complete set of their structural and relational properties define the rational process of agency.

2. THE FUNDAMENTAL TERMS OF AGENCY

2.1 Causation and Its Terms

Agency requires change, and causation is the explanatory relation of change. Therefore, causation will be the fundamental relation of this model of agency: it is a causal theory of agency. The relation of causation requires datable, locatable causal particulars as its relata, and the rôle of the relata of causation will be played here by events. 'Event' is used here in a broad enough sense so as to include changes (and lack of change) over time, states and processes. If events

It is plausible to consider a change of properties with spatial and temporal dimension to be an event. Similarly, a set of properties enduring at a place over
are to be particulars, they require spatio-temporal properties. One of an event's temporal properties is duration, and so events are the sort of item which may exist over time and endure either change of properties or constancy of properties. A state I understand to be an event described in terms of its properties. That is, a state is a time-slice of an event, a set of properties at a place at a time, a point along the space-time line that is the event, if you will. Or, since states too may endure, a state is some relevant or pertinent properties of an event which endure over time, a set of properties at a place over time, or, a segment of the event-line over which some properties remain constant.

2.2 The Representing and the Represented: The Types of Particulars Required for Agency

While we may assume that the relata of causation are sheer events, agency causation requires a cognitive particular which is capable of causation. That is, for agency we need a particular for the mental items integral to agency which can stand in both causal and intentional or representational relations. I here assume time may be an event; and so a set of properties at a place momentarily. Thus I want the following to qualify as causal relata: the mental states of desiring to F and believing that P; the mental process of coming to intend to F. Including processes as events is to accept J.J. Thomson's view of events. See her Actions And Events.

*I do not intend by the term 'agency causation' what others, such as Irving Thalberg, have, namely a sui generis causation. See Thalberg, "Do We Cause Our Own Actions?", Analysis 27 (1967): 196-201; R. Chisholm, "Freedom and Action," in Freedom and Determinism, ed. K. Lehrer (New York: Random House, 1966), 28-44.

*I don't want to assume that representation is necessary for agency, but cognition is.
that event, broadly enough construed, can subsume the mental items integral to agency as well as bodily events implicated in action. The mental items include states, such as the belief that P, the desire to F and the intention to G, and mental processes such as coming to intend to G and means-end reasoning about how to G.

The fundamental relations and terms of agency are:

1. Causation
2. Events of two types:
   i. mental events which are cognitive, or informational or representational, and which may stand in causal and intentional/representational relations to one another and to,
   ii. non-mental, non-cognitive worldly events.

The account of agency which follows is an information-theoretic account of the fundamentals of agency. That is, the rôle of the constituents of agency will be described in terms of the formation and manipulation of information. But since any such account would be required to be naturalizable, what I will describe in information theoretic terms is actually causation. It is an account of causation in terms of the passing of information from one state, the cause, to another, its effect. If the primary function of an intention is to cause certain events which are, ceteris paribus, actions, that relation is describable in information theoretic terms as the intention passing certain information to the world. Similarly, the belief function in this model is causal. That is, beliefs are states which are affected in a certain way by the world, and have certain causal effects of their own, most notably for agency, certain effects on desires and other allied states of agency. These causal rôles too have an information theoretic translation: beliefs pick up information
from the world and, in agency, pass that information on to an active attitude such as desire. I intend nothing I say here about the functions of belief and desire to beg any questions for or against any theories of belief and desire, except insofar as to claim that whatever beliefs and desires turn out to be, in order for agency to be viable, these functions, which I am about to describe, must be present.

2.3 Distinctions Among The Mental Terms of Agency

The fundamentals of our model include two separable information-passing functions which do the work of belief and desire, which, however, do different things with their content. The active attitudes, in “passing” content, cause what they represent. The cognitive attitudes merely cause their content to be passed, or, produce change in other mental states, while the active attitudes actually produce other states, mental and otherwise.

For agency we need motivation, and for motivation we need causation which is “directed at” an object. So, for agency we need a causal and cognitive state which is causal of what it represents. States of this type we call active attitudes, whose species include desires and intentions. For rational agency, agency “informed” by reason, we need information which is put in the service of the motivational states. That is we need an informational state with the function to inform active attitudes. That rôle is here performed by states of the type epistemic attitude, a type which includes belief.
The category to which a mental state belongs is a function of its causal rôles and the structure of its content. The active attitudes, which include desires and intentions, wishes, hopes and compulsions, have one set of causal rôles and their content is structured appropriately for their rôles: the content of an active attitude represents, or is “about,” a change or a causing, and the attitudes themselves are causal with respect to what they represent or are about. The epistemic attitudes, which include beliefs, doubts and wonderings, have another set of causal rôles and their content is propositional in structure: they are “about” states of affairs, and they represent what caused them. The functional differences within a category will be due to structural differences in the content of those states. Although desires and intentions share generic causal rôles, the actual relations into which the two may enter differ in virtue of their respective cognitive structure: an intention is a more rationalized attitude than a desire and the effect of that rationalization is to give intentions causal rôles not open to mere desires. In our theory, intentions have certain evaluative content as the result of

7 Myles Brand, in *Intending And Acting*, 85 ff., discusses two general methods for distinguishing among the attitudes. According to what he calls “object assimilationism,” the attitudes all take objects of the same type and differ in virtue of their differing attitudes (causal rôles?) to their objects. Conversely, “attitude assimilationism” is the view that the attitudes of the various mental states are all of a single type and the mental states differ, rather, in virtue of taking type-distinct objects. In our model, the mental states differ in virtue of their causal function and the structure of their content which is suited to their causal function.

judgement, and it is this additional content which makes intentions fit for rôles for which mere desires are unsuited. Similarly, although beliefs and doubtful beliefs share generic causal rôles, the actual relations into which a belief may enter differ from those available to a doubtful belief in virtue of their respective cognitive structure: belief, like intention, has a certain structural feature—a certain evaluative component—which doubtful beliefs lacks, the effect of which difference is that standard beliefs and doubtful beliefs have different causal rôles. What this cognitive content and structure is, and how it may be affected, is discussed below.

3. THE STATES OF PRACTICAL RATIONAL AGENCY AND THEIR FUNCTIONS

3.1 The Structure and Function of the Active Attitudes: The Fundamental Practical State

We cannot make sense of the notions of action and agency without a mental particular which is capable of cognitive as well as causal properties. That is, for agency we need a particular for the mental items integral to agency which can stand in both causal and intentional or representational relations to their causes and effects. The causal function of the active attitudes will be represented by ‘→’, the causal arrow. The content of the attitudes will be represented in terms of events and their properties. Events will be represented with the capital letter E and numerical subscripts: thus, e.g., E₁. Properties will be represented by capital letters F, G, H and so on, placed “above” the event of which it is a property and
separated from its symbol by a horizontal or diagonal line, not unlike the arithmetic device by which numerators and denominators are separated. For example, ‘F/E₁’ represents the event, E₁, with property F. To indicate that ‘F/E₁’ is the content of an active attitude, the causal and representational symbols will be combined as: —F/E₁→. This is our notation for the state which is causal, represented by the arrow (— →), with respect to a description, F, of an event or state of affairs, E₁.

What common-sense folk-psychological state might —F/E₁→ be, if any? Although it has the causal function, represented by ‘→’, which all active attitudes have, it cannot be a desire due to the (simple) nature of its cognitive structure. Desires are states which are capable of causing states of affairs beyond the mind, which causings they represent. The causal object of a desire is, then, the coming to be or causing which it represents. For example, the desire to embarrass the government is a state causal with respect to the coming to be of the state of affairs of the government being embarrassed, and not simply causal with respect to the state of affairs of its being embarrassed. The key difference between fully-fledged desires and the simpler state, —F/E₁→, is that desires are “for” causings

---

⁹ Read “F of E₁.” I here adapt the arithmetical convention of expressing a ratio with a diagonal rather than a horizontal line to my purposes.

¹⁰ Actually, as will be pointed out below, this account of the nature of desire-to is inadequate in one very important respect: desires-to are states which are causal with respect to causings of a certain sort—namely, agency causings—and not just causings simpliciter. States which are causal with respect to causings of any sort are desires-that. More on this below.
rather than for simple states of affairs, and this difference is marked in the notation by the symbol for coming to be or causing: \( \text{ww} \rightarrow \), "which would cause." If the state of affairs of the government being embarrassed is represented as \( F/E_1 \), the desire to embarrass it would be:

\[
\begin{array}{c}
\text{D1} \\
\text{ww} \rightarrow \\
\frac{F}{E_1} \\
\frac{E_x}{x}
\end{array}
\]

which is a state of causation, \( \rightarrow \), with respect to an event \( E_x \), which has the property of being the event which would cause (\( \text{ww} \rightarrow \)) the desired state of affairs \( F/E_1 \).

Desires, like all informational states, structurally encode the information contained in their that-, or to-, clauses. The structurally encoded information in the desire-to represented above describes a causing of a state of affairs. That description is incomplete in a way which is of central importance to practical/means-end reasoning: it describes a causing of \( F/E_1 \). But before the desire could be successful, i.e., before it could cause \( F/E_1 \) in any way other than serendipitously, it must describe a particular causing. That is, the incomplete, indefinite description must be completed—i.e., made to identify the causing. The occurrence, within a practical rational system, of a state with an incomplete desire-to structure like that of D1 above, poses the question which it is the function of practical/means-end reasoning to answer; namely, how to cause the
object state of affairs; or, in other words, what is the identity of the $E_x$ which would cause $F/E_1$. Only states with this structure, then, can be inputs to practical reasoning, the function of which is to provide descriptions of the object at which the desire-to is directed which are sufficiently definite to maximize success. That is to say, only states with the structure of a desire-to are fully practical states.

A desire, then, in being a practical state, is an attitudinal or functional state which is informed in its structure by the belief that causation is necessary to make actual any non-actual and causable state of affairs. A desire encodes a representation, $F$, of its causal and intentional object, $E_1$, the goal-event of the desire, which functions in means-end reasoning to initiate identification of the event $E_x$, which is the cause or change in the world believed necessary and/or sufficient to produce the goal-event. Identifying $E_x$ by accumulating sufficiently identificatory descriptions of a cause of the goal which is available to the agent is one of the central functions of practical reasoning. Furthermore, only with this causal structure of the content of a desire, and the subsequent function of desire in practical reasoning, is the language of agency, practicality, teleology and means-end reasoning appropriate. Only an active attitude which is causal with respect to a causing of an event, the representation of which is incomplete in just this way, and which is completable by belief, poses the question which it is the function of

\[11\] The belief that the goal is non-actual and causable is a rationality condition on a desire proceeding in practical reasoning. See the next section for more on this.
means-end reasoning to answer. Such incompleteness is an incompleteness of the appropriate relational or functional description of the event of which the agent is to be causal.

3.2 The Rational Origins of Desires-To

It would seem from the above that a desire, such as D1 above, is not the simple state it may have appeared to be. The complexity of the structure of a desire which provides the rational framework for practical reasoning, suggests that that structure is the result, or perhaps the conclusion, of a fundamental piece of "pre-practical" reasoning. The pre-practical syllogism, were it to yield a state with the structure of D1 as conclusion, would consist of premises which themselves ought to be discoverable in their effect, the desire structure itself. First, there is that state which accounts for the agent's motivation toward that goal-event: a causal state with respect to F/E₁, tout court, rather than to its causing, which we represented above as —F/E₁→. When this "proto-desire" is combined in rationality with the belief that causation is necessary to make actual any non-actual causable, causation shifts, as it must, from the object F/E₁ of the antecedent proto-desire to the object Eₓ which it is believed would cause F/E₁. The content of a

---

¹²While all of the states in the processes of agency to be described in this and chapter 4 are conclusory, there is nothing here to prevent a practical state from being parachuted into an agent: nothing, that is, to prevent an agent simply finding himself in a desire-to state, for example, due to some non-rational process. Perhaps agents can be simply caused to have fully fledged desires—states which are ready-made for practical reasoning.
desire has this form: it is causal with respect to a causing. Thus the minimal
description of the object of a desire is: “that event with the causal relations to the
object of the proto-desire.”

We have thus excogitated two states which are causally and rationally prior
to a desire that, in our example, the government be embarrassed. The first is the
attitude or causal state whose object is the state of affairs of the government being
embarrassed: \(-F/E_1\). Such a cognitive and causal state might occur in a
practical rational system. However, a system which remained in this proto-
practical state, rather than performing the conversion outlined above, would be in
a causal state, but would be without the conception of means; that is, such a
system would be an “agent” without any way of representing how. The second
state we excogitated from the structure of a desire, the belief in the necessity of
causation, is necessary for the question of means to come forward for the agent.
But before this belief can have the effects in agency we have described, the agent
must also believe that the object state of affairs, F/E_1, is both non-actual and
causable. When this is so, causation, which had been attached to the
representation of the desired state of affairs, F/E_1, shifts to the means to that
object, the coming to be, or the causing, of F/E_1. This new object of the attitude is
the event E_x which would cause the event F/E_1. The rationally and causally
subsequent state is, then, the desire that F/E_1 come to be:

\[ \text{The belief that causation is necessary to make actual the non-actual and}
\text{causable.} \]
3.3 The Pre-Practical Active Attitude: Desire's Ancestor

The presence of such a simple causal and representational state as $\neg F/E_1 \rightarrow$ was excogitated from what the notion of action requires as antecedent. The state of being causal with respect to $F/E_1$, the state $\neg F/E_1 \rightarrow$, is a possible state of an agent, but it is importantly devoid of the content which the basic belief in causation would add in the derivation of the fully fledged desire. $\neg F/E_1 \rightarrow$ is a state which does not represent the world as in need of change or causation in order for $F/E_1$ to come to be. The content of $\neg F/E_1 \rightarrow$ is not an adequate representation of the sort of world to which agency must connect, since $\neg F/E_1 \rightarrow$ is comprised only of causation and a representation of some state of the world. Such a simple representation lacks the capacity to accommodate an essential feature of agency: that the means to the desired state of affairs must be addressed. Only beliefs about the causal structure of the world, and states which are open or susceptible to such content, allow that essential connection between goal and means to be made.

$\neg F/E_1 \rightarrow$ is a state whose representation is blind to the causal structure of the world, as yet incapable of accepting relevant causal beliefs. If an agent were to affect the world on the basis of such a state alone, and were to (serendipitously?) cause $F/E_1$, it would do so as motivationally as one billiard ball moves another:
that is, cause it wherein the representation of F/E₁ plays no rôle. In order for “F/E₁” to play a rôle in causation, it must be part of a causal representation, such as that of the D₁ above.

The simple state, —F/E₁→, being a state of pure nascency, is not, therefore, capable of entering the process of practical reasoning which searches for means to its goal. It is a state not yet structured to accept the effects of practical beliefs about how to cause F/E₁. It would be a state, as we shall see, also incapable of satisfaction, not because it lacks the representation which would be supplied by the relevant causal beliefs, but because it lacks the representation to which the belief that F/E₁ is actual, the belief which functions essentially in desire satisfaction, could be relevant. It is a state which, although a prelude to action, is not yet engaged in the rationality of that enterprise. This is not to say that —F/E₁→ could not have its structure altered or its content added to so as to render it fit for practicality; but rather that as it is, it is not yet a state capable of such rationalization. It is, that it is to say, a pre-practical state.
4. THE PRACTICAL FUNCTIONS OF BELIEF

4.1 The Belief Function in Content Concatenation

Practical reasoning is a cognitive process by which a desire’s incomplete representation of its intentional object is completed. That is, it is a process whereby an insufficient description of the event which would have causal relations to the goal-event becomes sufficient. In order for the identificatory content of a desire to be sufficiently completable, desires must be states which are capable of being affected in just this way by beliefs. Described in information theoretic terms, desires must be capable of taking on information or content from beliefs. It is just this function of being transformable by beliefs, or, alternatively, the ability to take on content from belief, that defines the relationship of desires to practical reasoning: desires can accept and store practically relevant information. Desires are in this way transformable by beliefs. It is desire’s impressibility by impressive beliefs which is the presupposition of rationality in agency.

---

14 The determination of the goals which our active attitudes will incorporate, *viz.* normative reasoning, is obviously important to practical reasoning. Rational practicality, as we know it, cannot proceed without normative beliefs. In this essay, however, we are not engaged in a theory of belief formation, either normative or factual. Although this is not an essay in epistemology or moral theory, its concerns do include the identification of the types of belief which a practical reasoner would use, since the main concern here is to identify the states, relations and processes which comprise such a reasoner.

15 Failures of those functions of desire and belief which are presupposed by the transformability of the one by the other, functions which include impressibility and expressiveness, mark important junctures in our account of impracticality and irrationality.
If practical reasoning were to take as input the simple state \( F/E_1 \rightarrow \), and if practical reasoning is to be practical, the belief that causation is necessary to make actual any non-actual causable must impress that content (or otherwise cause the desire) upon the active attitude \( F/E_1 \rightarrow \), thereby transforming that simple state into the complex state

\[
\begin{array}{c}
\text{ww} \rightarrow \frac{F}{E_1} \\
E_x \\
\end{array}
\]

in which the change in structure and the additional content is due to the belief that causation is necessary to make actual the non-actual and causable goal-event \( F/E_1 \). The occurrence of such a state, in which the causal object is as yet unidentified (indicated here by the ‘x’ of \( E_x \)), sets the task for means-end reasoning of providing a more complete identifying description of \( E_x \): that event is as yet described only as an event with causal relations to the motivating event \( F/E_1 \).

Suppose an agent desires to kill the King, and that we represent the goal-event of the King’s death as the event, \( E_1 \), of the King’s death, \( F \). If so, the agent, in the earliest stages of practical reasoning, will desire an event which is a killing of the King. Suppose further that our agent believes that the poison being poured in the King’s ear, \( G/E_2 \), would be sufficient and/or necessary in the circumstances to cause the death of the King. If our agent is practically rational, the effect of this belief will be to transform the antecedent desire for the death of the King into the desire for an event under a new and more particular description: the effect of this
means belief is to transform the agent's desire into a desire for an event, $E_x$, with causal relations, $\mathbf{ww} \rightarrow$, to the, $E_2$, poison being poured in the King's ear, $G$, which in turn has causal relations to the $E_1$, death of the King. We may represent this state which is a conclusion of means-end reasoning as:

\[
\begin{array}{c}
\text{G} \\
\text{ww} \\
\text{E} \\
\text{E} \\
\text{E}_x
\end{array}
\]

This is a desire for the event $E_x$, which would cause the event $E_2$, of the poison being poured in the King's ear $G$, which would cause the motivating event $E_1$, of the death of the King $F$. Thus, the concatenation of content and structure in desire is due to the informing function of belief in practical reasoning.

One of desire's functions which makes desire suitable for rational practicality, then, is to accept and store or concatenate the content of the beliefs which inform and affect them in practical rational systems. Another practical function of desire is to cause certain beliefs. The desire for the causing of $F/E_1$ could or would, in a practical rational system, cause the belief that the agent so desires. The representational content of this belief would be that of the desire. We have represented that desire as:

\[
\begin{array}{c}
\text{ww} \\
\text{F} \\
\text{E}_1 \\
\text{E}_x
\end{array}
\]
This structure will form the representational content of the belief. To complete our notation for the belief state itself, we must add our symbol for the belief attitude or function.

In this model of practical rational agency, beliefs function causally, and their rôle in agency is limited to affecting the content of desires and other beliefs by passing or transferring information to them. Thus, with Hume, beliefs here lack the capacity to affect the (rest of the) body directly. Beliefs differ from desires not only in the form of their content, but also in their function. We represent the belief attitude or function as causation, just as we did for desire, but mark the functional difference with an arrow directed downward, as it were, upon states within the system and not at anything external to it, thus: \[\downarrow\]. We intend this to indicate that the causal rôle of belief is entirely absorbed in the transmission of content to other internal states of the system.

To complete our representation of the belief that the agent desires the causing of \(F/E_1\), we add our symbol for the belief function at the place in the above structure of that belief's content which is analogous to the place of the symbol for the desire function in our representation of desire:

\[
\begin{align*}
ww &\rightarrow \frac{F}{E_1} \rightarrow E_x \downarrow
\end{align*}
\]

To complete our representation of belief, we must add a symbol for the causal object of the belief: the object upon which it will function. If beliefs are causal and
transfer the representation they contain to their causal objects, then those objects, as distinguished from their contents, are the states to which beliefs may cause their content to be passed. We may suppose that the causal object of this belief state is some other, as yet unspecified, state of the system, $E_y$. If so, the completed structure of the belief that the system desires the causing of $F/E_1$ is:

\[
\begin{array}{c}
F \\
\overrightarrow{ww} \rightarrow E_1 \\
\overrightarrow{E_x} \rightarrow E_y
\end{array}
\]

The belief state bears this entire structure. It is a state which is causal with respect to the transferral of some representation to another object in the system. If truth and causation are to be co-preserved, then the causal object of the belief ($E_y$) must itself be an object capable of and susceptible to the amalgam of efficacy and cognitivity. Thus, the causal objects of beliefs are other intentional or attitudinal states.

The representational content of belief is derivable from both non-cognitive and cognitive states of affairs, from the physical as well as the mental. And although the content of a belief may be caused by the non-cognitive physical, they apparently cannot reciprocate in kind. The psychology of these matters embedded in natural discourse tells us that we have cognitively efficacious states with quite different causal capacities. States of one sort—beliefs—have their function absorbed in affecting the content and structure, and therefore the function, of the
states of the other main sort—desires. A desire is capable of causing change in the physical world according to its representational content. That content is affectable by belief. These capacities are clear and familiar and close to the surface of common sense, folk-psychological discourse. However, as we proceed with this account of the antecedents of action, subtleties will emerge, as they should, which are not visible from the surface.

4.2 Content Decatenation

We have so far identified, from among belief's repertoire, the cognitive function of "informing" an active attitude, such as desire, by supplying practically relevant content—i.e., content relevant to causing the object of that active attitude—which is concatenated in the cognitive structure of the active attitude. This is one of the ways beliefs transform active attitudes. It was also mentioned above\(^{16}\) that belief is implicated in the (cognitive) satisfaction of a desire. In order for this to be so, beliefs must have another function besides the addition of content. For suppose that an agent desires the coming to be of F/E\(_1\). In rationality, we argued, this state presupposes the belief that F/E\(_1\) is both non-actual and causable.\(^{17}\) Now suppose our agent comes to believe that F/E\(_1\) is actual. The content of this belief contradicts content already embedded in, or presupposed by, the desire that F/E\(_1\)

\(^{16}\)See section 3.3 above.

\(^{17}\)A plausible weaker rationality condition is the absence of the belief that the F/E\(_1\) is actual and/or non-causable.
come to be; namely, the need for a cause of $F/E_1$. If our agent is practically rational, the function of this newly acquired belief is to "decatenate" the appropriate content from the desire that $F/E_1$ come to be, the effect of which is that causation is no longer directed at the coming to be of $F/E_1$. That is, the effect of this belief is to remove the "which would cause" clause from the desire. In our notation, that clause comprises:

$$\text{ww} \rightarrow \ldots \quad \frac{E_x}{E_x}$$

The deleted or decatenated content is that causal description of $E_x$. The result is an active attitude which lacks a representation of the event $E_x$ which was believed causally necessary and/or sufficient for $F/E_1$. That representation, it will be recalled, had been derived from the practically presuppositional belief in the necessity of causation to make actual any non-actual causable, together with the belief that $F/E_1$ was both non-actual and causable. The effect of the addition of this content was to transform the antecedent desire for $F/E_1$ into a desire for the causing of $F/E_1$, the causal object of which was the event, $E_x$, which it was believed would cause $F/E_1$. With that content now deleted from the desire for the causing of $F/E_1$, the active attitude may stay with the heretofore desired state of affairs, $F/E_1$, and comprise the pre-practical state $-F/E_1 \rightarrow$, the proto-desire for $F/E_1$, or perhaps the mere liking of $F/E_1$, an attitude unimpressed by the belief in the necessity of causation to realize the favoured state of affairs.
In this way, then, the belief that the goal-event of an active attitude is actual deactivates that attitude, renders it non-causal with respect to a causing, by removing the content which is essential to the practical process which leads to action. This "de-effication" is an essential component of desire satisfaction. For the belief that the goal-event of an active attitude is actual to fail to function so as to shut off the pursuit of that goal-event, so that the attitude proceeds through the practical process toward action, we have a form of irrationality, perhaps a form of obsessive behaviour. This, like the other irrationalities, categorizes various aberrant relations between desires and beliefs, between active and epistemic attitudes, and therefore also between our cognitive and causal states and the rest of the world.

4.3 The Practical Effects of Certain Beliefs

A belief is an item which may play several practical rôles, each of which may be seen as a distinguishable but related causal function in the process which may or may not culminate in action. A belief may interact with a desire in practical means-end reasoning where the function of the belief is to cause its content to be added to that desire, thereby further identifying the causal and intentional object of the desire, in what amounts to an increase in rationality. One condition of rationality is that only relevant content is so added. In practical reasoning, there

\[18\] See chapters 3 and 5 for more on obsession.
is an underlying question at work which determines which content is relevant. The first and fundamental question which is posed for practical reason by the occurrence in the system of a desire is about which event or events are believed to have causal relations to the goal-event. This being the practical question at this stage of practical reasoning, only beliefs with this content are relevant. Should such a belief function in the way described above and add its practically relevant content to the active attitude, that addition is enabling of the desire, providing, as it does, information necessary for the getting of the goal. Beliefs thus have this enabling function, while desires have a motivational function.

Beliefs also have a disabling function, as outlined above in the section on decatenation of content. That is, certain beliefs can disable or stop a desire state from proceeding through the process to action. Imagine an agent with a proto-desire for the end of the war, \( \rightarrow W/E_1 \rightarrow \). Suppose further that he believes that causation is necessary to make actual every non-actual causable. This belief, together with the beliefs that \( W/E_1 \) is both non-actual and causable, transforms the proto-desire into an active attitude with respect to the event \( E_x \) which would cause \( W/E_1 \):

\[
\begin{array}{c}
\text{ww} \\
\text{W} \\
E_1 \\
E_x
\end{array}
\]
Now, an agent may have four typical beliefs about the likelihood of causing \( W/E_1 \); that is, four beliefs about the likelihood of there being an event \( E_x \) which would cause \( W/E_1 \). He may believe that:

1. it is impossible, nomologically or logically;
2. it is possible but improbable;
3. it is already actual;
4. it is causable and non-actual.

We will suppose that our agent has each of these in turn and describe their effects on agency.

1. If we accept that desires concatenate, or take into their structure, the information from relevant beliefs, then when impressed by the belief that causing \( W/E_1 \) is impossible, the above desire will be converted to a state which is causal with respect to a state of affairs the causing of which is believed impossible. On our model of agency, this is a wish for \( W/E_1 \). In rational practical systems, wishes do not function as do desires; do not, that is, function as motives to action. Wishes do not have the structure which allows the right practical questions to be asked; namely ‘What is a further specification of the causing of \( W/E_1 \)?’ The structure of a wish indicates that the system contains the belief that causing \( W/E_1 \) is impossible, and this belief is incompatible with beliefs about how to cause \( W/E_1 \). It is a condition of rationality that a system could not be in a state of assent to both beliefs. It is a further condition of practical rationality that a wish is shunted aside, off the practical reasoning rails, as it were, and “parked” in such a way that its occurrence does not
prompt the system to engage in practical reasoning to service it, the way a desire prompts a practical system. A wish is not an irrational state but is an impractical one, since it cannot function in practical rationality.

2. If an agent believes that causing what he desires is possible but improbable, the effect of this belief on the desire for the causing of his goal-event should be to transform it into a mere hope for that state of affairs, a state which is impractical, and therefore not a proper input to practical reasoning, but one which, unlike mere wish, is more likely to have its content revised in light of one's beliefs about the causability of the motivating state of affairs.

3. The belief that the motivating state of affairs is actual ought to disable an active attitude, since practical reasoning would be otiose. This belief may function as a premise to desire satisfaction, which was sketched above, typically when the desired state of affairs is caused by the agent. But that need not be the only origin of this belief, desired states of affairs sometimes occurring in other ways. The effect of the belief in the actuality of the motivating state of affairs either causes the desire for that state of affairs to revert to a proto-desire, \(-W/E_1\rightarrow\), or causes its extinguishment, unless, of course, our active attitudes are obsessive, an irrationality marked by insatiability, or only momentary satisfaction, of a desire in the face of the belief that the goal-event has in fact occurred.  

19 Again, see chapters 3 and 5 below for more on obsession.
4. Finally, if the desire for the causing of W/E₁ is accompanied by the belief that it is both non-actual and causable, then this state will have satisfied a rational and practical condition for its proceeding in practical reasoning and it will engage the system further for identification of the event E₂, which the system believes has the causal properties sufficient for W/E₁.

In this model of agency, the function of belief is to add content to, or subtract content from, the motivational states of the system which themselves drive the process through practical reasoning and which may ultimately affect the world. The informational content of a motivational state, which is, in whole or in part,²⁰ the result of the belief functioning on that motivational state, allows those motivational states to be identified as states which the system ought either to shunt aside, as non-practical states (e.g., hope), convert to non-motivational states (e.g., wish), or allow to progress toward action (e.g., desire-to). The belief functions of picking up and passing on content or information are fundamentally functionally distinct from the motivational function of the active attitudes, which is causation, rather than information transfer.²¹ Motivation is treated here as causation with respect to some actual state of affairs which may be either non-

²⁰How much of the informational content is determined by belief will depend upon the level of autonomy achieved by the agent.

²¹This description of the functional difference between cognitive and connative states assumes that causation is not itself an informational relation. Perhaps the causal function of a desire or other cognitive active attitude is describable in information theoretic terms as the passing on of information from states of the agent to the world.
cognitive states of affairs beyond the body, as they are in typical cases of overt action, or cognitive states of the system, which would be the case in so-called mental acts. This functional bifurcation is what fundamentally distinguishes desires from beliefs. Only if we find that bifurcation unnecessary because we find, for example, the functional rôles non-distinct\(^{22}\) in a theory of agency and action, may we denigrate belief-desire psychology.

### 4.4 The Structure of the Content of Belief and Desire

If it were true that each of the main types of practical attitudinal state, the active attitudes and the epistemic attitudes, had typically distinguishable content, then we might have simplicity of theory as a reason to take those differences as explanatory of the differing causal rôles of the attitudes. That is, we might take difference in structure of content as explanatory of a state's unique positions in the process of practical rational agency. Each attitude-type's rôle would, then, differ in the practical rational process according to its type of content. And differences in content would instantiate rational progress in that practical process. The rôle of belief in practical rationality is, as most would agree, to provide the information or structured content whose addition to a previous stage constitutes increased

\(^{22}\) One type of single function theory is the view that desires really are beliefs. The desire as belief theorists include T. Nagel, *The Possibility of Altruism*, Part II; J. McDowell, "Are Moral Requirements Hypothetical Imperatives?," *Proceedings of the Aristotelian Society*, Supplementary Volume, 1978, etc.; M. Platts, *Ways of Meaning*, Chap. 10.
rationality, that is, practicality, and causes the sequent state in the process. Suppose, for example, that a desire state harbours the following informational structure: “the event of my agency which will have causal relations to my goal-event of eating that apple”. The form of the structure of the content of the desire is that of a definite description and not a full proposition. Let us agree that a desire is a state with a certain structure to its content and causal relations which define a part of the practical rational process. Suppose that the practicality of the process is to sufficiently identify a motivated event the agent will cause. Then the “active” state within that process will be a state whose causal direction is being determined by a content whose structure is in the form of a definite description of that motivated event (whatever in the further development of cognitive theory that will turn out to be). Desires and the other “active attitudes” will therefore be information- or identification-hungry. Intending, as we argue later, will be both identification- and evaluation-hungry. The process of practical reasoning adds to the description of the event the agent will cause, but since the practical state is, we assume, finally causal with respect to an event in the world, the state requires completion of its incomplete definite description. For such causation to occur, the active attitude requires a description which is sufficiently identificatory: a definite description.\(^{23}\) It is the function of belief in practical means-end reasoning to add

\(^{23}\)Identification is one component of cognitive sufficiency—the other being evaluation—which is a property of a fully-formed intention.
further descriptions to the definite description structure of the active attitude that is being processed by practical rationality. Beliefs about particular events with definite descriptions as their subjects serve this end since their function is to add information to an existing definite description structure. Thus if the agent were to desire to eat that apple and believed it could be eaten by first biting into it, then, in rationality, that means belief would cause to be added to the desire the further information that its object is also the event with (causal) relations to the biting of it. Each such increase in information is an increase in practical rationality and constitutes therefore a sequent state in the process.

We may see from the function for the beliefs of practical reason that their content will typically differ in structure from that of desire and the other "active" states as well, deriving as it does the content for its subject term from the definite description which is the content of the active state which causally precedes it. It is the rôle of belief to add the content/information of its predicate place to, and thereby cause, the next practical state. The source of the additional content of the predicate place is a matter for an epistemic theory. The contents of practical beliefs are therefore truly propositional in form: a derived definite description as subject plus a supplied further description of that subject as predicate. As we hope to have shown, no active attitude will have this form and therefore only the

---

24 See next section for more on this point.
epistemic attitudes are properly named propositional attitudes. The others should be called practical rather than propositional attitudes.\(^\text{25}\)

4.5 Content Concatenation and Practical Reason

The representational content of the proto-desire —F/E₁— is essential to the transformations which occur in practical reasoning. It is the determinant of relevance of all subsequent stages of the process which leads to action. The essentiality of this representation points to its rôle as motive, as the constituent of every mental module around which content continues to concatenate in practical reasoning. The idea of a motive is that of a term which is an anchor at one end of the nomological span we employ in the explanation of practical rational agency. At the other end, action anchors the explanatory span. There are other terms set into this sequence which allow us shorter nomological journeys within the span. These other terms, as we shall see, in addition to ‘motivant’, include, on the side of agency, nodal points in the process of practical reason such as ‘desire’ and ‘belief’, from which we have just come and which are explanatory because of their content and function, and ‘intention’ to which we shall shortly come. On the side of action we shall find the objective but still functional correlatives of the content of these

\(^{25}\)H.-N. Castañeda, an “attitude assimilationist,” argues for just such a division among the objects, or contents, of the attitudes in virtue of their different structures. On his account, the object of a belief is a proposition, while the objects of desires and intentions are “practitions.” A practition is, roughly, the mental content correlative of a thing to do. See Thinking and Doing, (Dordrecht: D. Reidel, 1975).
nomological anchors in such terms as 'goal event', 'means event' and 'event of agency', as well as their correlative functions.

The property of desire which allows it both to concatenate and to decatenate content from beliefs puts into relief the presence, in the history of a desire, of an attitudinal state, the proto-attitude or proto-desire for $F/E_1$, namely $\neg F/E_1 \rightarrow$. These states bring causation into the process and are rationalizable into desires. No desire module is without such a term and neither is any subsequent state which is effective in action possible without the inheritance of that representation. Where $\neg F/E_1 \rightarrow$ has been the occasion of an action schema or process, then the representation, $F/E_1$, will occur in all subsequent states leading to action including desires, hopes, wishes, beliefs and intentions. The inheritance of the first term by all subsequent mental states which lead to action is the basis of our notion of concatenation of content in that each segment in the process leading to action concatenates content around this representational anchor. The point of concatenation is to provide sufficient identification of the event with causal relations to $F/E_1$. To represent the event of agency which will (best) stand at the head of a causal chain which leads to $F/E_1$ is the rôle of practical reason. But such identification cannot be made out unless $F/E_1$ is part of each subsequent practical description and the description is not sufficiently completable without concatenation unless, as may only sometimes be the case, the motivating event, $F/E_1$, is identical with the event the agent will directly cause.
movement of your body would be such a case. The motivational rôle of \( F/E_1 \) is expressible in syntactical terms as the necessity of its content to the formation of all representation which occurs in pre-action intensional states. \( F/E_1 \) is a representation of the state of affairs for which reason is being practical.

We saw above that practical reasoning commences with an initial active attitude, a desire for some state of affairs, \( F/E_1 \), together with the basic causal belief that causation is necessary to make actual every non-actual causable and the belief that \( F/E_1 \) non-actual and causable. The effect of this pair of beliefs is a state with a structure which has as object the event, \( E_x \), which would stand at the head of a causal chain leading to \( F/E_1 \). This representational structure poses the problem for practical reason: identify that event which would best head such a causal chain. Identification of this causal chain must make essential use of the description, \( F \), which must therefore be retained in all pre-action intensional states of which \(-F/E_1\) is the motive. Concatenation of descriptions of the agency-event which would initiate the causal sequence leading to \( F/E_1 \) continues until that event is satisfactorily identified for the agent.

### 4.6 Full Decatenation of Content—An Exercise in Rationality

Antecedent to the desire that the government come to be embarrassed, there is an attitude to the motivating event contained in the desire: that is, there is the causally and rationally prior state which is causal with respect to the
representation of the government's embarrassment. We described above the processes by which a desire may be causally and therefore practically disabled by the belief that the object of the desire is either actual, impossible or improbable. Let us now explore the possibility that the pre-practical proto-attitude state \( \rightarrow F/E_1 \) is further dissociable. Consider the meditative processes, perhaps some Buddhist technique, of "de-contentifying" one's mind in which one is taught to empty the mind of all content which could be tied to causation or attitude, perhaps by using the "Zen" belief, "All is worthless anyway, and all striving with it". The objective, according to our model, is to expunge the \( F/E_1 \)'s from all \( \rightarrow F/E_1 \)'s; to purge oneself of teleology; to develop the ability to make mental content fade so that cognitively efficacious states are left with only efficacy, attitude, or "empty" \( \rightarrow \), energy not wasted on effort or striving. Nirvana! This Zen-like process takes us from desire back to its necessary condition, \( \rightarrow F/E_1 \), dissociates the elements of that antecedent state, expunges content and could leave us with a non-cognitive self-suppressing state of efficacy. If this process could continue and expunge all efficacy as well as content we should have intellectual death. Perhaps some meditative results approach this. Notationally, this de-agency process takes us from desire for the coming to be of \( F/E_1 \),

\[
\begin{align*}
\text{\( w \text{\rightarrow } F \)} & \quad \frac{E_1}{E_x} \\
\end{align*}
\]
back to $-F/E_1 \rightarrow$ and with the representational $F/E_1$ removed, to $\rightarrow$ alone. This is accomplished presumably through the effects of certain anti-agency beliefs and the use of certain pacifying or agency-unrelated content. We should notice the similarities between this process of content removal and that which occurs during the cognitive satisfaction of a desire.

The point of this Zen-like exercise is to make plausible, from an account of how agency goes forward, a process which employs the functions of practical rationality yet may take one back to the *simples* of agency. The total regression of this active causal function, which requires that it be totally denudable of content, leaves the control of its content, and therefore of that part of our actions, in the domain of beliefs. That makes plausible the idea of an autonomy wherein the content of motivation is totally determinable by reason. This would be a governance of the content of any origin of action by the mind rather than the acceptance as a given by the mind of at least some existing contentified active state of the system which therefore always supplies some content, however rarefied, as an unconsidered premise toward a practical conclusion. An argument about the scope and nature of autonomy is an argument therefore really about the nature of practical syllogistics.

---

26 "Think of the sky!" Although we are not practitioners of Zen, the above process surely has sometimes not seemed entirely unattractive to authors.
5. DISTINCTIONS AMONG THE PRACTICAL STATES OF THE PROCESS OF AGENCY

It has emerged that the processes of rational agency require states with the following properties:

1. type-distinct structures or contents, which determine a state's
2. type-distinct causal relations with other states.

The states in the process are attitudinal states such as liking, disliking, desiring, believing, doubting, knowing, hoping, wishing, intending and planning. To describe a state with an attitudinal term is to describe it as having structural and relational properties which constitute a part of the process of practical rational agency. The names of the attitudes are thus just names of states with relational and structural properties, a set of which defines the process.

That each attitude-type has typically different content or structure is not universally acknowledged. That this possibly typical difference of structure is a determinant of the distinct causal rôles of the attitudes themselves is also debatable. We shall argue for the truth of the first and offer a reason to accept the second.

---

27 See John Searle in both *Intentionality* and *Speech Acts.*
28 These causal rôles may be distinct only in their position in the process. See chapter 4 section on “theoretical smoothness.”
5.1 Desire-that vs. Desire-to

To begin, compare the desire that the Tirpitz be sunk with the desire to sink the Tirpitz. They would be analyzed by some\textsuperscript{29} as two distinct attitudes (functions), the desire-that and the desire-to, respectively, with a capacity for identical content, occupied, in this case, by “the Tirpitz’s sinking.” We have argued that the structure of the content of both of these attitudes is: the causation of the event of the Tirpitz’s sinking, not that sinking itself. This follows if a general belief in causation and a belief in the non-actuality of the object of a practical active attitude are conditions of such a state’s being part of the practical rational process. If the attitude is set in a rational system it will then be directed at the event, or type thereof, which would cause the sinking of the Tirpitz. In the case of the desire-that the Tirpitz sink, the structure of the content of the state is given by the language: “any event with causal relations to the event of the sinking of the Tirpitz.” Contrariwise, “the event of my agency which would cause the sinking of the Tirpitz” gives the structure of the content of a desire to do so. In a desire-to, the practical enterprise is focused on the identification of an event of agency, an event the agent himself may cause, and not just any event with the right causal relations. We cannot therefore treat a desire-to and a desire-that as completable by the same, and therefore interchangeable, content.

\textsuperscript{29}The so-called “object assimilationists” argue that the different attitudes take objects of the same type, and therefore differ in virtue of their unique attitudes, perhaps causal roles, which they take to those objects. See section 2.3 above.
5.2 Desire-to vs. Intention

The case for typical difference of content is the same between intentions, which are always intentions-to, and desires-to. Although both attitudes are held with respect to the same sort of event, namely an event of agency, and not just any event which might have causal relations with respect to a goal-event, other parts of their content nevertheless differ. As we shall argue in the section on intending, a state of intention is a more rationalized state and therefore positioned "later" in the rational process than a desire-to. Its greater rationality consists in the fact that certain positive normative beliefs about the goal-event are necessary for an active attitude state to assume the causal position of intention in the process. The normative beliefs will, of course, have to be believed justifiable in relation to a normative theory. It is in virtue of this normative judgement being part of its content that intentions are distinct from mere desires-to.

If we accept the above, then the difference between the structures of the content of a desire to A and an intention to A would be as follows. To desire to sink the Tirpitz is to be causal with respect to an event of one's agency which would cause the sinking, i.e., to that action. On the other hand, to intend to sink the Tirpitz is to be in a state which is causal with respect to an event of one's agency which would cause the sinking—that action—and that causal property of the agency event one believes is desirable. It would therefore be false to portray
the syntax of the above cases as "the desire to A" and "the intention to A" as if their contents were intersubstitutable.

In this account of a portion of practical reasoning, we have seen the fundamental terms and relations of agency functioning to an ideal; namely, goal-functionality. At the center of this account are the information transactions which are presuppositional to rational agency. With these fundamental terms and relations of agency before us, we turn, in the next chapter, to an account of the fundamental failures of these terms and relations.
CHAPTER 3

THE FUNDAMENTAL AGENCY FAILURES:

A NOSOLOGY OF AGENCY, PART 1

1. INTRODUCTION

In the previous chapter, a portion of practical reasoning was presented and analyzed in the fundamental terms of our theory. In that model, the functions of the agency states themselves jointly constitute the agency processes. It is within that processional setting that the functions of desire and belief are distinguishable. Thus we distinguished two broad types of agency state, each with its type-distinct content and causal rôle. The epistemic states have their functions absorbed in being information or content conduits from both the world and other states of the agent and to agency states of the type necessary for action, the active attitudes. Without states which perform the rôles ascribed to belief—without, that is, a state the rôle of which is to control the content and function of the motivational states—there could be no rational action, if to be a rational action requires at least that it be under the control of the agent's beliefs. The active attitudes include desire and intention, whose functions include the ability to accept information transferred from belief and to cause the event which they
represent. When the informational transactions which occur between the epistemic and active states are goal-functional, the state which is produced in the process of intention formation will cause the goal. When the informational transactions fail to be goal-functional, the resultant state will be non-goal-functional, a state which could not cause the goal: it would be a state which lacks the cognitive sufficiency for successful action.¹

In this model of agency, progress through the process is marked, in informational terms, by the addition (or subtraction) of information the content and structure of which are necessary for practical rational agency and action. Thus, wherever an information transaction is required for agency to go forward, there is a place a failure might occur. The result of such a failure would be that the agency process would fail to produce the subsequent practical state in the agency process, and may thus fail altogether to produce a state which is cognitively and causally sufficient for action. The task of this chapter is to produce a taxonomy of failures of practical rational agency which will identify and classify all the ways in which the informational transactions necessary for action may fail.

First, then, an agent may fail to get his goal because of the nature of the content itself of one or more of the agency states antecedent to action. Secondly, agency may fail due to the failure of one or more of the individual functions

¹Since action is an ideal, and thus implies success, 'successful action' is a pleonasm.
required for the exchange of content. A taxonomy of agency failures will therefore include a set of content *malfunctions* together with a set of *dysfunctions* for each of the functions we have distinguished in the agency processes. Among the class of cases in which agency fails, there will be those which are explainable by one or another of these types of failures of the individual agency states, but there will be those which require more complex explanation. For example, there will be cases of non-goal-functionality due solely to the fact that one's belief about how to get one's goal is false. In the alternative, that false content itself may be due to some malfunction of the belief, in which case the explanation of why agency failed would have to advert to both content and function.

If content and function exhaust the constituents of agency, then content malfunctions and the function dysfunctions will occur in the *explanans* of all of the failures of agency. A test for whether a taxonomy is complete would be whether an explanation for every case may be found within its resources. Therefore, once we have in place a taxonomy of the "pathologies" of agency, which have been put in sharp relief by our fine-grained model of agency, we may turn our attention to the manifestations of agency failures in agents whose complexity approaches that of our own, with the hope to gain some purchase on how and why agents such as ourselves may be less successful agents than we might otherwise be.
2. THE INFORMATIONAL DOMAIN OF AGENCY

Systemic dysfunctions of the sort mentioned above would be powerfully disruptive of agency. Our model of the process of intention formation is one in which intention emerges out of the bringing together of the two ingredients necessary for action: causation and information. This bringing together we described as the concatenation, by the active attitude, of the goal-functional information supplied by belief. Thus, a necessary condition of action for agents which keep separate their epistemic and their active agency states is that their active attitudes be informable or impressionable by impressive epistemic attitudes. Only when the effect of this passage of information is sufficient to explain causation in agency terms is the causal active attitude which causes action cognitively sufficient for action.

That the active attitudes readily accept information from belief is clearly not alone sufficient for goal-functionality, but constitutes only a minimal, or presuppositional, condition of goal-functionality. Other conditions, we have seen, will include the nature of the content which is passed—whether it is true, false, relevant, etc.—and what the states so informed do with their content once they

---

2 The measures that an agent may take to mend such faults will vary with the nature of the aberration and the resources available to the agent. Among the measures available to agents (see chapter 5 below) are those cognitive means by which a dysfunction may be ameliorated. Other dysfunctions may be correctable only through causal intervention—various drug therapies, for example.

3 These are things you really can’t do without.
have it. The total set of information transactions which constitute intention formation comprise the informational domain of agency.

We may distinguish three areas within the informational domain of agency, in each of which information is involved in key agency functions. The first is the set of interrelationships of the epistemic and the motivational states necessary for agency, a portion of which was described in the previous chapter. These interrelationships comprise the locus of rationality in agency and will be addressed more fully below. While the interrelationships are at the center of rational agency, an agent with these internal functions but without the ability to connect with the world would be incapable of agency. That is, these interrelationships, while necessary for rational agency are themselves insufficient for successful agency, or goal-functionality. What more is needed is the agent's ability to connect informationally with the world. An agent's informational connection with the world has two aspects: the ability to take information from the world and, in being causal, the ability to give information to the world. The item we have in our folk-psychological arsenal designed to perform the rôle, among others, of information uptake is belief. A precondition of a belief performing this central rôle in agency is its ability to stand in the right relation to the world. In

---

4This, again, is an information-theoretic analysis of causation. If we are systems in which cognitivity emerges at one of our (causal) interfaces with the non-cognitive, then what is to prevent the thesis that cognitive causation necessary for action is analyzable as the correlative interface of the cognitive with the non-cognitive: cognitive causation is non-cognitive "uptake" from the cognitive.
order to provide goal-functional information to the active attitudes, beliefs must have goal-functional information, and they can have it only if they can get it. Therefore, a pre-condition of goal-functionality is that one's beliefs must be impressionable—they must be able to take on content at those interfaces of belief formation. And, if beliefs are to function in successful action, they must "represent" what caused them—must, that is, be informational states the content of which has "fidelity" with the world one's active states are to be causal in. An agent whose beliefs lacked this function could not hope to act successfully. This is not to endorse the view that one's beliefs must represent or "mirror" the world in any strong sense. However, the concept of successful action, of "works," presupposes that the agent's beliefs have fidelity, in some sense, with the world in which the agent acts.

In order for an agent to act at all it must have the ability to produce change in the world. The folk-psychological items we have designated to do this job are desire and intention, members of a class we have dubbed the active attitudes. Without states one function of which is to produce change in the world beyond, as well as in, the body, one could not act fully in that world. But not any change counts as an action: actions are not mere effects. An action, we have argued, is an event, with certain properties, caused by an intention which represents and positively evaluates the change it produces.\(^5\) Without this event which has the

\(^5\)See *Agency In Action*, chapters 1 and 4.
properties intended of it we cannot make sense of action. In informational terms, action is the giving of information to the world, precisely that information contained in the state which is causal, and thus the information which explains the effect. Correspondingly, then, without a state which is both cognitive and causal we could not make sense of intentional action.

We have distinguished a trio of informational functions necessary for agency which stand along a continuum which defines the logical space of agency: at one end stands belief, at the other stands intention, and the middle ground is occupied by the interrelations of the epistemic and the active attitudes which comprise intention formation:

1. The relationship of the world to the agent’s epistemic states.
2. The interrelationships of the epistemic and the motivational states.
3. The relationship of the agent’s motivational states to the world.

Each of these rôles requires certain functions for the states involved. We have identified two of those functions: the function of belief to “uptake” information, and the function of intention to “express” information. The interrelationships of the epistemic and the active attitudes require certain other functions of the attitudes: beliefs, if they are to perform their informing rôle, must be able to give information to the active attitudes, and they, in turn, must be able to uptake information from belief. Furthermore, in order for the informational transactions to be practical, the content of the attitudes must have a certain structure or
syntax. In our model of agency, the epistemic attitudes have propositional content and the active attitudes have practical content: desires and intentions are "practitional" attitudes. Of course, syntax alone will not guarantee that the informational transactions are practical: practicality requires that content satisfy semantic conditions such as truth and relevance. The structure of goal-functionality or agency, then, requires a set of conditions on the content of one's agency states, and another set of conditions on the functions of those agency states. First, we will look at the ways in which content may fail to be goal-functional.

3. THE CONTENT MALFUNCTIONS

3.1 Semantic Malfunctions

Consider those cases in which the explanation of why an agent failed to get his goal adverts solely to the content of the relevant agency states—cases, that is, in which the content of one of the practical attitudes induced non-goal-functionality. The reason for 'solely', in the previous sentence, is to distinguish these pure cases from those in which the non-goal-functional content is in place due to failures of other sorts: for example, due to some functional failure. If we could speak of the

---

6 The mere fact that content is non-goal-functional is not sufficient for irrationality. However, irrationality, of one sort or another, may be implicated in the having of such content.
content of states as itself having a function in agency, that function would be to have fidelity with the world in a fashion which allows agents to cause their goals. Thus, a semantic malfunction will be one source of non-goal-functionality. The semantic malfunctions will all be cases in which the content of the agency states lacks that fidelity with the world in which the agent will act, which is a condition of successful action. Semantic malfunctions may have their origin in certain other failures of agency, such as ignorance, inattention, and selective or biased attention. Should one of these be the source of an occasion of non-goal-functionality, the explanation of why the agent failed must advert to those functional failures which explain the content malfunction. For the moment, however, we are concerned with the “pure” content malfunctions, and not the possible sources of those malfunctions.

Perhaps the most important of the content malfunctions is falsity. If agents are to act in the world, then their epistemic states must connect with the world in the right way: that is, their beliefs must be true. Making a mistake is a case in which the agent’s belief about means supplies false content to his intention which, as a result, causes an event which does not have causal relations to the agent’s

---

7 For some account of ‘true’.

8 What is the desire correlative of a false belief? A “false” intention, or a “mis-intention”—one which will not cause what it represents. Should a desire/intention inherit from a false belief about means false content about the causal route to the goal, the result of adding the false means content to the active attitude is that it will not, with that content, represent a causal route to the goal: that is what it is for means content to be false.
goal, but rather to some other non-goal event. Another source of agency failure is lack of true relevant content. An *accident*, as distinct from a mistake, is a case characterizable as an event which lacks the causal relations to the agent's goal, not now because of false means beliefs, but because of a *lack* of true belief about an intervening and goal-preventing event. And an *inadvertence* is an event with unforeseen causal relations to a non-goal event which is due to the lack of true belief, not about a goal-preventing event, but about effects.

While false beliefs and lack of true beliefs may be the more commonly recognized causes of agency failure due to content, they do not exhaust the content malfunctions. To this set we may add that content malfunction in which the agent's concepts are incoherent or garbled. The man whose concept *my hat* takes as its extension his wife, suffers from a semantic dysfunction. The explanation of this concept-aphasia may be his "faulty wiring"—perhaps his perceptual mechanism is deviant in taking input from his wife to his concept *my hat*—or, to brain damage. We should be careful to distinguish this case of semantic malfunction at the concept level from other dysfunctions. Suppose an agent's concepts are semantically sound, but that beliefs formed about the agent's states are, because of faulty belief-formation at *this* point, semantically unsound. Alternatively, imagine the agent with semantically sound states *reports* of which

---

^9 See *Agency In Action* for a full account of the unintentionalities, *mistake*, *accident* and *inadverntence*. 
are semantically garbled due to some form of aphasia. Non-pathological cases of garbled or incoherent concepts would include those in which the agent simply learned the concepts incorrectly. In addition, or as a result of some other aberration, there are those cases in which, while the concepts themselves are sound, the information transactions are non-goal-functional because the content employed is irrelevant. Perhaps some other forms of aphasia are describable this way.

Thus we have the following categories of semantic malfunction:

*Semantic Malfunctions*

1. falsity\(^{10}\)
2. insufficiency\(^{11}\)
3. irrelevance
4. incoherence

3.2 *Syntax Malfunctions*

The semantic properties of the content of the agency states are not the only properties of content which may produce non-goal-functionality. Our model of agency includes two broad types of attitude—the epistemic and the active—each with type-distinct causal rôles and type-distinct content. The relations between the agency states and the extra-intentional presupposes a certain structure to the content of the various states. We described the structure of the content of the

\(^{10}\) False belief is implicated in *mistake*.

\(^{11}\) Lack of true belief is implicated in *accident* and *inadvertence*. 
epistemic attitudes as propositional, and that of the active attitudes we described as practitional. The epistemic attitudes represent states of affairs: ideally, those states of affairs which caused them. Their syntax, therefore, is properly propositional. The active attitudes represent events which they are causal with respect to. An intention is a state which, under certain conditions, would, ceteris paribus, cause that event which is described by its content. Therefore, the structure of the content of the active attitudes is of a description: ideally, a description of that state of affairs with respect to which it is causal.

The interrelations we described between and among the agency attitudes presupposed the above structure or syntax of the content of the agency attitudes. Only with the structure of a description with an empty “slot” to be filled by content from a belief can an intention so structured pose the question it is the function of means-end reasoning to answer. And only with that slot open, as it were, may the intention “take” content from belief and thus be informed as to the identity of the event which has causal relations to the goal. If this is so, then syntactical malfunctions would be as disruptive of agency as would the semantic malfunctions mentioned above. Imagine ersatz beliefs the content of which is non-propositional; and ersatz desires the content of which is non-practitional. These would be states incapable of goal-functional information exchange, and therefore incapable of producing action. This sort of malfunction may be what explains syntactic
aphasia, the inability to produce syntactically correct sentences. Mental states with garbled syntax could not interact goal-functionally, if to do so requires, as we have described, states with propositional content informing states with practitional content.

Thus we have the following syntax malfunctions:

Syntax Malfunctions

1. Belief-like states which are non-propositional
2. Desire-like states which are non-practitional

4. THE DYSFUNCTIONS

In addition to the pure content sources of non-goal-functionality, there are those failures of agency which are due to failures of the individual functions of the agency states themselves. In our model of agency, each stage of the process towards action is a product of the individual functions of states of the two broad types: the epistemic and the active attitudes. Both states are required to pick up content from their respective sources, store it in characteristic ways, and to pass it on to the their respective targets. Thus we have distinguished the following informational functions for states of the types epistemic and active:

Information Functions

1. Uptake
2. Storage
3. Expression

Syntactic aphasia is also known as cataphasia.
The uptake function is what allows belief and desire formation and revision. The storage function allows states once formed to endure. The expression functions are those functions we have identified above as necessary for action and rational action. The expression function for belief is its ability to pass information to the motivational active attitudes. The expression function for desire and intention is their ability to cause what they represent.

I assume that each of the agency states is a causal state with a unique function which is determined in part by the content of the state. Like other causal items in the causal nexus, the causal powers or properties of agency states are mutable. Although I will not argue for it here, it may be the case that cognitive agency states are, by their very nature, less "steady" or constant in this respect than the mutable non-cognitive causal states. Whether this is so or not, in what follows I will assume that agency states are susceptible to causal flux and that this unsteadiness of the agency states is the source of many of the dysfunctions. Each of the functions of a state will have a standard which will reflect the typical strength of the function for most of us, most of the time. Alternatively, a function may be overly strong or too strong, or overly weak or too weak, again measured against some standard we have for the functions, and which is arbitrated finally by the goal-functionality of the strengths. Aberrations of strength may be specific to individual states, they may afflict state types, or they may be global and affect all states.


4.1 Recalcitrance

Imagine an agent who, under conditions under which we would say that he ought to have formed a belief, fails to do so. Suppose, that is, that conditions exist at the interface of the cognitive and the non-cognitive which standardly are sufficient for belief formation, yet no belief is formed. Or, imagine that the conditions which are standardly sufficient for belief revision are present, yet no revision occurs. There may be several possible explanations for a failure to form a belief or to revise a belief's content, ranging from faulty perceptual mechanism to the failure to make an inference justified by one's (other) beliefs. However, our identification of the fundamental functions of belief in agency makes possible the following explanation for failures of these types. The omission may be due solely to the failure of belief's uptake function, a species of the state dysfunction we will call recalcitrance. An agent who was belief-dysfunctional in this way would be seriously impaired qua agent. We may presume that recalcitrance is a dysfunction of agency states which may affect only individual states, one at a time, or it may be a more global dysfunction and affect many or all of one's agency states. It may also be a content-specific dysfunction for the agent who cannot form certain beliefs.  

Recalcitrance may also afflict the active attitudes desire, intention, hope, and wish. Suppose that an agent's desire that $\Phi$ come to be the case is

---

13 An agent who suffers from global belief recalcitrance is an agent it is “hard to make an impression upon.”
dysfunctionally recalcitrant. For an agent whose active attitudes are under the control of his beliefs, the ideal result of the presence of the belief that there is available to him a causal route to $\Phi$ which he himself may initiate would be to convert the desire that $\Phi$ come to be into a desire to $\Phi$, an important practical step, marking, as it does, the first fully practical state. If the agent's desire-that is recalcitrant, he would fail to form the desire-to, and remain in the pre-practical state of only desiring that $\Phi$, rather than moving to the fully practical state of desiring to $\Phi$. Should recalcitrance prevent the conversion of the desire-that into a desire-to, the agent would remain in a state which his practical reasoning could not address directly.

Amenability to belief is essential to the formation and revision of hopes and wishes. A recalcitrant desire may thus remain a desire despite the presence of what otherwise would be sufficient conditions for its reversion to a mere hope or a wish. We argued above that a hope is a causal state with respect to an event the causing of which the agent believes is improbable, and a state which, as a result, is not in the agency stream: a hope is a non-practical state. But the relevant change in belief, to the effect that the causing is now not improbable but possible, ought, in practicality, to convert the hope into a desire, and thereby place that active attitude en route to action. A hope's recalcitrance would prevent such a conversion and deny to the agent the search for possible means to what should have become a desired state of affairs. Conversely, the practical effect on a desire
of the belief that the desired state of affairs is impossible to cause is to convert that desire to a wish. The desire's recalcitrance would leave the agent in the non-goal-functional state of desiring to cause what he believes he cannot.

Finally, if full or ideal autonomy is the ability of an agent to form his own desires by supplying to desires all and only that content which accords with his theory of the agent he would prefer to be, desire recalcitrance would be enough to thwart the full exercise of autonomy.

### 4.2 Incontinence

Once formed, states may suffer from a range of dysfunctions. Imagine an agent whose beliefs or desires were only very short-lived and suppose the explanation is that his states are "unsteady" or vaporous and unable to store content once they have it. As a result of such a dysfunction, states would vanish soon after their formation. We will call the inability of a state to store content incontinence. Incontinence, like recalcitrance and the other state dysfunctions, may affect only individual states, or it may be more global and, say, affect all states of a type. It is also a dysfunction which itself may be strong or weak. Thus we can imagine states which lack the ability to hold content altogether, and those which may hold content, but only briefly. An agent whose means beliefs could not hold content

---

14 Unfortunately, this term is in wide use to describe the inability of an agent's beliefs to control his desires. We will be careful to distinguish the inability of agency states to hold content from the failure of belief to function so as to control which desires are efficacious for the agent. More on incontinence below.
would be an agent without the ability, or at least impaired in the ability, to identify causal pathways to his goals. Incontinence may be content-specific for agents whose beliefs about, say, *modus ponens* simply vanish, or for the agents whose best intentions simply disappear.

We may also suppose that the dysfunction of incontinence may afflict only certain parts of the content of a state. Thus, imagine a desire to Φ which was only incontinent enough to lose content related to available causes of Φ, specifically, that content to the effect that there is available a causal route to Φ which the agent himself may initiate—the very content which distinguishes desires-to from desires-that. The effect of this dysfunction would be to prevent the desire to Φ from being the practical state it ought to be. Alternatively, suppose that an agent’s desire to Φ is incontinent enough only to lose content related to goal. The effect would be a causal state with content about “means” to a goal the agent no longer has, clearly an impractical state. An agent whose desires were so afflicted would be forced to pay the additional cost to agency of constant surveillance of his desires, and he may be forced to resort to various maintenance and restoration techniques, perhaps including those taught by the psychotherapist, to circumvent the dysfunction.¹⁵

¹⁵These techniques include constant “re-minding,” and stressing the desirable aspects of a goal. It is arguable that rationality would require some such measures of the agent who is dispositionally dysfunctional and unable to ameliorate the dysfunction through self-control. I return to the issue of control and the distinction between ideal and non-ideal control in chapter 5.
4.3 Garrulousness

In our account of agency, states of the two broad types necessary for practical rational agency have an expression function. The differences in the nature of the two distinct expression functions integral to rational agency is part of what makes a state the type of state it is. Beliefs, we have argued, pass their content to active attitudes in practical reasoning, and active attitudes in turn are causal with respect to the content they have absorbed from belief. Expression, then, is a function which is necessary for practicality: without the informing rôle played by belief, active attitudes would be without the information necessary for goal-functionality, and without the causal rôle played by desire and intention, there would be no action. So, while practicality would demand that states of each type be expressive, it also supplies a condition on expression: not only must expression occur, but it must occur practically; that is, goal-functionally. What goal-functionality requires of the expression function is that it occur, and occur in the right way. Thus, another way for agency states to be dysfunctional is for them to be non-goal-functionally expressive. Imagine, then, a desire that $Ψ$ which was causal, despite the fact that its content is insufficient to describe a causal route to the desired state of affairs which begins with an event of agency. We described a desire-that as a state which was causal with respect to the causing of a state of affairs, but a state as yet uninformed by beliefs about the availability of an agency-causing of the desired state of affairs. Such an outburst of causation,
uninformed, as it would be, by belief about how the desired state of affairs, $\Psi$, may be caused, would be unlikely, to say the least, to be goal-functional. It would be a causal state which lacked the cognitive sufficiency for goal-functionality. A causal desire may have any number of effects, limited only by the state's connections with the (rest of the) body, and would therefore be a non-practical exercise of causation. Such an unrationalized outburst of causation would thus be premature, occurring as it would before certain cognitive conditions necessary for goal-functionality were in place. It would be an instance of the expression function of a desire which was unamended by belief. We will call such a dysfunction *garrulousness.*

Desires-to may likewise be garrulous. Imagine an agent with a desire-to, a state which is a proper input to means-end reasoning about how to cause the desired state of affairs, but a state, let us suppose, which has yet to enter that process. If so, it is a state the content of which lacks a sufficient description of the causal pathway to the desired state of affairs, including a description of the event of the agent's body with which agency must connect if the desire is to initiate that causal chain to the goal. Should such an unrationalized state be causal, it would be prematurely so, and unlikely to be goal-functional, since it lacks cognitive sufficiency for goal-functionality. It would be a state which exercised its expression function, but unamended by belief. It would be a causal eruption without the benefit of cognition. A garrulous desire is commonly called a
compulsion, and garrulousness may explain some forms of compulsive behaviour, obsessive-compulsive disorder and impulse control disorder.\textsuperscript{16}

What would it be for a belief to be garrulous? If the account of garrulousness is to be univocal, a garrulous belief would be one which exercised its expression function prematurely. For there to be premature belief function there must be mature or ideal belief function, and so we have to ask what the ideal is for the belief's expression function. In our account of the nascency of agency in the preceding chapter, an important step in the process toward practical reasoning proper was marked by the effect on a proto-desire of the beliefs in the non-actuality and the causability of the state of affairs represented in the proto-desire, together with the general causal belief that non-actual causables require causing for actuality. The effect of this belief complex was to convert the proto-desire into a state whose expression function was directed towards the causing of that state of affairs, rather than towards that state of affairs itself. It was just this additional content which marked the origin of a desire-that. Suppose, then, that an agent believes that some state of affairs, \( \Phi \), is non-actual. This is a belief, therefore, which may, in conjunction with the other relevant beliefs and the proto-desire for \( \Phi \), contribute to the formation of a desire that \( \Phi \) come to be. We may presume that the expression function of the belief that \( \Phi \) is non-actual is triggered by the

\begin{footnote}{\textsuperscript{16}}\text{However, the latter two syndromes invite more complex explanations which include a deficiency of the function of beliefs to control desires. More on this in chapter 5.}\end{footnote}
presence in the system of the proto-desire for \( \Phi \), and not otherwise. This presumption is justified by the fact that only when so triggered is the expression function practical. But suppose that our agent has no proto-desire for \( \Phi \), and has instead a proto-desire for \( \Psi \). Should the belief that \( \Phi \) is non-actual express that content in the presence of the proto-desire for \( \Psi \), it would do so irrelevantly and impractically, and produce a state whose content would prevent it from being further rationalized. Such an occurrence may be explained by the over-expressiveness or garrulousness of the belief that \( \Phi \) is non-actual.

Garrulousness of belief may explain certain forms of obsession, which is commonly described as an uncontrollable thought or idea, and one for which awareness of its inappropriateness may have little or no effect in attempts to control it. An agent with garrulous beliefs and desires, and prone, therefore, to irrelevant causal outbursts, would thus be an agent much less likely to be goal-functional than the agent whose agency states functioned ideally.

4.4 Reticence

The expression function of agency states is a function without which there would be no action. In addition to the expression dysfunction of over-expressability, or garrulousness, we may distinguish a dysfunction characterizable as the failure to express content. We will call this dysfunction reticence. Reticence, like the other state dysfunctions, may be limited to individual states, or it may be more global,
or content-sensitive. Reticence, too, is a dysfunction which admits of degree: reticently dysfunctional states may range from the weakly reticent to the totally reticent for agency states which lack their expression function altogether.

Reticence of belief may result in the outright failure to supply content to active attitudes in practical reasoning. Thus suppose that an agent desires that the ship he sees on the horizon sink, and suppose that he believes that one way to bring about that desired state of affairs would be for him to sink it. But suppose that this belief is reticent, and so fails to convert the desire that the ship sink into a desire to sink it. The effect of such an impractical dysfunction of belief would be to deny to the agent one, let us assume, promising avenue of endeavour.

With the introduction of the dysfunction of reticence, we have an alternative explanation for the failure of practical reasoning to produce the next practical state in the process, despite the presence of what would otherwise be sufficient

\[17\] I am not here claiming that every failure to form a desire-to would be due to belief reticence. Rather, I am offering an analysis of one sort of such a failure in terms of one sort of functional failure of belief. Clearly, there may be other causes of a failure to form a desire-to, including those caused by other state dysfunctions, as well as those caused by certain forms of interference. In addition, there will be cases in which reasons, such as fear, for example, prevent the move to the next state. Where fear is judged to be justified, we would not describe the omission as a failure. Fear may prevent, not only actions, but intention and desire-to formation. An analysis of fear and its function is beyond my present purposes, but if fear is an active attitude, perhaps the desire to avoid harm which is believed to be imminent, fear may interfere causally in the production of other states. A dispositionally fearful or timid person may be one whose agency processes are too easily interrupted by the prospect of harm, either because his agency states in general are dispositionally weak or reticent, and so easily thwarted, or because the desire to avoid harm is overly strong or garrulous. A fuller account of these and other interesting aspects of agency must wait for the fuller working out of this programmatic project.
conditions for the production of that state. In our discussion of the dysfunction of recalcitrance, we found that an agent may fail to form a desire-to despite having the desire-that and the relevant belief about there being a possible agency causing of the desired state of affairs, conditions under which an agent ideally moves from a desire-that to a desire-to. There the explanation of the omission was that recalcitrance of the active attitude rendered it unimpressionable by the belief. The belief dysfunction of reticence introduces an alternative explanation of such a failure: an agent may fail to form the next practical state in a practical rational process, despite the presence of what would otherwise be sufficient conditions for that formation, because the relevant belief is reticent, a dysfunction which renders it unimpressive, and therefore denies it a function necessary for progress through the process.

A dysfunctionally reticent active attitude would be a state whose expression function, which is to cause what it represents, is diminished or absent altogether for the state which lacked that causal sufficiency. Since the point in agency at which the expression function of an active attitude ought to occur is at the point of action, and since we have not yet proceeded that far in our account of the agency processes, we lack the context which would best display reticence of desires and intentions. But perhaps some foreshadowing would not be out of place. Imagine, then, a fully formed intention which, were it not for its reticence, would be fully equipped to cause an action. It would be a state which had been fully rationalized,
and so cognitively sufficient for action, but due to its reticence, it would be a state which lacked causal sufficiency for action, an impractical state par excellence.

Reticence is a dysfunction of active attitudes which may occur at any stage of the agency process. Thus, imagine an agent who desires to \( \Phi \) and who is now deep in the process of means-end reasoning about how he may best \( \Phi \), when the process is brought up short by the sudden diminishment to zero of the causal strength of his desire to \( \Phi \). Since the causal function of desires and intentions is one essential component of motivation, the agent would no longer be motivated to \( \Phi \), and reasoning about how to \( \Phi \) would be otiose. Reticence is thus a dysfunction which may play havoc with goal-functionality.

5. CONCLUSION

The above dysfunctions of agency states are all informational dysfunctions. Given the central importance of information transactions in our model of agency, the dysfunctions will be powerful sources of non-goal-functionality and may, in certain cases, result in non-agency. We have seen how the individual dysfunctions may disrupt certain early stages of practical reasoning, such as the formations of desires-that, desires-to, wishes and hopes. The dysfunctions, however, may occur at any point in a reasoning process which involves the epistemic and the active attitudes. Before we may fully appreciate the extent to which the dysfunctions may disrupt agency, we need a complete account of intention formation for agents
which approach the complexity of human agents. That complexity will be a complexity of a goal-functional process and will therefore reveal a number of additional ways in which an agent may fail to be goal-functional. In the next chapter, the account of intention formation will be completed. Once that is done, we will consider an agent with this structure and in possession of more content, including more goals, than we as yet have considered. That content will include the agent's beliefs about his own agency and the world in which he must act, and a theory about how to be an agent given the contingencies of his nature and circumstances. We will then be in a position to address how and why he may fail to meet the standards of agency such a theory would impose.
CHAPTER 4

INTENTION FORMATION: THE FUNDAMENTAL PROCESS OF AGENCY

1. INTRODUCTION

With the fundamental terms and relations of agency having been set out in chapter 2, we are now placed to consider the agency process which produces a state of the agent which is both causally and cognitively sufficient for action. That state is an intention, a state which has been rationalized in agency to such an extent that it may play the rôles required of it by our explanatory action discourse: those rôles include causing what it represents, and thereby explaining what is caused.

The account of the agency processes which follows is an idealized account of information processing which moves states through the process which culminates in a state the function of which is to initiate action. That process will be described here as conforming to a certain sequence. This is an idealized model, and actual agency processes may exhibit considerable plasticity in the order by which a fully formed intention is produced and yet still qualify as practical rational agency just so long as that process both includes the important nodal points of practical rationality about to be set out and produces a fully formed intention which is both
cognitively and causally equipped for action. Getting clear on just what is required of practical rationality, especially in fundamental terms, will set the stage for our consideration of the failures of practical rationality for complex agents such as ourselves.¹

The process I will describe will consist essentially of agency states interacting according to the functions set out in chapter 2. Their interactions are describable by a set of laws of interaction which describe goal-functional information processing between and among states of the process, and the failures of agency will consist of failures of just these functions. It will be specific departures from the ideals of agency which constitute failures such as impracticality and irrationality.

Theorizing about agency requires us to remain within the realm of the cognitive, since such a theory is meant to describe the nature and relations of the cognitive states required by a theory of behaviour on the part of agents such as ourselves. Thus what follows is a theory of that process which takes an agent from those cognitive states which are, or may be, emanations of the physical to those states which, in overt action, emanate in the physical. An action, in being the expression of cognitive causation at that end which emanates in the physical, implicates a cognitively causal state, such as desire, at the other. Such a state

¹Because what follows is an ideal of intention formation, each step in the process will be practical. Each step will also, therefore, be a condition of rational practicality.
must have the properties of cognitivity and causation in order for what emanates in action to have a mental etiology. We begin our account of the process, then, with the occurrence of what we above termed a proto-desire, a state whose nature contains the seeds of a fully formed intention.

2. PRE-PRACTICAL REASONING: THE CAUSAL AND COGNITIVE ORIGINS OF DESIRE-TO

2.1 From Proto-desire to Desire-that

In our examination above of the structure and function of the desire states which function in action, it proved useful to trace the rationality of such a state back to its more primitive causal and cognitive origins, a state which we dubbed a "proto-desire." Our project here is to describe the agency processes by which certain key agency states are produced. That account should begin, therefore, with a description of the construction of a desire from its humblest beginnings. Assume, then, the occurrence of a proto-desire for the death of the King. Even though our account of the, possibly, Zen-like process by which the content of any motivating state may be expunged made plausible the occurrence of the content-less state of causation, the bare arrow of causation \( \rightarrow \), the state which would, therefore, be found at the commencement of the processes which generate a desire, I here leave aside that part of the process which would generate a proto-desire, and simply assume its occurrence.

It is worth noting also, that an account of the rational process which yields these important states of agency does not preclude the possibility that an agent may find himself in any one of these states, but which is unconnected to any antecedent cognitive process: agents may be caused to be in agency states in any number of odd and non-cognitive ways, including having them "parachuted into" the process.
analyze such a state as one of causation, \( \rightarrow \), with respect to the event, \( E_1 \), under the description 'death of the King', \( F \). This cognitive and causal state, the proto-desire for the death of the King, we represent as \(-F/E_1\rightarrow\). The death of the King itself, the causal object of the proto-desire, we represent as \( F/E_1 \).

A proto-desire is a state not yet fit in its structure for practical reasoning. Should our agent, therefore, now engage in reasoning about how to cause the event \( F/E_1 \) at this stage, without further rationalization (which I am about to describe), he would be guilty of a form of impracticality. Furthermore, since a proto-desire is a pre-practical, rather than a practical state, an explanation of why reasoning is occurring about how to cause the event it represents would be inexplicable. That is, certain conditions have yet to be satisfied before a state with \( F/E_1 \) as its object can function as input to practical reasoning. The active-attitudinal state, the desire that the King be, or become, dead may be seen as a state which is conclusory, the result of a process which includes among its cognitive and causal ancestors, the proto-desire for the King's death, a state which is uninformed by beliefs about the actuality and causability of that state of affairs. A proto-desire is, unlike its rational descendants, consistent with the beliefs that the represented state of affairs is either actual or non-actual, causable or non-

---

\(^9\)It would be impractical to reason about how to cause \( F/E_1 \) without first believing that it is non-actual and causable. It would also be irrational to judge the doing desirable without first believing that it was do-able.
causable. Thus, an agent may be in this proto-desire state and hold either of the above beliefs without contradiction.

If a practically rational system has the proto-desire for the death of the King, and holds the general belief in causation, the belief that causation is required to make actual any non-actual causable, then with the addition of the beliefs in the non-actuality and causability of that state of affairs, the system will be caused to be in a causal state with respect to a new object, one with articulation derived from the above mentioned active and epistemic states which serve as its practical premises. This new active state has as its intentional object the event which would cause the death of the King, and would be the desire that the King die, an active attitude toward the coming to be of that state of affairs, and so an attitude with the representational content the causing of the death of the King.

This desire—that is the first conclusory state in the practical process which, with further content, may yield an intention to kill the King. It has, in contrast to its proto-desire ancestor, a structure which allows it to be open to the practical question of how the object F/E\textsubscript{1} may be caused; the question, that is, of the identity of the causal pathway to F/E\textsubscript{1}. However, such reasoning about the cause of the desired state of affairs would not yet be practical reasoning proper, a term which we commonly reserve for that part of the process of intention formation which identifies, not just any causal pathway to the desired state of affairs, but one

---

\textsuperscript{4}To fail to be so caused would be a practical failure.
which the agent may initiate: a causal pathway to the agent’s goal which begins with an event of his agency.

What would practical rationality require of an agent with the desire that the King die, together with the belief that such an event was not causable? Clearly it would be impractical to devote the resources of agency to searching for a causal pathway to an event believed not to have one. Thus, the belief in the non-causability of the desired state of affairs ought to remove the desire-that from the practical process, and “park” it as a wish or a hope; as a state, that is, which is causal with respect to a state of affairs believed to be non-causable, and therefore a state which cannot enter the practical process. Alternatively, the belief in non-causability may cause the desire to revert to its active attitude ancestor, the proto-desire for the King’s death. In either case, our agent would no longer be in “action mode” with respect to the death of the King, although just so long as his beliefs remain revisable, that attitude toward the death of the King may return to the practical rational process with the belief that the King’s death was causable.

Suppose, in the alternative, that our agent believes that the King’s death is causable. What ought the effects of this belief be on the desire that the King die?

---

5 Impractical because non-goal-functional.
6 The goal-functional ‘ought’ again.
7 It may be that, according to our standards for practical rationality, entry into the practical reasoning process requires that the system believe that a causing of the motivating event is possible. Alternatively, our standards may employ the weaker condition, that the system not harbour the belief that such a causing is impossible. We have here opted for the stronger condition.
We may consider the belief in the causability of \( F/E_1 \) as opening a gate, as it were, to the further practical rationalization of the desire, and the belief in its impossibility as closing that gate and opening another to the attenuated states of mere wishing or hoping, states which are not in the practical process. The belief in the causability of \( F/E_1 \), together with the desire that \( F/E_1 \) come to be, function together as inputs to the enquiry about whether that causation is possible for the agent; i.e., whether there is a causal pathway to (the causable, non-actual) \( F/E_1 \) which begins with the agent himself.

Our description of intention formation is a portion of a theory of agency. As was noted above, the theory admits considerable plasticity on such things as the order in which certain of its steps occur. Moreover, there seems nothing in rationality to require which of the possible non-practical states above ought to be the consequence of the belief that the motivating state of affairs is not causable. That is, rationality may be satisfied when the effect of this belief puts the agent into a state of wishing the King were dead, hoping the King would die, or merely liking the prospect of the King's death. This is to say, the model of agency before the reader admits of plasticity here, as well at other points we will note. More important than plasticity is the point that if the desire that the King die is to have a cognitive history, or itself be a cognitive state, then the agent must believe that
the King's death is causable. It cannot be practical to continue in the practical vein without this belief.\(^8\)

2.2 From Desire-that to Desire-to

We have arrived at the point in the practical rational process where the desire that the causable and non-actual F/E\(_1\) come to be comprises an input to further practical reasoning. Suppose that the agent believes that there is an event which he can cause which would have the causal relations sufficient for F/E\(_1\). The effect of this belief is to transform the desire that the King die into the desire to kill the King. This is the desire that the agent himself be the cause of the death, the agent, that is, of the change in the world sufficient for that death.

The belief that there is for the agent an "entry point," as it were, into the world which would allow the agent to cause the object of his desire is an important further articulation of the causal object of the agent's desire. This description—'event of my agency'—further narrows the range of possible causes of F/E\(_1\), beliefs about which the agent will soon come to survey. The search for the cause of F/E\(_1\) has now been greatly narrowed. The agent now desires not merely some cause of

\(^8\)Alternatively, it cannot be practical to proceed in the practical vein with the belief that the causing is impossible.

What form of irrationality is committed by the agent who proceeds to process a desire which fails this condition? There are three possible forms: 1. The agent fails to vet the desire on this point, and so neither believes the causing is possible or impossible; 2. The agent believes it is non-causable (or lacks the belief that it is causable) but proceeds to process the desire (a) out of weakness of rationality, or (b) because the active attitude is too strong for rationality to control on this point.
Chapter 4

F/E₁, but his causing of F/E₁, an event of his agency, which we symbolize as Ag, which would initiate a causal sequence including the death of the King. Heretofore, in merely desiring that the King die, there was too little direction from that description of the causal object of the desire to involve practical means-end reasoning. Thus, only desires-to are proper inputs to means-end reasoning since only with desires-to is the agent attempting to identify an object of his own agency, a truly practical enterprise.

The causal and intentional object of a desire-to is an event the agent himself can cause, and that description, although incomplete and in need of completion by practical reason, remains constant until action has occurred. The essential and steady question of practical reasoning is: What event causable by the agent will satisfy the description of being the sufficient condition of the goal? The desire to kill the King is the active attitude whose causal and intentional object is an event of agency, Ag, with causal relations to the goal-event of the king's death, F/E₁, and we represent that desire as:

\[
\begin{align*}
    Ag & \xrightarrow{\text{w}w} \frac{F}{E_1} \\
    & \quad \xrightarrow{E_x}
\end{align*}
\]

a state of causation, \(\rightarrow\), with respect to the event, \(E_x\), under the description 'event of agency, Ag, which would cause the goal-event, F/E₁.'

With this structure to the content of a desire-to, practical reason has its final direction and specification of the description it now seeks. This description
'event of agency which would cause the goal' will exercise means-end reasoning for further identification of $E_x$ until that indefinite description is made definite. In means-end reasoning, the agent canvasses his beliefs about possible entry points into the world which would have the desired effect, selects, from among possible alternatives, the one judged by him to be desirable, and adds that description to the appropriate place in the structure of the desire-to.

The causal function, $\rightarrow$, with which we describe the performance of desires-that, desires-to, and their ancestors, is, despite differences in their objects, the identical function. All these states are causal with respect to the states they represent. If the agent should come to believe that the object of its desire-to is not an event of agency, i.e., not causable by him, then that desire-to must, if the process is to be practical, disengage from the practical process as did hopes and wishes. The agent must cease desiring-to, and with that the process of agency with respect to that goal must cease, although a desire-that may persist. Thus only those states which have been impressed by the belief that there is an event of agency which would cause the goal-event may proceed into practical reasoning. The set of necessary conditions for agency will include the satisfactory identification and evaluation of the causal route to the goal. These are described in the next section.
3. PRACTICAL REASONING, FROM DESIRE-TO TO EFFICACIOUS INTENTION: THE CAUSAL AND COGNITIVE ORIGINS OF INTENTION

There are three major features at work in practical rational agency, two of which we have seen at work above. The first is causation, without which there would be no motivation, no practical rational efficacy or action. The second is identification or representation of that which the agent is motivated with respect to, or of what the agent will cause, without which agency would be blind. The third feature of practical rational agency is evaluation of a proposed course of action without which an agent would be at the mercy of his desires. Identification and evaluation contain the logic of practical reasoning: they comprise the process of practical reasoning and are themselves causal or functional roles of certain types of beliefs. Thus the entire process of practical reasoning consists of two informational functions, the accumulation of information within a causal state—a desire or intention—and the transfer of information by states of another sort—belief—to states of the first sort, the interactions of which are governed by a logic which comprises practicality.

The occurrence of a desire-to, such as the desire to kill the King, prompts an agent to evaluate that prospective action according to other of his goals. Evaluating an action is a complex process and may proceed in a variety of ways. It might begin with an examination of the consequences of such an action for the rest of the agent’s world. But whatever scope rationality requires of an agent’s evaluation, whether he is required to have a broad or only a narrow range of
concerns, to be a selfless or self-interested agent, moral or otherwise, evaluation is one of the functions which sets rational agents apart from agents who are at the causal mercy of their desires. That one's cognitive and motivational states are amenable to belief is a presupposition of rational agency. Since our purpose here is to describe the evaluation function, and not any particular normative theory of agency, we can proceed without taking a stand on how far down the causal chain of consequences rationality requires an agent to look and evaluate before acting. That requirement will be part of the larger theory of rational agency, found perhaps in moral and legal theory, and accounts of the reasonable man.\(^9\) For our purposes, then, which, to repeat, are to describe the evaluation function in agency, we may, for convenience, limit our considerations to the self-interested agent and expand that scope of concern only as far as a minimal theory or ideal of rationality would require.

In the discussion above of the belief function we saw content of the means sort being passed from belief and concatenated in desire. With the evaluation function, content of another sort will be likewise concatenated in desire as the active attitude passes through the process toward action. The first instruction given to agency by the presence in the system of a desire-to is to further identify for the agent the causal object of that desire. The second instruction is to evaluate that object. Is the desired object desirable? Identification and evaluation proceed

\(^9\)These issues will be addressed in the sequel.
together in practical rationality until the agent acts. With action the two main functions, causation and cognitivity, must come together: all the content concatenated through the process which defines practical reasoning is added to the function which affects the external world, and only then is action possible; only with the confluence of content or information and causation is action possible. The state penultimate to action, an efficacious intention, is an amalgam of these two functions: an external-facing causal function and content. Agency *simpliciter* is that function which can be informed by belief, can have content affect it in the ways we are describing, and when so informed, and only when, lead to action.\(^1\)

Suppose that our agent positively evaluates the desire to kill the King, believes, that is, that such a killing is not only desired, but desirable, all things considered.\(^2\) The effect of this evaluation is to convert the desire to kill the King to a state which is similar in content but with the significant additional content that the doing is now judged to be desirable. The content of this belief is thus added to the desire-to, producing, with this first subjection to, and affirmation by, rational (evaluative?) scrutiny of a state within the process, some commitment\(^3\) to

---

\(^{10}\) An intention is external-facing, in one sense of 'external,' only for overt actions, which include the events of the body. For mental actions, what the intention faces is “external” to itself, and to practical reasoning.

\(^{11}\) It is plausible to consider the function of belief in practical rationality as condition-remover on the efficacy of an active attitude.

\(^{12}\) This is a *conditional* judgement in favour of the doing only so described, and may be overturned in the light of further descriptions of the doing.

the doing: that is, an intention to kill the King. We mark the effect on an active attitude of a judgement that a doing is desirable with the symbol ‘Ev+’, for ‘positively evaluated’. The intention to kill the King we represent as:

\[
\begin{array}{c}
\text{AG w} \rightarrow \\
\frac{F}{E_1} \\
\frac{E_x}{Ev+}
\end{array}
\]

The occurrence of a state with this structure is a threshold worth marking in the process toward action and is distinguished with the name ‘intention’. It is the point in the process of practical reasoning at which the agent is causal with respect to an event of his agency which has causal relations to his goal and which has been judged by the agent as desirable. It is a prospective doing that is now not simply desired, but believed desirable. And the agent may now be said to be committed to the doing, rather than merely motivated to do it. The source of the belief in the desirability of what is represented in the active state may be derived from a theory the agent holds about which doings are desirable, about which desires are desirable, and could be as simple as the judgement that, in relation to his other beliefs, desires, and values, this doing is desirable. On the other hand, its source could be as complex as an objective moral theory. In either case, the effect is to move the agent from a state of merely desiring to kill the King, into a state of intending to do so. Before this, no evaluation of the doing had occurred.\textsuperscript{14}

\textsuperscript{14}Of course, there may well have been, and, perhaps, goal-functionally ought to have been, evaluation of the beliefs which have functioned so far in the process.
but only the fact of the agent's attitudes and beliefs about actuality and causability. As a result of the combination of the desire-to and the belief that the represented doing is desirable, the agent will cause the event $E_x$, of his agency which he believes is sufficient for the goal event, in our example, the death of the King, just so long as he believes that desirability is maintained.\(^\text{15}\) With the addition of these causal properties, the agent will, *ceteris paribus*, kill the King. Before the occurrence of this intention, that could not have been said, if we believe that evaluation, or the exercise of normativity, is an essential ingredient in agents whose agency is capable of being fully under the control of belief. Thus the importance of the belief in the desirability of the action, creating as it does, this crucial difference in practical state.\(^\text{16}\)

With the occurrence of an intention, practical reasoning enters an iterative phase of means-end and evaluative reasoning, each step of which adds content to

---

\(^{15}\) As this suggests, the question of desirability remains open throughout the process of intention formation, is askable at every point where a further description of the proposed doing is generated, and is answered in the affirmative at those stages which move the process along toward action.

\(^{16}\) Evaluation of this sort constitutes an important gate for causation. Its function prevents causation with respect to events which the agent desires, but which he judges undesirable. Suppose for a moment that the evaluative step fails, that despite an evaluation to the negative, causation proceeds. This phenomenon may be due to two sorts of aberration. Suppose that causation proceeds, that the active attitude with respect to the desired but now undesirable state of affairs remains in the action mode despite the negative evaluation because the evaluative function is too weak to prevent it. Either this will be so because rationality is abnormally weak, or because the causal attitude (desire) is abnormally strong. In the first case we have what deserves to be called a form of weakness of will: if will includes judgement as what one should do, that judgement is here impotent, or at least too weak to work. In the second case we have a compulsion: an active attitude that remains causal despite the best efforts of rationality to diffuse it.
the description of the action. As has been noted, on this model of agency the content of each relevant belief state in the process passes that content to the active attitude it is in service of, and the active attitude concatenates that content. Therefore, the final intentional state, the state which is causally and cognitively sufficient for action, has as its content the full practical history of practical reasoning, which has identified and positively evaluated the event intended. That event will be the one which is believed to have the causal relations which are sufficient and necessary for the death of the King. In our example, it is the event believed to be a moving of the agent’s hand which will have causal relations to the poison being poured in the King’s ear.

The content of an efficacious intention can thus be seen to have the following structure. There is content which identifies the goal-event, the means, and the event of agency. We represent that structure with the following template for an efficacious intention:

\[
\begin{array}{r}
\text{[Agency]} & \text{[Means]} & \text{[Goal]} & \text{[Ev]} \\[E_x]\end{array}
\]

This is the template which is “filled in” by practical reasoning. In our example, the goal slot has been filled by the event of the King’s death, \(F/E_1\). The means which was selected was the event of poison being poured in the King’s ear, \(G/E_2\). That event stands in causal relations both to the goal-event and the event of agency, \(Ag/E_x\), the event this state will, \textit{ceteris paribus}, cause. Reference to the means relation of the event of agency may serve to identify it as the represented
causal object of the active attitude: it is the event believed will have causal relations to the goal. For means, only the relationality of the means event is essential and desired, and while some non-relational description of that event may be necessary to identify it as one with the appropriate causal relations, any description of that *relatum* may do for the identification of its functional rôle: as cause of the goal-event and effect of the event of agency. In contrast, a goal-event is motivational only under its goal description.

As practical reasoning proceeds to address the intention to kill the King, that intention becomes modified, through accretion of, or concatenation by, beliefs about which causal strings of events containing the goal-event \( F/E_i \), have "entry points" for agency—that is, which causal strings with \( F/E_i \) as a terminus have an event of agency at their source. Suppose, as does this example, that the agent believes that the pouring of the poison in the King's ear is the best way to achieve the death of the King. This of course presupposes an evaluation of possible causal routes to the goal, in which the poison pouring ranks highest. The function of that belief in practical, means-end reasoning is to pass its content to the intention to kill the King, one function of which is to concatenate that content in the means position. The result is the new causal and representational state, the intention

---

\(^{17}\) It may be a feature of our epistemic abilities that we only identify relational properties of events by their non-relational properties.

\(^{18}\) An interesting dysfunction of an intention would be for it to concatenate that content in, say, the goal position.
to kill the King by pouring poison in his ear; or the intention to cause that event of agency which will cause the death of the King by causing the poison to be poured in the King's ear.

The event of agency, $Ag/E_x$, at this point in practical reasoning, may be represented by the agent as an event merely of this type, or as an event token. But, at this early stage in the identification of the event which is to be his killing of the King, it is unlikely that the event will be identified beyond its type, an event of agency which would cause the death ... . In any case, a plurality of chains of agency-caused events believed to culminate in the King's death by poisoning would entail a plurality of events of agency to be considered and evaluated. The various causal pathways believed to be available to the agent may not come into play for the agent until he has faced certain planning considerations. That is, the process toward action may cut into planning with the occurrence of an intention. Once there is an intention, or some commitment to action, the question of which particular causal pathway to employ may be deferred pending the larger and more complex question of how best to arrange behaviours so as to serve one's various commitments.\(^{19}\)

One particular causal pathway to $F/E_1$, and therefore one particular event of agency $Ag/E_x$, will finally however have to be fixed by the agent for intentional

---

\(^{19}\)See Michael Bratman's seminal work on planning, in, for example, *Intentions, Plans, and Practical Reasoning*. 
action to occur. The event, $E_x$, which is believed to have the properties to initiate a chain of causes to $F/E_1$ will then finally be identified through a sufficiently singular referring expression. In our example, it will be the event, $E_g$, which is the movement of his hand, $H$, which causes the poison to flow into the King's ear, $G/E_2$, which causes the death of the King, $F/E_1$. And thus the final descriptive content of the efficacious intention which, with the *proviso* that evaluation remains positive, will cause the event of agency described and positively evaluated as the killing of the King. That state is representable as:

$$\text{Ag} + H \quad \frac{G}{E_2} \quad \frac{F}{E_1} \quad \text{Ev}^+ \quad \frac{E_3}{\;}$$

This is the structure of a possible efficacious intention. And with it, this sketch of the function of practical reason in the formation of an efficacious intention is complete. I now turn to discuss some issues which this model raises.

### 4. THE EFFECTS OF PRACTICAL REASONING

#### 4.1 The Practical Syllogism: Desire

The cognitive and causal agency states we attribute to agents figure importantly in our appraisals of them and of their actions. Since there is some controversy over which states practical reasoning yields as conclusory, this topic is worth discussing. First, the practical syllogism with a desire premise.
Suppose that an agent desires to kill the King and believes that pouring poison in his ear is the best means available to him. What practical conclusory state should an agent be in as a result of these two premise states? Ought he desire to pour poison in the King’s ear? We know that we must not take this state to be the liking of the event of poison being poured in the King’s ear. That would be the wrong object of practical reasoning.\(^\text{20}\) We must take it as the desire to pour poison in the King’s ear, representable, then, not as causation with respect to that event, \(-G/E_2\), but as causation with respect to the causing of that event:

\[
\begin{align*}
\text{Ag} \quad \text{ww} & \rightarrow \\
\frac{G}{E_2} & \rightarrow \\
E_\chi
\end{align*}
\]

This desire-to state is supposed by some to be the practical and rational result of the above practical premises. To assess this claim about the process of practical reason, we ought to recapitulate our account of desire. We think of a desire as a state with a causal function which is determined by the structure of its content or information. The event about which a desire-to state contains information is an event of agency, an event the agent may cause, such as the movement of his arm. What our agent desires when he desires to kill the King is an event, \(E_\chi\), which he represents as of the agency type, Ag, which has causal relations, in our example through another event, \(G/E_2\), to the event which is the death of the King, \(F/E_1\).

\(^{20}\) It would be one way for agency to be impractical.
That event, and not F/E_1 itself, is the object of his desire to kill the King. To desire F/E_1 directly, and without causal relations to an event of his agency is impractical so long as causation is necessary for actualization.\(^\text{21}\)

The desire to kill the King—the desire, that is, for an event of agency with causal relations to the death of the King—does not change its object\(^\text{22}\) just because the agent believes that the event, G/E_2, is a means. If we need a principle by which to distinguish desires, it would be that desire identity is determined by its object; and so the identity of a desire-to would be determined by the event of agency being represented as the event with causal relations to the goal-event, F/E_1. The identity of the desire does not therefore change as the agent continues to articulate the nature of the relations which he believes the event of agency must have in order to be the event with the causal relations required to cause his goal.

The fact that the agent has such an attitude to this event of agency is explained by the causal relations that event is believed to have to his goal. Recall that the causal attitude, \(\rightarrow\), was transferred from E_1 to E_x because the agent believed in the necessity of causation for the causable and non-actual F/E_1. The causal attitude was subsequently connected to an event of agency, Ag/E_x, with the belief that F/E_1 was causable, albeit indirectly, by the agent. The causal relations

\(^{21}\) Another form of impracticality for agents in worlds in which non-actual events require causings is to fail to be in a causal state with respect to the causing of an event one desires would occur.

\(^{22}\) Its object is the event of agency with causal relations to his goal.
to his goal-event are essential to the attitude remaining tied to the right event of agency—an event of his agency which would cause that goal. That description of $E_x$ shows the line of causation, and therefore the reasoning, from the liking of $F/E_1$ to the desiring to $F/E_1$ and explains why $Ag/E_x$ is the event which has inherited the attitude of desire. It is the perceived causal relations of $Ag/E_x$ which cause its fuller representation to become the subject of practical reasoning. The relationships which exist among the structural components of a desire-to, that is, among the representation of the goal, the representation of the event of agency, and causation, provide the structure for a significant part of the practical rationality of agency. No inference from a desire premise to a desire conclusion in practical reasoning will be allowed which is a transference of the causal attitude to an object which is not the event of agency believed to have causal relations to the goal. Unless the representation of the event of agency as one which has causal relations to the goal controls the representational content of desires-to and intentions and therefore ultimately of what the agent will cause, practical reason and practical rational agency are not possible.

What do the above remarks mean for the inference from the desire to kill the King and the belief that pouring poison in his ear is the best means available, to the conclusion: the desire to pour poison in his ear? The impracticality of this conclusion is that it has lost, and so does not preserve, the relationship to the goal, the death of the King, and this change in causal object, this practical distraction,
may lead one to cause a poison pouring which does not have causal relations to the
goal.\textsuperscript{23} It would be impractical, that is, non-goal-functional to be causal with
respect to an event of agency with causal relations to the pouring of poison in the
King's ear without the \textit{proviso} that it be a pouring which would cause the goal, the
death of the King. Therefore, one does not come to \textit{desire the means} because one
desires the end. In fact, one does not desire the means at all. Rather, in practical
reasoning, one comes to desire the means \textit{as means}; but that is only to make
reference in a new guise to the identical causal relations which were referred to by
the description of the event of agency which would cause the goal. What follows,
in a "valid" practical inference, is not a state with a new causal object: it is a state
with a new description which further articulates the causal properties of the event
of agency. The practical process is one of sufficiently identifying that event so as
to make successful action possible. Therefore, we do not come to desire new
properties of the event of agency; rather, we desire the same properties—its causal
properties for the goal-event—under new descriptions. Only in that way is
practical rationality and therefore motivation preserved. And, \textit{vice versa}. It is of
course possible independently to desire the means not as means. But to do so
would be a practical distraction and not an effect of practical reasoning.

\textsuperscript{23} Since some poison pourings may not be death producing, although some are.
And, this inference would make the \textit{pouring} the goal—the event which would
satisfy the intention—whereas practical reasoning began with an altogether
different goal.
We must be careful to mark such distinctions if we are to have available a fine-grained and accurate map of the causal psychology of agents.

4.2 The Practical Syllogism: Intention

What should follow from my intention to kill the King and my belief that pouring poison in his ear will turn the trick? Do I intend to pour poison in his ear? Do I, that is, intend the means? If so, then if I successfully pour the poison in his ear, I do so intentionally. Since action discourse, which is informed by our beliefs about the causal psychology of agents, has consequences for our assessment and treatment of agents, something of importance is at stake here.

An intention to pour poison in the King’s ear is causal with respect to the event which would have causal relations to that event. Furthermore, it is a state which represents that causing as desirable. Without the positive evaluation in which the agent judges the action desirable, the agent merely desires to, in this case, kill the King. An intention is a state of desiring to do what the agent judges is desirable. Thus, in order to intend to kill the King, the agent must have positively evaluated the event of agency under the description ‘which would cause...

---

24 One might quite naturally be led to answer in the affirmative, failing to notice that what makes this seem plausible is the plausibility of describing what one intends as ‘to pour poison in the King’s ear so as to kill the King.’ Notice that intending to pour the poison is only a part of that description. Thus the description of the intention in our question is incomplete.

25 Among theorists who argue that one does intend the means is Myles Brand, Intending and Action, esp. p. 24.
the death of the King; that is, to have judged desirable the event of agency for those causal relations; that is, to have positively evaluated that goal-functionality of the event of agency. Since evaluation is tied to intensional contexts, the event of agency has been judged desirable essentially or non-substitutionally under that description, or, for having those relations to the death of the King. The causal relations from the agency-event to the goal which the agent has found desirable may be equally identified or represented as the causal relations which the poison pouring has if it has both effect relations from the moving of his hand and cause relations to the King's death: if, that is, the poison pouring is means. The causal relations which the event of agency has to the goal are, then, identical to those which the poison pouring has when that is the means. So the causal relations one intends and values are referable to as those which the means to the death of the King from the moving of his hand has. The causal relations one intends are equally the causal relations of the pouring of the poison in the King's ear. And those relations are identical to those which the killing of the King has. Since the death of the King is the desired effect along the same line of causation from the moving of the agent's hand, its position on that line of causation may also be used to refer to that line of causal relations. Similarly, we may use events otherwise positioned as *relata* among the set of desired causal relations to refer to those relations. There are, however, only three types of distinguishable position along this practical causal chain: an event as event of agency, an event as means, and
an event as goal. The concepts of action, means and goal name these types of relata in the line of causation.

This identity of reference is at work in the practical syllogism. The agent intends to kill the King and believes that pouring poison in his ear is means. With this means belief he has introduced an event which stands in the means relation to the event of agency and the goal event. The first premise used the position of the goal event to refer to the desired causal relations of the agency event. The means premise allows the creation of a co-referential description of that set of causal relations. They are now describable as the causal relations which the event of agency must have to the poison being poured in the King's ear, since that event is also describable as the cause of the King's death. So the same span of causal relations is encompassed in the conclusion as was encompassed in the first premise. The practical reasoner is thus entitled to conclude that the intention to kill the King is the intention to pour poison in his ear, and, with the further belief that moving his hand thus and so will cause to poison to be so poured, to conclude that the intention to move his hand in just this way is the intention to kill the King. The requisite means beliefs supply descriptions of events which are relata within an identical causal chain from the event of agency to the goal, from the movement of his hand to the death of the King, and allow reference to that same causal chain by a description of the relative position of the relata within that chain. Practical reasoning from an intention premise and a belief about means
yields a conclusion, then, which is a new description of what was identically referred to in the intention premise: the goal-functionality of the event of agency. New conclusions do not produce new referents, but only new representations of the original causal relations intended. These new representations which practical reasoning concatenates in the intention allow the agent to sufficiently identify the particular functionality of the event his agency will cause.

The agent will of course be aware that the event of agency will not only have the desired and, *pro tempore*, desirable causal relations to the goal-event, but also at least the properties his other representations of these relations indicate. These causal relations are not only those which span the route from the event of agency, his hand moving, to the goal, the death of the King, but also those which take us through an event which is a pouring of poison in the King's ear. So his event of agency has this property as well as its causal properties. Just as it was a requirement of rationality that the agent evaluate what he desires, it is a further requirement that he evaluate the additional properties of the causal relations he desires. And that is how he may choose among alternative causal routes to his goal. The rational agent will thus evaluate the object of each conclusory intentional state in which he may find himself as the result of the process which

---

36 Evaluation is one component of autonomy, and distinguishes the rational and autonomous agent from the non-rational, heteronomous agent.

27 This claim presupposes that the agent has certain goals about *how* he causes his goals, such as efficiency and lack of untoward effects. This requirement is thus beyond the core of rationality, and we will return to it, and similar expansions of rationality in the sequel.
identifies the event he will cause through additional representations of that event.\textsuperscript{28}

Despite these evaluations prompted by the additional descriptions of the causal relations to the goal-event that are desired and intended by the agent, it is practically rational for him to continue to desire and intend only those relations; i.e., only the goal-functionality of the event of agency, $Ag/E_x$, so long as his appraisals of the additional properties of these relations remain positive. In prudence, which is an exercise of rationality, he will evaluate these additional properties which he represents the killing of the King as having. But it is the killing of the King which has engaged his agency, not any other of the properties of the event of agency. That is his goal. And while there is reason to address how something was evaluated, action itself concerns what was intended and that is determined by motive, by the F-ness of $E_1$.

\subsection*{4.3 Minimal Means}

An active attitude is potentially and/or actually causal and represents what it will cause, its intentional object. The represented or intended causation is of a pathway whose function is to link what the intention will directly cause—the event of agency—with the agent's goal. What the agent represents as means is

\textsuperscript{28}Failing to evaluate the action under each of its new descriptions, or intending despite a negative evaluation of the doing under a new description, are forms of impracticality, with corresponding forms of irrationality.
just that causal pathway. Points on that pathway by which the agent identifies it do not have other of their properties incorporated within the intention but are to be taken only as markers or identifying points of the particular causal route intended between what the agency state directly causes and his goal. Thus the essential generic practical content of an intention, or of what the agent essentially represents in practical reasoning is:

1. the event of agency—what the agent will directly cause,
2. the particular functionality of the event of agency which will link him to
3. his goal.

5. INTENDING

An intention, we saw, is a state in the process of practical reasoning which culminates in action. As such, its nature would be revealed in the nature of the process of which it is a stage. The generic features of agency which precede and culminate in action include two generic causal functions: the function of the epistemic attitudes, including belief; and the function of the active attitudes, including desire and intention. Together these functions take content to and from the world, and thus are the functions which define our relations with the world. The belief function takes on content from the world via perception and, in practical reasoning, gives off content to desire and intention, which, when causally and cognitively sufficient for action, has the capacity to cause what it represents in the world. The function of these cognitive-cum-causal states is not explicable without
some notion of representation.\textsuperscript{29} The representational function of the active attitudes is to identify the event such states will cause. Since the only events directly causable by such states are events of agency, including body events,\textsuperscript{30} the content of the active attitudes has a structure which allows representation of an event of agency which the agent believes will have the properties and relations he judges desirable. One of the practical functions of belief is to service an active attitude by providing it with representational content sufficient to identify the event which will be the agent's action. The other belief function central to practical rationality is to evaluate that represented object as its identification proceeds through practical reasoning. Identification and evaluation constitute the rational aspect of practical reasoning.

In our account above of the origin of intention we argued that it was attractive to postulate an early and fundamental move which structured the content of an active attitude so that the question of the identification of the object of that attitude was raised and that this fundamental syllogism creates the state of agency we call desire. Desire is thus the first rationalized, cognitive state of agency. States of cognitive agency may, however, exist which are not so rationalized, which do not have the structure or informational content which

\footnotesize
\textsuperscript{29}This is not a metaphysical plea, but only a structural remark: we do not here urge any particular view of the metaphysics of the agency functions.

\textsuperscript{30}We do not intend to exclude so-called mental acts in which the effect of the intention is a mental event. These remarks apply, \textit{mutatis mutandis}, to mental acts.
would allow identification of the cause of the goal-event. We held, then, that an unrationalyzed state of agency, nevertheless one with cognitive content, was conceivable and could be represented as $-F/E_1 \rightarrow$, a state of agency or efficacy with representational content but with no practical representation of its object: no representation yet which allows it to be related to causation and therefore to agency. It is merely a causal state for an F-event but with no representation of the cause of that event and therefore no representation of an event which is linkable to agency. For practical rational agents, the belief in causation formulates the basic identification of the event which it is practical for agency to address, as the event which would cause the goal:

$$ww \rightarrow \frac{F}{E_1} \rightarrow E_x$$

This is a state of desiring that $F/E_1$ come to be. The object of the attitude now has a basis in causation. A desire is a rationalized state just because it is the result of the belief in causation having functioned upon an active attitude so as to begin the process of identification of the object which would cause the goal-event. The previous state, $-F/E_1 \rightarrow$, was uninformed by belief in causation and would result in "behaviour" which was not practical, because not informed by belief about how to connect the agent with his goal. The desire that $F/E_1$ come to be is not a non-rational state but it is a non-practical state: it does not yet allow the agent to
connect with the world. Furthermore, it, like the previous state, is one that could be satisfied in any number of non-agency ways.

The process of practical rationality described thus far has made use of the causal functions of belief and desire and representational content. Evaluative beliefs, which are of central importance for rational practicality, have not yet come into play. We assume that evaluation by the agent of any of his states with a causal function is practical and therefore rational. Evaluation may occur at any point in the process so far described.

The desire that $F/E_1$ come to be is a state which has not yet made the connection between an agent's causal capabilities and the goal-event. The only worldly event which may stand in the effect relation to an agent's internal states of agency are events of the agent's body. A system which lacked beliefs about, and therefore representations of, the limitations on the causal relations to which agency states are restricted would not be a system capable of action. Its behaviour

---

31 See remarks in chapter 5 below about the utility of evaluation for agents which can fail to function goal-functionally in as many ways as we.

32 In our account, evaluation first appears with the move from desire-to to intention. However, it may come in earlier. If evaluation of content has the essential role which I have argued it ought to have for agents like us, then evaluation ought to function at every step in the process. Thus, wherever belief functions, there too evaluation ought to have functioned antecedently, giving assent or commitment to the content, without which the belief would not have passed its content on in the process. Positive evaluation of belief's content makes it available to the process. The beliefs in non-actuality and causability, which function very early on in practical reasoning, require evaluation. Therefore, these are places for impracticality, and therefore, irrationality, to occur.

33 Again, we do not intend to exclude mental acts: those actions, the events of agency of which are mental.
might be like that of the infant who lacks beliefs about which causal routes are available to it and which of those lead to its goals. Where those beliefs do exist and the agent believes that there is an event he can cause which would cause his goal event, their effect is to convert a desire that such and so come to be into a desire to do that thing. The desire-to, we saw, is a practical state, and the agent has entered that part of the practical process which will identify his action. Heretofore, his active attitude had been first directed at a motivating event, $F/E_1$ simpliciter, and subsequently toward whatever might cause that event, viz., $E_x$. With desire-to, however, agency is directed at an event of his agency, $Ag/E_x$.

Agents can intend only these objects, and cannot intend merely that something be caused. Intentions thus have as their object the same type of object as desires-to, namely, doings, whereas a desire-that is causal with respect to causing of a "larger" sort. What separates an intention from a desire-to is the judgement, the belief, that the doing is not only desired but desirable. Where the agent believes that the desire to $F/E_1$ is desirable, believes, that is, that the event of agency which would cause $F/E_1$ as represented is desirable, which is equivalent to the belief that so acting is desirable, then, in a properly functioning system, the agent is in a state of intending to $F/E_1$. This is the first positive evaluation of a

---

$^{34}$ The ideal agent could not both desire to $F/E_1$ and believe that to do so would be desirable and not intend to $F/E_1$. The function that desirability adds is commitment to the doing. An agent's failure to be moved to a state of intending would therefore indicate some dysfunction or other; either belief reticence, or desire or intention recalcitrance, or etc.
doing, of an action, and with it, a threshold in the process toward the sufficiency for action has been reached.

It has been claimed by action theorists that an intention has certain properties which allow it to play its various rôles in agency. In the following sections we argue that the conception of intention we are here developing has the properties to play all the rôles required of it.

5.1 An Intention's Relations with Desire and Belief

According to the model of agency being presented, both desires and intentions are cognitively causal states whose presence one is able to report upon. Thus, the existence of a state of either type may comprise the content of a belief. Both desires and intentions may have their content determined by beliefs and both have a structure which is "identification hungry," which raises the question of the identity of the event of agency which would cause the goal event and to which the efficacy of the state is therefore directed. This similarity of structure shows that they are close neighbours in the enterprise which leads to action. That the


\[36\] Note the connection here with autonomy, and with degrees of autonomy: that desires and intentions may have their content wholly determined by the belief that such is desirable, etc.
difference between desires and intentions is constituted by the belief in the desirability of the doing shows that that difference is one of added cognitivity.

5.2 Commitment

In this model, an intention is a state which is the result of the first positive normative evaluation, following the first representation, of a doing: an evaluation of, therefore, a desire-to, the previous distinguishable state in the practical reasoning process. Since the difference between a desire-to and an intention is the belief in the desirability of the doing represented in the previous state, our claim must be that the first commitment to the doing by the agent is associable with this first evaluation. On our view, commitment is either the result of the increased rationalization of the doing due to the agent’s belief in its desirability, or it is just identifiable with that increased rationale for the doing. In either case, this model allows commitment to be a property of intention which is due to the increased rationale, something we should expect and want for such a pivotal change of state in agency as ‘intention’ marks. It should be noted that scalar increase and decrease in belief of the desirability of an action match the scalar increase and decrease in commitment to a doing of which an intention, and so an agent, is capable.
5.3 Purity

Purity is the property which precludes the possibility of the reduction of intentions to desire-belief complexes.\(^{37}\) In this model, an intention is the result of the belief that A-ing is desirable functioning on the desire to A, and therefore is not identical with those causal and cognitive antecedents. This is due to our view of content concatenation within active states of agency and we find the economy of states which that view entails attractive. Moreover, intentions are states which one might find "parachuted" into one, rather than caused in the usual cognitive way: they may simply occur as the result of non-cognitive rather than cognitive antecedents, as the result of some process other than practical reasoning. This possibility emphasises their purity since in neither case would antecedent desire and belief states be necessary conditions for intention. Actually, it is possible to imagine an inference to an intention from states which do not include a desire-to: the plasticity of practical cognitivity would allow evaluation of causation of an event before one has come to desire it and the subsequent direct move to an intention without the intermediary desire.

\(^{37}\)Most notable, perhaps, among the theorists who argue that intentions are reducible to desire-belief complexes is R. Audi, in "Intending," The Journal of Philosophy 70 (1973): 387-403.
5.4 Relations with Planning and Other Intentions

In our model of agency, intentions occur at a crux in practical reasoning. At that crux, and certainly no sooner, planning becomes practical. That is, as Michael Bratman has put it, one of the functions of intentions is as input to planning. Since the desirability of an intended doing may have been made relative only to its own properties and consequences, the intention would still require ordering in its desirability in relation to other intentions as well as to other planning considerations such as space and time might require.

5.5 Sufficiency for Action

One of the roles which intention must play is that of being causal in the world beyond the mind. What properties must such states have to be sufficient for action? On this model, that is the same question as when the processes of identification and evaluation of the content of a causal state are sufficient for action. If one is to be an agent, and a rational agent, there must come a point at which, for at least some intention, he need not, and will not ask of his beliefs how to implement that intention, or for further evaluation in the light of an added representation. To have an intention which is rationalized to a certain point is to

---

38 Myles Brand, following Sellars, who finds Davidson lacking on this matter, calls this question "possibly the fundamental problem...in philosophical action theory." M. Brand, Intending And Acting, 33.

39 For these questions to persist constitutes a form of impracticality and thus irrationality.
be prompted to act if agency is possible. In order to pour the poison in the King's ear our agent must, if he is an agent, finally become capable of affecting the world in just such a way. Where he is finally so capable is precisely where the question of means or of further identification and evaluation does not enter. Thus for an agent there must come a point where, for at least some intention, he need not ask of his beliefs how to implement that intention. In order to pour the poison in the King's ear our agent believes, say, that he must move his hand thus and so. But in order to move his hand thus and so he need not have any beliefs about which moves would be means to his hand movement. It is sufficient to have had the intention identified and vetted as above: sufficient, that is, to have the object of the intention sufficiently identified and evaluated. Our basic means of affecting the world thus must occur directly, as it were, and without further rationalization. Just as there are actions antecedent to which we perform no other actions, there are mental events after which occur no relevant others before we act. We have an event of agency when it is the effect which the agent represented in his causally sufficient intentional state as the event he with that intention would cause. The event of agency is then an object which is not capable of being waywardly caused.

The factual question of when an intention attains sufficiency for action is the question of when the practical reasoning process of intention construction is

---


41 See Myles Brand, *Intending And Acting*, chapter 1.
complete. Since the changes in this process are those of content and its structure, it will be these changes which determine when sufficiency of an intention occurs. The point at which the process actually satisfies the conditions of sufficiency may be determined or signalled by the occurrence of the belief that practicality—identification and evaluation—has been satisfied. When that belief occurs will vary from case to case and agent to agent. When that belief should occur, when the agent should stop thinking and start acting, and when not, are further questions which require for their answers a theory of rationality, one purpose of which is to provide the conditions of justification for a belief that an intention is complete. The notions of standard of care and the reasonable man are attempts in the legal world to address this question which is surely partly normative. We will address these issues in chapter 5.

Desires-to and intentions, and all the other active attitudes, share the function of being causal with respect to the body and other mental states of the agent. The various active attitudes, such as proto-desires, wishes, hopes, desires-that and -to, and intentions, are distinguishable by their content and their functional position in the process. We may regard all the practical reasoning in

42 An alternative is that the structure of the content of an intention is such that so long as their remains a slot to be filled with content, be that identificatory or evaluative, that intention has not reached causal and cognitive sufficiency for action.

43 The active attitudes are also the states which, in concert with beliefs, are causally functional with respect to the processional aspects of practical reasoning: that is, they drive the process from state to state.
practical rational systems as the satisfaction of a set of ceteris paribus conditions on efficacy. Thus, in rationally practical systems, active attitudes would be causal but for their being under the control of belief. \(^4^4\) As the process of identification and evaluation unfolds, cognitive content is added to the active states, marking the increasing satisfaction of the ceteris paribus conditions of the doing. When the event with the properties to achieve the agent's purpose is believed by him to be identified in the particular and still positively evaluated, then practical rationality, and therefore the ceteris paribus conditions on the doing, are believed by him to be satisfied and sufficiency simpliciter for the doing exists.\(^4^5\) Until practical reason is complete what exists is only causal sufficiency ceteris paribus.

5.6 Relations with Intentional Action and Responsibility

A fully formed intention, one which has satisfied all the ceteris paribus conditions of a rational system and is thereby both causally and cognitively sufficient for

\(^4^4\) This point marks a significant difference between our view and Davidson's. Davidson endorses the view that if an agent judges that it would be better to do \(x\) than to do \(y\), then he wants to do \(x\) more than he wants to do \(y\). One way to read this is the view that judgements may increase the causal strength of wants. On our view, belief may not affect the causal strength of active attitudes, although it may prevent active attitudes from functioning, regardless of their strength. See "How is Weakness of the Will Possible?", Essays on Actions and Events: 23.

\(^4^5\) Compare the nature of the intentional state at the point of causal sufficiency for action according to our view with Davidson's "all-out" or "unconditional" value judgements ("Intending," Essays on Actions and Events, 83-102) which are meant to be states of causal sufficiency on his view. His "all-out judgement" will only be an endorsement of the event he intends to cause under as few as one of its descriptions. This cannot be an account of the evaluation which must take place in order for practical reason to be considered rational.
action, has been called a present directed or immediate intention, a state which stands in the causal relation to the action. In contrast, a future directed intention is only causal ceteris paribus, or subject to further conditions of practical reason, such as that evaluation remains positive up to the point of efficacy, or to planning. That an agent's intentional state has caused an external event is only a necessary condition for the agent to have acted. Agents and their actions are related explanatorily as well as causally. An intentional action is, on our view, that set and only that set of properties and relations of the event caused by the agent which correspond to the representation or information with which it was caused; i.e., which correspond to the properties represented and intended in the content of the intention. That object makes essential reference to its explanandum, the efficacious intention, and so fits it for its rôle as an object for which we hold the agent responsible and to which normative considerations may be addressed. Only with the nature of the intention of the agent kept fully alive in the conception of his act can the idea of an action contribute usefully to these matters.

6. THEORETICAL SMOOTHNESS

The concept of agency requires the working presence of a state which is both cognitive and causal at the point at which agency and its object, action, come

---

46 These are Bratman's and Brand's terms, respectively.
47 See Agency In Action, especially chapters 1 and 4.
together in the causal relation. That presence is recognized by most writers and has borne the name of ‘present directed intention,’48 ‘immediate intention,’49 ‘intention in action,’50 and ‘all out, or unconditional judgement,’51 and it is that state whose presence and function both marks agency or the causal sufficiency of persons and explains the emergence of those objects called actions which we attribute to agents. On our account of these matters, the cognitive efficacy of the agent does not emerge suddenly at the point of action. Our view recognizes a "smoothness requirement" according to which there is both cognitive content and causal efficacy present antecedently and throughout the process of intention formation in practical reasoning. Views which see causation emerging suddenly and not until the point of cognitive sufficiency for action may feel the need to inject causation, not yet present, they believe, into the proceedings. Thus a willing, volition, or intention might be taken to be a required causal state not reducible to desires and beliefs since those states apparently lack the causal function to affect the body. This break between desires and intentions leads to a discontinuity in the theory of practical agency. The cognitive efficacy or sufficiency which functions at the point of action is best seen, however, as the culmination of a process which is explicable in terms of the states and processes of practical

---

48 Michael Bratman.
49 Myles Brand.
50 John Searle.
51 Donald Davidson.
reasoning. Since that process is entirely cognitive, entirely the function of reason, except for the active attitudes subject therein to reason, and since there is no property of reason which can create or add efficacy or sufficiency, the causal sufficiency which exists with the intention at the point of action must be attributable to earlier active states in the process. And the function of cognitivity or belief is to control its release by functioning as ceteris paribus conditions upon the already present, rather than the reason created, causal sufficiency for action. This presence of causal sufficiency under the control of reason has the further virtue of accommodating, with simplicity, incompletely rationalized outbursts of causation.\(^\text{52}\)

Thus, while it is true that any theory of action must account for why causation occurs at the point of action and neither earlier nor later, this explanation is not confined to the claim that causation just occurred at the point of action after practical reasoning and planning—that only then did cognitive and causal efficacy emerge. A superior theoretical option, we claim, is that causation is present as a power initially in the process with proto-desire or desire. The route to the expression of this power is controlled by beliefs which, we must imagine, may either open or close gates in the process so that the causal power already

\(^{52}\) A non-rationalized outburst of causation is an attractive account of some forms of non-goal-functionality and irrationality, such as compulsive and impulsive behaviour. See the discussion of these functional failures in chapter 2 above and in the sequel.
present therein, and reportable upon, may emerge, or not, depending upon the judgement of the agent.

While beliefs, as they function in practical reasoning, may control causal sufficiency for action so that it is expressed only if certain beliefs, and therefore certain contents, are in place, they may not create that sufficiency. Nevertheless, beliefs seem capable of affecting desires and intentions. For example, beliefs may create mere hopes or wishes out of desires. But this may be seen as the closing of a firmer gate upon the causal sufficiency of the active attitude which could however be removed later and the *ceteris paribus* process defined by practical reason resumed toward action. More problematic for theoretical smoothness is the fact that a belief seems capable of expunging an active attitude completely. If beliefs may undo efficacy or causal sufficiency entirely, why may they not create it also? This is really the question of whether we need two separate causal functions for action theory or only one.\(^{53}\) We here openly assume that two distinct functions, distinguished by their causal rôles in agency, must be made out however apparently unitarily they are packaged. The point to hold on to in this matter is that it is the belief function, with its injecting or withholding of information, which

---

determines what rôle in the process a state will play and therefore what state it may be.

This is what is meant then by theoretical smoothness. We introduce an object—an active attitude—at the beginning of the process which culminates in action, an object which has causal sufficiency and which explains not only the negative and positive effects which beliefs may have upon the prospects of cognitive causation, or action, but also why the process moves through its stations. The active attitude placed at the beginning of the process thus offers us a motivational aspect, through causation combined with information, and a rational aspect, through beliefs. But since both aspects of the account are essentially in causal terms, an additional causal concept, such as will or volition, is otiose. Just as ‘intention,’ then, marks an important juncture in the process of practical reasoning, namely, the first occurrence of commitment to do what was theretofore merely desired, a juncture which is the result of an affirmation by the belief in the desirability of the desired doing, so does ‘desire’ mark an important crux in agency: a desire is a cognitive state which is causally sufficient, ceteris paribus, for action. A state which was desire-like, but which lacked causal sufficiency, would not be a real desire. If an agent reports that he desires to A, but fails subsequently to perform as agents so motivated ought to perform, then, assuming other things are equal, that failure is due to the fact that, contrary to what he believes, he doesn’t
“really” desire to A: the state falsely reported upon lacks causal sufficiency for action.⁵⁴

Another feature of smoothness is the concatenation which occurs in the cognitive content of an active attitude. Practical reasoning begins with a state, in our example, the proto-desire for the death of the King, the content of which is a representation of such an event. As the active attitude proceeds through the process of practical reasoning, it gathers content. Later, when a means to the goal of the King’s death is fixed, the intention to kill the King becomes the intention to kill the King by pouring poison in his ear. And when that intention is finally efficacious it has become the intention to kill the King by pouring poison in his ear by moving one’s hand thus and so, which is believed to be desirable.

There is one further aspect to the smoothness hypothesis: there may really be only one active attitude throughout the practical rational process, and thus no distinction, qua attitude, between desires-to and intentions. That attitude would

⁵⁴This feature of our account marks a significant departure from those of other action theorists, most notably, perhaps, Donald Davidson, who finds it self-evident that “If an agent judges that it would be better to do x than to do y, then he wants to do x more than he wants to do y.” (“How is Weakness of Will Possible?,” in Essays on Actions and Events, p.23.) See also A.R. Mele, Irrationality, p. 33ff. This feature of their theory allows an account of akratic action according to which the desire to A, together with the judgement that it would be best to A, comprise insufficient motivation for A-ing. An explanation available to them is that the desire was insufficiently motivational, and that insufficiency remained, despite the increase in causal strength supplied by the judgement that A-ing was what one ought to do. In cases of akratic action, this insufficiency would allow a desire to B, which is sufficient for action, to itself be efficacious, despite the presence of the judgement that one ought to A, and despite the absence of the judgement that one ought to B. An account of akrasia occurs in the next chapter.
be a causal state, or, in computational terms, a state with the function of addition—the function to move to the next state provided by the structure of the process. Thus, desire and intention would be generically the same in their function: move to the next state in the process. Their difference would consist in where they performed that move-to-the-next-state function in the process. Desires perform it at an earlier stage, intentions later. Progress in the process, or the continued functionality of the states within it, is administered, however, by the increasing rationality of the content—definable as a certain structure—of the state.
CHAPTER 5

RATIONAL AND IRRATIONAL AGENCY

1. INTRODUCTION

The previous chapter presents an account of the practical process of intention formation which comprises the core of agency for complex agents such as ourselves. That such a process is at the heart of our agency is only a contingent fact about us: had we been constructed differently, such a process may have been unnecessary for action. Thus, imagine an agent who need not cause his goals indirectly, as it were, by traversing, as we most often must, sometimes complex causal pathways from our body to our goals. Imagine, that is, an ideal agent all of whose desires are desirable, and who can cause his goals directly, without the intermediary of means. Or imagine a lesser god who must cause his goals “through” the causal world, as it were, but whose intentions occur fully formed and goal-functional. Such agents would be spared the rigors of the practical rational process of intention formation and the malfunctions and dysfunctions of that process with which we must sometimes contend, together with the inhospitality with which the world sometimes greets our attempts at agency. Our predicament
is to be required to "construct" the goal-functional state which is both causally and cognitively sufficient to cause the change in the world which will cause the goal-events which motivate us. Thus, we distinguish a second goal which arises with every goal-event we would produce through agency. That necessary and antecedent goal is the goal of the practical rational process itself, which is to produce, through cognitive means, a state which may perform the tasks we have assigned to intentions. The construction of such a state is a project which may easily fail. Part of this chapter's project is to show how the process may fail in this regard and what steps agents may take to ameliorate or avoid such failures.

Imagine an agent who instantiates a process similar to the process described in the previous chapter, except that for this agent every step in the construction of a state which is fully equipped to cause this agent to "behave" is a purely non-cognitive one. Unlike the agent who has some control over which of his desires become goals, let us suppose that what this agent is caused to "do" is solely determined by the relative causal strength of his desires: it is his strongest desire that motivates him to action. This is to suppose, furthermore, that it is the

\[1\]
If 'construct' is an action term, then it may be misleading to say that agents construct their intentions. Intention formation is not an action, in the ordinary sense of the term. The process is self-sustaining, the result of the structure of agency and the functions of the constitutive agency states. Just as there are actions before which we need perform no other actions (Danto), there are intentions before which we need form no others. Although we may: see below for an account of the measures an agent may take to repair a failed intention. Intention formation provides the goal around which we may form intentions should the process itself fail.
motivationally strongest desire which becomes the subject of practical reasoning and which the system supplies with content about means and event of agency until sufficient identification of the causal route to the represented "goal-event" is reached, at which point causation occurs. Every step of this construction is simply determined, then, by the relative causal strengths of the relevant states. Which states get into the process is therefore not determined by beliefs, such as beliefs about desirability, truth and relevance. Now, whether such a causal process would produce goal-functional states would depend upon such considerations as how well the agent's functioning beliefs map actual causal relations among events, how reliable the process is at placing true and relevant content in appropriate places in the syntax or structure of the state under construction, and so on. Whether or not such an agent would be optimally or maximally goal-functional would depend solely upon the consistency and compatibility of his desires-to, since for him, the only determinant of goal is causal strength of his desires-to, rather than a judgement about compatibility and consistency. This would be an agent, that is, without cognitive means at his disposal either to address the possibility that some of its beliefs about causal relations may be false, or to prevent irrelevant content from functioning in the process, or to plan when and where its desires-to would be causal, in consideration

\[\text{\textsuperscript{2}This is not the \textit{cognitively sufficient} state of intention, since it has only identificatory completeness, and lacks the evaluative component of intentions which accounts for commitment.}\]
of satisfying more of, or the more important among, his desires. Furthermore, it would be an agent without the cognitive resources to address the agency failures we discussed in chapter 3: if all of the information-processing is “hard-wired,” its failures will be reparable only by repairing the hard-wiring. This would be an agent, therefore, that could be neither rational nor irrational.

In the first chapter I argued for a distinction between practicality and rationality. I argued that a process is practical just in case it is goal-functional. Thus, intention formation is a practical process just in case it produces its goal—an intention, which is a cognitive and causal state fully equipped for its rôle in action. For a cognitive process to be rational, I argued, is for it to be controlled by states which function so as to promote the goal-functionality of the practical process. Rational intention formation occurs when the content which enters and functions in the process is determined by beliefs about how to cause the goal which motivates the process. To vet content for goal-functionality wherever that is (contingently) possible, is a means to producing goal-functional intentions, and a means available only to those agents with states which may control the content of such processes. Vetting presupposes some standard in place with respect to which content is evaluated. The beliefs which govern intention formation will be guided by general beliefs about how to construct goal-functional intentions, and will include, for example, the belief that only justified, true beliefs about causal relations be allowed to function as means beliefs, together with a set of beliefs
about causal relations which meet these conditions. These overarching and
guiding beliefs will comprise a theory for the agent about how an agent of his sort
can be goal-functional: they will be beliefs which concern the means to goal-
functional intention formation. Thus rationality is the cognitive and evaluative
means to maximizing the goal-functionality\(^3\) of a practical process: it is the
functioning of beliefs about how to be goal-functional upon a practical process to
maximize the goal-functionality of that process. These governing beliefs will be
informed by a theory about the goal-functionality of the process so governed.
Where the process is intention formation, that theory will comprise an ideal of
agency.

In contrast to the agent who is merely caused to form states which are
causally sufficient for “behaviour,” imagine an agent who, because of a functional
bifurcation between his cognitive and active attitudes, must instantiate a process
which constructs a co-functional state—a state of cognitive and causal sufficiency
for action—out of the cognitive and causal means at its disposal. Suppose further
that those means include the ability to control the ebb and flow of content in that
process. That is, imagine an agent with desires and beliefs, one type of which has
the ability to determine what content gets into the process and what states
function in the process. Rationality of such a process concerns the function therein

\(^3\)Alternatively, it is the minimizing of non-goal-functionality of a practical
process.
of this governing belief: the process will be rational just in case the controlling beliefs allow to function only that content which passes evaluative muster—only that content which is believed to be goal-functional, and only those states which are judged to be goal-functional, according to the agent’s theory about how to produce goal-functional intentions.

Rationality is thus seen to be part of an ideal of agency. That ideal is seen to be at work in the agent who is in control of his agency states, and who is thereby in control of what he does. What makes practical reasoning rational is that it is controlled by belief, and guided by the goal of the process, which is to produce a goal-functional intention.

In our model of agency there are two aspects to control. The first is a structural feature of agency. Just as for the primitive agent, the occurrence of a motivational state in a rational agent prompts the supply of information from belief about whether and how satisfaction may take place. The second aspect of control itself has two features which distinguish rational from non-rational agents. First, for the rational agent, there is a certain structural feature which content itself must have before it is permitted to function in the process. That required structural feature of content is that it be licensed by belief. In our model, licence by belief is represented with our notation ‘Ev.’ The second feature of control which distinguishes the rational from the non-rational agent is the licensing function itself. Licensing of content, which we may think of as assent, presupposes
beliefs according to which content is judged to be appropriate to the process it would enter. The rational agent, in being rational, lets no content function that has not been licensed by belief.

We have argued that the function by which beliefs may affect other agency states is limited to adding content to, or subtracting content from, other mental states. Thus, the first appearance of an intention is the result of the addition to a mere desire-to of the evaluative content that the desired doing is desirable, an addition which marks commitment to the doing and prepares the active attitude for means-end reasoning. In the means-end component of practical reasoning, beliefs about causal relations supply causal information to the process, but only when that content is judged to be means to the goal, or to be the best means available. The distinctive rôle of evaluative beliefs, then, is to control, by granting or withholding positive evaluation, which other contentful states, be they beliefs, desires or intentions, function in the process—which is to say, add their content to the formation of an intention. Intention formation will be rational just in case those controlling evaluative beliefs control the content according to the agent's theory of goal-functionality for intention formation. Intention formation will be rational and practical only when those evaluative beliefs allow into the process only goal-functional content.

---

*Causal beliefs function as means beliefs when they are judged to be so.

*In general, a process is rational just in case evaluative beliefs control content according to the agent's theory of goal-functionality of the process being governed.*
Chapter 5

If rationality is a means to promote the goal-functional performance of agents, then rationality must involve, crucially and critically, the evaluative function of belief. Evaluation is a powerful antidote to the kinds of non-goal-functionality to which we may be subject, and it falls within the ambit of the belief which may "arm" or disable other agency states. Without such a filter on intention, an agent may find himself acting on mere desires and so thwarting more goals than he achieves. In general, rationality would require that there be states that let no state function the content of which has not been judged appropriate to the state it informs. Thus, a rational agent would have no belief function in his practical processes unless its content has passed some evaluative test for beliefs: unless, that is, the content has been judged to be believable and relevant. Similarly, because agents may have any manner of desires, some or all of which they have been caused to have, the rational agent would have no content function in desire—have no desire function, therefore—the content of which has not been judged to be desirable. And, perhaps most importantly, considering the consequences for goal-functionality in general, the rational agent would intend no action the doing of which has not been judged to be desirable.

Suppose that our agent has good reason, perhaps derived from experience, to believe that he is not an ideal agent. He may believe, for example that he may have false beliefs, or that he is not immune to error in selecting means from among his causal beliefs; or he may believe that some of his desires are
undesirable, while others are incompatible, and so could not be co-satisfiable. His desires therefore require selection if he is to satisfy more, or most, or the most important of them. He thus recognizes that the means he has available to produce goal-functional intentions is to let cognitive states function in the process only when their content is judged to be goal-functional. He will, if rational, let no content function that has not met the standards of goal-functionality. His goal is to produce a goal-functional intention; his means are to select from among his contentful states, those states the content of which he judges to be goal-functional according to some normative standards for believability of means beliefs and desirability of desires. It would be a condition of rational agency that only believable beliefs and desirable desires be allowed to function.

It is, then, a general condition of rational intention formation, that only tried or proven content be allowed to function, that only content which has passed evaluative muster be allowed to enter the process. Thus, while the evaluative function of belief—the gate-keeper for content—is essential to rationality, without further beliefs which inform and justify these evaluations, the evaluative beliefs could not perform their essential rational rôle. Those further and essential beliefs set standards for beliefs and desires and comprise a theory for the agent about how to be an agent, about how to be goal-functional. That theory will contain standards for believability and desirability which comprise the agent’s account of ‘justifiable’. These standards will be contingent and revisable in the light of
experience, and they will be what determines content of intention formation when agents are rational. An agent’s theory about how he should manipulate his contentful states so as to be a good agent may well differ from community standards for how agents in general ought to function. Justification will be relativized to whatever system of beliefs are operative. As a result, an agent may be subjectively rational but judged to have been irrational by a community’s standards. *Objective* rationality, tied as it is to goal-functionality, is the ideal to which both the individual’s and the community’s standards aspire.

Merely as a matter of engineering, a rational goal-functional system, built along the lines of the model in the previous chapters, will be a better agent than one that lacks control over the content and function of its states that the evaluative function of belief affords. A system which can supply and withhold assent to both beliefs and desires, according to its criteria of believability and desirability, and so determine which beliefs, desires and intentions it has, would be a better, because more goal-functional, system than one which lacked that function of belief. A system which can expunge non-goal-functional content and supply goal-functional content is a goal-functionally better system than one that cannot. And an agent that can revise his theory about how to maximize goal functionality, or to minimize non-goal-functionality, would be a superior agent to one that cannot.
That one's agency states are amenable to belief is thus seen to be a presupposition of rationality and irrationality precisely because without the ability to control the information transactions of intention formation, agents would be without the ability to control what content gets into the process, and so would be without the ability to intend, and ultimately do, as they judge they should. Furthermore, without beliefs about how to construct goal-functional intentions, the function of which is to guide or control the information transactions of agency, agents who sometimes harbour false causal beliefs, inconsistent or incompatible desires, and dysfunctional agency states, would be agents without the means necessary to be good agents in spite of their cognitive and connative shortcomings. That agents desire to be agents, and so desire the means to construct goal-functional intentions, is shown in the following. There is one desire which is presupposed by the occurrence of any "actionable" desire. It is something an agent finds desirable if he finds any doing desirable. That is the desire to be an agent, the desire to be a goal-functional intentional system, and thus an agent which produces goal-functional states such as intentions. The cognitive means to that goal is rationality, which employs a theory about how to manipulate the information at its disposal so as to construct goal-functional intentions; a theory, that is, about how to be goal-functional.

What the supreme value of rationality reveals is that belief is exactly that kind of state whose function is to serve goal-functionality. Belief is the supremely
adaptive state, sensitive to the world in which agents must act, and thus ideally suited to its rôle as cognitive means to goal-functionality. That is why belief is our best and bottom line in furthering goal-functionality. And that explains the intuition that the ideal agent is the agent in which belief reigns.

If rationality is an ideal of cognitive means to goal-functionality according to which all and only goal-functional content is allowed into the practical process, then irrationality occurs where those cognitive means are available, but do not function as means. Intention formation will be irrational, therefore, not because it fails to produce a goal-functional intention, but because, despite the presence of the goal to produce a goal-functional intention, and despite the availability of means to do so, the content which functions in the process does so non-ideally, bypassing, overriding or functioning in spite of those means. Intention formation is irrational, then, because content functions without the sanction of beliefs about how to construct a goal-functional intention. Irrationality in general is a departure from the ideals of rational goal-functionality. Where a practical cognitive process proceeds uninformed about how it goal-functionally ought to proceed, there is irrationality.

Given the sorts of agents we sometimes are, liable to false belief and errors in judgement, and capable of having dysfunctional agency states, it may seem that rationality would be a necessary condition for the process under belief’s control to be goal-functional. Is it also sufficient? If it were sufficient, then one could not
satisfy the conditions of rationality and be non-goal-functional. Where the process
is intention formation itself, if rationality were sufficient for goal-functionality of
that process, rationality would guarantee that a goal-functional intention would be
produced, which in turn would guarantee successful action. And a failed action
would, if rationality were sufficient for goal-functionality, entail that intention
formation had been irrational. Both outcomes are implausible, and therefore the
relationships between rationality, irrationality, goal-functionality and successful
action are not as straightforward as they might have seemed and need to be
addressed.

The goal of rationality in practical reasoning is to produce a goal-functional
intention. The means to goal-functional intention production is to evaluate the
processes of intention formation at every step or node which is amenable to the
belief function. Therefore, for every step of the process which employs the belief
function in the production of the subsequent state, there is a goal and a means to
that goal supplied by one's theory of rationality. Thus we saw the following
rationality requirements at work in intention formation. Given the causal rôles of
active-attitudes, it is rational, because goal-functional, to form desires-that only
for those events one believes are causable and non-actual; it is rational to form
desires-to only for those events which are non-actual and agency-causable.
Similarly, it is rational to form intentions to do only what one desires to do and
believes is desirable. And, in means-end reasoning, it is rational to select from
among one's means beliefs, that means which is believed to be the best means available.

Now suppose that an agent satisfies all the requirements of rationality in action, and thus has been rationally practical. All the content in process has passed scrutiny, and has thus been judged goal-functional according to the agent's theory of goal-functionality. However, despite functioning according to his own ideal of intention formation, suppose that the intention fails to cause an event of agency with causal relations to the agent's goal: suppose it turns out that, despite having been judged to be believable, one of the functioning means-beliefs was false. As a result, what the intention causes is an event of agency with causal relations through some non-means event(s) to some non-goal event(s). Thus, intention formation was both \textit{ex hypothesi} rational and impractical, or non-goal-functional. If this is possible, rational intention formation is compatible with impractical intention formation and, therefore, with unsuccessful action. An agent may satisfy the conditions for rational practicality, and therefore satisfy the goal or ideal of being a rational agent, but fail to act successfully: intention formation may be both rational and impractical.

As a momentary aside, it is worth noting that it would be a condition of rationality, specifically a condition of rational means-end reasoning, that agency failures, because they reveal that what was thought to be "ideal" was not, prompt appropriate belief revision. A mistake, which is an agency failure resulting from
the use of a false means belief, ought to cause the rejection of the false means belief culprit, with a corresponding revision of the agent's general causal beliefs. And, it would be a condition of rational agency that successful action reinforces or confirms or increases the commitment to the relevant operative beliefs.

There are two questions which rationality may require answers to: first, 'Is this goal-functional?,' and second, 'Is this the best way to be goal-functional?.' Both questions seem to be part of our conception of rationality. The second question is open, and serves as a constant test on one's theory of rationality. That the question as to the best remains open means that without the first, much less exacting and answerable question—the question of whether, for example, some particular course is, not the best unconditionally, but the best way to be goal-functional all things considered—rationality could not serve the purpose behind the concept. Without the first question, an agent could never get an answer to the question 'Is this rational?,' could never get closure on rationality, and belief would be deprived of its rational function. It is the first question, therefore, which allows rationality to be practical. That such a question govern intention formation is sufficient, then, for core rationality, and is not merely a compromise: without it, the goal in the service of which rationality has such great value for us, namely, goal-functionality, could never be achieved, since with only the second, infinitely open question, intentions would never get the guidance of a theory of rationality.
It cannot be a consequence of rationality that one never acts. Rationality must be satisfiable.

Rationality is belief functioning as governor or monitor over potentially goal-functional processes so as to maximize the goal-functionality of those processes, or to minimize their non-goal-functionality. Because it is informed by the agent’s theory about how to be goal-functional, what rationality requires of the agent will be contingent, revisable, evolving, conventional, open and historical. Although the standards may change, although what rationality requires of the agent will differ according to what beliefs and desires the agent has, and according to the world he inhabits, there is at bottom an unchanging concept which anchors rationality: rationality is not, as some theorists have argued, a purely contingent concept in all its aspects. It is bound as a theory about how to be goal-functional. The different and sometimes competing conceptions of rationality will, despite their differences, therefore all be theories about how causal, information processors such as we, may be goal-functional. Differences in what states rationality requires one to be in, then, will be due to differences, not in rationality, but differences in the contingencies of the processes over which rationality reigns—different beliefs, desires and intentions, various non-goal-functional

---

6 Was Hamlet an agent stopped by the open-endedness of the first question?
propensities, and differences in the world in which agents live. Goal-functionality is the commonality across competing theories of what rationality requires of one. Whatever beliefs and desires a theory of rationality endorses will be a function of goal-functionality given the other members of the set of beliefs and desires an agent has.

2. THE CORE OF RATIONAL AGENCY

Each step in our idealized account of the reasoning process which produces an intention was practical *vis à vis* the goal of that process, which is to produce a state which is causally and cognitively sufficient for action. What would make each such step also rational, given our account of rationality, would be if it were a requirement of the agent's theory of goal-functionality for that process. Thus, in our account we claimed, at various points, that a practical inference would be rational on the assumption that the agent was engaged in an activity the goal of which was to produce a goal-functional intention, and who had the means to promote that goal: namely, the ability to control the content of the process according to further, higher-order beliefs about what content in process would be goal-functional. For such an agent, that core practical process of intention formation would also be the core of *rational* practicality. In this section I will try to justify the claim that each practical step is a rational step. Some of these justifications will not take us beyond the minimal account of the agent with which we have so far been working. This is an account comprised only of those states
required for the agent to be engaged in intention formation under the guidance of beliefs about how to do so. However, some other justifications of the rationality of certain steps in the process of coming to intend will take us beyond the minimal agent and towards a richer agent, one with more goals than those to be satisfied by forming an intention and acting. This section, then, will begin our expansion of the ideal of rationality beyond its core or primary concern with satisfying the single desire which practical reasoning is in service of, and into rationality's further concerns with other and higher-order desires. This will take us to what more rationality may require of agents than the mere practicality of intention formation, a topic more fully addressed in section 3 below.

2.1 Proto-desire to Desire-that

Our analysis of the terms and relations of agency allowed us to trace the origins of agency back to the primitive state we called a “proto-active attitude” which is the source of motivation for agents and which served as a premise state to desire-that. The proto-desire was transformed into a desire-that with the addition of the content of the belief that the represented and motivational state of affairs is both non-actual and causable. The practicality of this inference is seen in the fact that the desire-that is a state which is causal with respect to what it represents, and it would thus be impractical to construct a desire-that for a state of affairs which was either actual or non-causable. What would make this a rational inference would be that it was believed to be practical by the agent who desires to be practical.
The rational principle which would justify such a conversion would be that an agent ought to desire—ought to be causal with respect to—only the causable and non-actual. Otherwise agents would find themselves in non-goal-functional, because unsatisfiable, states. What rationality would require of a state which was causal with respect to a state of affairs which the agent believed was not causable would be to remove such as state from the practical process: that is no place for an impractical state. A “parked” impractical state, such as a hope or a wish, ought, in rationality, to remain out of the agency stream, as it were, but to be “reactivateable” with the appropriate revision of the relevant beliefs.

2.2 Desire-that to Desire-to

In our account of intention formation, a desire-to is produced from a desire-that with the function of the belief that the desired state of affairs is not only causable, but causable by the agent—the belief, that is, that the causing is doable. The conclusion of this piece of practical reasoning is a state which may function as premise to means-end reasoning about how the agent may go about the doing. What makes it practical to desire to do only what one believes is doable is the impracticality of desiring to do what one cannot. What would make this restriction on desire-to formation rational would be the agent’s desire to be practical and, again assuming that the functional means to rationality are present, the belief that it would be impractical to desire to do what one cannot. Alternatively, an agent may have the weaker condition on rational desire-to
formation—that he should desire to do only what he does not believe he cannot. This is likewise seen to be rational, given the function of the desire-to state: it is premise to the search for means, and it would be impractical to search for means one believes are not available. It would be irrational because impractical to devote the resources of agency to the search for a causing which one believes is unavailable, and likewise irrational and impractical to fail to devote those resources to a causing one desires and believes is of a type which is available to agency. Thus it is rational to move to a state of being intentionally causal only with respect to agency causings, causings available to the agent.

### 2.3 Desire-to to Intention

An important node of practical rationality is the first occurrence in that process of an intention, a state which carries with it commitment to a doing. In our account, commitment occurs as the result of the function of the belief that the desired doing was desirable, a judgement made in the light of the agent’s beliefs about what sorts of actions he should perform and about what other desires could be causal in its stead, according to which what he intends is what he ought to do. The effect of such a judgement is to commit the agent to the doing he only theretofore desired. The ability to select from among one’s desires which will be efficacious of action, and which not, is at the crux of freedom and autonomy for agents. In our account, we marked the (processionally) earliest point at which an intention could be formed, where this required only the desire to F and the belief that F-ing was
desirable, and where nothing further need be believed about the F-ing. This is not to say that one could not come to intend to F from a desire-to whose content was much more complete, say, even with all the means already in place.

Why is it practical to intend to do only what one desires to do and judges to be desirable? Why isn't desiring to do something sufficient for the practicality of commitment to doing it? The answer lies in the nature and origin of desire. For any system like us, caught in the causal impress of the world we inhabit, we will have among our desires those which we have been caused to have. If so, there can be no consistency requirement on one's desires, there being no consistency requirement on causation: an agent may be caused to have incompatible, contradictory and undesirable desires—those which are inconsistent with his beliefs about the sort of agent he would prefer to be, and thus inconsistent with the sort of desire he would prefer to have. Furthermore, desires differ in causal strength, and their strength may vary quite independently of their conduciveness to overall desire satisfaction: an agent may have very strong desires which, if causal, would prevent many or all other desires from being satisfied. For such an agent, it is practical to become committed only to those doings he judges to be desirable according to what else he desires and believes. A simple agent with only one desire may, of course, be practical without such considerations.

We may assume, then, that any agent complex enough to be of interest to us here will have multiple desires that vary in causal strength and which may be
incompatible with one another. We may also assume that the complex agent has preferences regarding his desires. Thus, there will be among his desires those which he judges to be more important than others, and these judgements will be independent of the desire's inherent strength. His preferences will reveal themselves in which desires would be favoured in cases of conflict, and which would be sacrificed in favour of more important desires. Thus we have (slightly) enriched the agent by giving him a complex of goals surrounding action: he not only has goals in the pursuit of which he will (overtly) act, but he also has goals regarding which of his actionable desires will make their way through to action, which will be held back or sacrificed in favour of the first, and which desires may never see the light of day.

What would make it possible for our agent to satisfy these higher order desires would be if he had the means to determine which of his desires would be efficacious, and when. Our agent has beliefs which will turn this trick. Thus, given his higher order goal to act on some desires rather than on others, it would be practical vis à vis this goal, and therefore rational, for the agent to vet his desires so that only those which are judged worthy of satisfaction, according to his own lights, are allowed to proceed toward action. That is, it will be rational, because functional with respect to this higher-order goal, for the agent to act only on those desires he judges to be desirable, rather than for his desires to be causal without the discriminating mediation of belief. If desires-to were left to be causal
of behaviour without having to pass through some filter, such as a judgement of desirability, then an agent may act so as to satisfy a desire which may thwart other desires, some or all of which may be ranked higher by the agent than the efficacious desire. For such a complex agent, it is goal-functional to act in ways which maximize the number and desirability of those of his desires which are efficacious of action.

2.4 Means-end Reasoning

With the formation of an intention, means-end reasoning has its first and enduring direction: find a means or causal route from the body to the goal-event. We claimed that an intention would, in a practical system, first prompt a search among the agent’s beliefs about causal relations for those which would be means to his goal, and then, if necessary, to select from among the set of available means, that means believed to be the best. The search for means is clearly practical, and it would therefore be rational for the means search to be controlled by the agent’s beliefs about causation. For the search for the “best” means to be practical and rational requires another enrichment of the agent. We need another goal, in addition to the goal of producing a goal-functional intention, for the reasoning to be functional with respect to, in order for the search for best means to be practical. Suppose, for example, that ‘best’ here means most efficient. That would imply that our enriched agent has the desire to act cost-effectively, to get the most return on his agency investments. Imagine, then, an agent who desires to get the most
desire satisfaction for the least amount of effort. This is an agent with a goal to be agency-cost effective. Means-end reasoning will serve this goal of his when it selects from among the possible means available to him, those which are the “best,” because most efficient, means available to him, and otherwise not. An agent with the standing desire to be efficient, and with the means to be so, namely the ability to determine which of his means beliefs functioned in the process, and with the belief that to so select from among available means was how he must reason if he is to be functional with respect to the goal of efficiency, would be an agent which was rational for doing so. The goal of cost effectiveness thus impinges on means-end reasoning: it is a goal which concerns how means-end reasoning is done. Reasoning will be functional with respect to this goal only if it selects a means which would produce the agent’s goal with the least cost to the agent. Thus this goal, and others like it, create more rigorous standards for means-end reasoning than truth.

These same considerations make it rational to vet the process whenever content is added. Thus, for example, the question of desirability remains open throughout the process.
3. RATIONALITY BEYOND THE CORE

3.1 Thickening the Agent

The rôle for rationality will reflect the complexity of the system in which it operates. What rationality requires of an agent, in addition to what it requires of one type of his reasoning, will be relativized to the agent’s nature and his circumstances. In general, what rationality requires of agents, and, thus, the ways to fail to be rational, will vary with the complexity of the agent, his desires and beliefs, the contingencies of his nature, including the sorts of non-goal-functionalities to which he is susceptible, and the means he has to ameliorate those failures, and the degree to which the world in which he acts is friendly to his pursuits. The simpler the agent, and the friendlier the world, the less room there will be for impracticality, and so the less exacting rationality’s demands may be. We are complex agents. We have multiples of goals, and we have goal-producing beliefs about which (sort of) goals we ought and ought not have; we have goals about how to pursue and achieve goals, about minimizing cost and maximizing effectiveness, about maximizing satisfaction and minimizing frustration. Each of these goals generates its own goal-functional process, and with it, its own portion of rationality. It will be on the hypothesis that we have an agent of such complexity that certain of the more interesting forms of irrationality occur.

So far we have considered what rationality amounts to for stretches of the core of practical reasoning. We now want to expand the rôle of the ideal of
rationality to include more of agency than the abstracted process of intention formation. Our analysis of rationality sees it as part of a larger ideal we have for agency. According to this view, the ideal agent is the agent in which belief reigns. Belief is the agency state ideally adapted for its rôles in rationality: it is the perfect (cognitive) means to goal-functionality. We can therefore expect an expansion of the rôle of rationality beyond the core process of intention formation to require an increase in the scope of belief's control of the content of an agent's contentful and causal states. With each increase in the rôle of belief comes a condition of rationality; and with each such requirement comes the possibility of failing to be rational. To see what more rationality may require of us, we need to enrich our working conception of the agent.

In addition to core rationality and depending upon the sort of agent one is—depending, in particular, upon the agent's proclivities for non-goal-functionality—there will be a set of steps or measures that it is rational for an agent to take, given that he has the goal that intention formation succeed in producing a state which is causally and rationally sufficient for action: given the goal, that is, that he be an agent. The nature and extent of this set of requirements on agents comprises a theory of what in the Law is called the reasonable man, a theory about what standards of agency we may be entitled to require of each other. The reasonable man is a standard against which we may be measured, and the weight of the Law may be brought to bear to compel agents to measure up to that ideal.
With the ideal expressed by the theory, and the means to that ideal set out in the theory—the so called “standards of care”—there is a set of non-goal-functionalities generated. These include all the ways one may fail to live up to the ideals the concept expresses. Failing to meet the standard of care set out in the theory of the reasonable man is itself a goal-functional failure; not necessarily a failure to get the goal of some action—not, that is, necessarily an impracticality with regard to the goal of action, but a failure to meet the goal of the ideal for the sort of agent one is. A failure to meet the standards of the reasonable man will be an irrationality beyond the core.

We need, then, a thicker conception of the agent. Thickening of the “first order” will have us add content and function which bears on the practical process itself. This minimal thickening will be enough for us to generate an account of what more rationality requires of agents than the cognitive control of intention formation itself. One could go on to add modules of content which would support theories of prudent or moral behaviour, for example. And one may have arguments to show that, given certain other goals of the agent, just such content is required by rationality. That, however, is beyond my present purposes.

As a first step in this thickening of the agent, let us suppose that our agent is self-conscious, that he has the ability to have true beliefs about himself. Among the self-conscious agent's beliefs, we would expect to find beliefs about his other beliefs, beliefs about his desires and intentions, about his preferences regarding
his desires, about his hopes and wishes, and so on. We would also expect to find the belief that he is an agent, together with some conception of what that means: we will assume that our agent believes that he has the ability—believes that certain of his states have the function—to control the content and function of his other agency states. Assuming that he has desires that are only satisfiable through action, we may also expect to find the desire that he be a good agent, one who is skilful at satisfying his desires.

Let us also assume that our agent has beliefs about how good an agent he is, how well his agency processes take intentions through to action. He will therefore have beliefs about his potential for false beliefs, beliefs about the compatibility and realizability of his desires, and beliefs about any state dysfunctions he may suffer from, such as reticence and recalcitrance, which may thwart or frustrate agency. Let us also suppose that this thickening of the agent takes us as far as the ability to theorize about means he may have at his disposal to ameliorate non-goal-functionality. Among these means there will be those fundamental cognitive measures we have described as part of the ideal of the rational agent: the control, by belief, of the content and function of any practical, goal-functional, process. What other means rational agents have available to

\^[8] Perhaps there are agents who, while self-conscious and aware of the world they inhabit, are unable to form beliefs and theories about their successes and failures: perhaps this is true of children at some early stages in their development.
maximize goal-functionality and to minimize non-goal-functionality will become apparent as we proceed.

As we have seen, the very least rationality will require of our agent is that he monitor his agency for those forms of non-goal-functionality he may correct. Thus, in rationality, the agent would form beliefs about the process and vet content for goal-functionality. The rational agent is one in which no content functions that has not been judged to be goal-functional. The question before us now, is what more, beyond the core uses of rationality, would be required of such an agent in the way of measures to minimize non-goal-functionality. Earlier, with the justifications of the requirements that agents intend only what they find desirable, and select as means only what would be most efficient, we saw the beginnings of what more rationality would require of the enriched or thickened agent. Both justifications adverted to further goals: the goal of satisfying more highly ranked, over less highly ranked desires, and the goal of satisfying desires with the least cost to the agent. In general, what more rationality requires of our agent will be a function of what other goals he actually has as well as those the current theory of the reasonable man requires him to have.

Thus suppose that our enriched agent has the goal of satisfying as many of his actionable desires as he can. One means to this goal we have already encountered: that was the requirement to intend only what one judges to be desirable, a means to avoid actions which would thwart other desires. Another
means to the goal of maximal desire satisfaction is *planning* which allows one to satisfy otherwise incompatible desires. Rather than permanently sacrificing one desire for the satisfaction of another, incompatible desire, planning to do A first, say, and then B, is a means to doing both. If so, planning would be practical *vis à vis* the goal of maximization, and a requirement of rationality for an agent equipped with the goal and the cognitive means to it.

If our enriched agent is to approach anything like the complexity of the adult human agent, he will have desires about what sort of events in general occur in the world, since worldly events may impact on him. He will also, therefore, have desires about what sort of events occur in the world *as the result of* his actions. An action with fewer undesirable side-effects and/or end-effects would be preferable to one with more, all else being equal. It would be rational, therefore, for effects of actions to enter into the calculation of the desirability of a proposed course of action, in much the same way that efficiency of means did. That the question of the desirability of an action remain open throughout the process, as we claimed it practically and rationally must, allows effects which are brought to light by the agent's beliefs about causal relations, to impinge on what agents do.

An agent's sense of self—the belief that *he* has an identity across *his* actions—generates further normative considerations for actions, and therefore an extension of the rôle of rationality. This sense of an enduring self who authorizes its actions over time creates its own ideal: each prospective action is an action of
his, not merely an isolated action, and thus it ought to be evaluated in this context. Each action will be an index of the sort of agent he is, and something for which he may be held responsible.

In order for desires about effects to function in the reasoning process there must be some effects to judge. Thus it will be functional with respect to the goal of minimizing undesirable effects of actions for an agent with this goal to search his beliefs about causal relations for effects of events he would cause. Failing to do so would be impractical and irrational. How far down the causal pathways an agent is required by rationality to search before concluding upon the overall desirability of action is a contingent matter, and would depend upon such considerations as agent’s abilities to so reason, time and urgency, predictability, and so on. When an agent ought to stop thinking and start acting will be a matter for an evolving theory of rational agency. Similarly, what sorts of events ought to count as undesirable, and what weight an agent ought to give certain sorts of undesirability, is a matter for a fuller normative theory.

An agent such as this, who is bound by a standard of care with respect of effects, whether by his own theory of how he ought to act, or by his community’s standards of the reasonable man, may fail to meet such standards in a variety of ways. Thus suppose an agent fails to identify effects of his action when he ought to have, and suppose that his action produces untoward effects. The diagnosis of

---

9 That he ought to have considered effects removes from our consideration here extreme cases of urgency or necessity in which the extent to which an agent is
exactly why he failed to identify effects remains to be made, and would bear on our
description of this departure from the ideals of practicality and rationality, but
even without it, we may, on the assumption that he ought to have considered
effects, charge such an agent with that type of negligence called carelessness in his
reasoning about effects. His reasoning was non-functional with respect to the goal
of minimizing harmful effects produced through exercise of his agency, and given
that he had this goal, and the means to it—namely, the function of vetting content
in process for harmful effects—his reasoning was impractical in this respect, and
he was prima facie irrational for failing to do so.¹⁰

Alternatively, suppose that our agent has satisfied the epistemic dimension
of reasoning about effects and has identified undesirable effects of his actions.
Suppose further that, by his own normatives, or by the normative theory of his
community which is embodied in its theory of the reasonable man, the
undesirability of those effects ought to have had the agent judge the action, tied as
it would be to those effects, itself not worth doing. Nevertheless he acts. He is ex
hypothesi not guilty of failing to search his beliefs for undesirable effects of his

¹⁰The reason for ‘prima facie’ is that the diagnosis of the failure may have us change the charge: if the desire upon which he was causal was compulsive or garrulous and prevented the agent from considering effects, we may, if we judge that he ought to have been able to control that state, add a charge of negligence with regard to his own state-dysfunctions. Not only would he be careless about consequences, he would be negligent with respect to the control of his own aberrant agency states. More on this below.
action, so he is not guilty of that sort of carelessness in his reasoning about effects. He is, however, guilty of failing to consider their undesirability in his calculations of the desirability of his action, or of failing to give their undesirability due consideration, or proper negative weight. As in the previous case, the diagnosis of this failure remains to be made, but on the assumptions that he believed there would be undesirable effects, and that he had the means available to avoid those effects, he is, prima facie, guilty of the charge of recklessness, of not giving appropriate negative weight to harmful events he was aware he would cause.11 His reasoning was non-functional with respect to his or his community's goal of minimizing harmful effects of actions, and he was therefore irrational for the omission.

Suppose that our agent is bound by morality, either because he inhabits a community that so binds its members, or because he has his own theory about what sorts of actions are acceptable and unacceptable on moral grounds. His practical reasoning will be functional with respect to these goals, self-imposed or otherwise, just in case they play some rôle in the outcome of his reasoning. Suppose that our agent takes the morality of his community into account—allows its goals to weigh on his calculations of what he will and will not do—not because he shares those goals, but because he desires to avoid the sanctions others may

11 Recklessness is a charge against one's normative reasoning, or lack thereof; carelessness is a criticism of one's epistemic performance.
impose for behaviour contrary to their moral code. If so, his reasoning would be practical with respect to the goal of avoiding punishment— it would be prudential rather than moral—and he would be rational for so reasoning. In the alternative, should our agent share the community’s moral theory, or have one of his own, his reasoning will be practical with respect to the goals of the theory— say, minimizing harm to others—and not merely practical with respect to avoiding its sanctions, just in case that goal impinges on his reasoning about what he will and will not do. For the moral agent, it is rational to reason morally and irrational not to.

We can see, by the addition of modules of content, what more rationality may require of agents than to govern the core of practical reasoning for goal-functionality. With every additional goal that bears on reasoning there will be a requirement of rationality: serve that goal by the cognitive means available. The general instruction given to rationality, whatever the specific normative it serves, is to vet content and function for desirability according to that normative. Rationality is the exercise of control based on an instruction from normative beliefs or ideals for the process being governed. Which normatives, beyond the goal-functionality of the core process of agency, rationality is brought to service will be a function of the specific “thickenings” of the agent.

The feature that would make intention formation rational can be generalized for any practical process, whether it be intention formation itself, means-end reasoning, selection of goals, or belief formation and revision.
Rationality is the exercise of control over the content and function of a process, based on an instruction from a normative belief or ideal regarding the goal of the process itself and any instructions from normatives or ideals for how the goal of the process is to be achieved. These ideals comprise the ideal for the process, and they give rationality its direction. Before rationality can do its work, therefore, the normatives for the process must be in place.  

But the normatives alone are insufficient for rationality: it is possible for an agent to have a full-blown theory of ideal agency, but be unable to translate those ideals into action. For example, an agent may have the belief that he ought to minimize undesirable consequences of his action, and yet that belief may not operate as it must in order for the practical reasoning process to have been governed by it. An agent who lacks the ability to control his agency according to his ideals may nevertheless be a good critic of himself and others, but he would be an agent who would forever fall short of his own ideals, except possibly serendipitously. What rationality needs besides its normative dimension is the means by which those ideals may impact on agency. That means remains constant across whatever process rationality would govern. Rationality's means is the employment of belief in its evaluative capacity in service of the ideal for the

12 Those theorists who believe the rationality is, at bottom, a purely contingent concept, may have been misled by the contingency of the ideals that function within rationality. If so, they have identified only one, albeit essential, component of rationality with the whole of rationality, and overlooked the common ground across all particular theories. That common ground is goal-functionality, whatever the particular goals may be.
process it governs. Thus rationality presupposes normative beliefs or ideals and the means to govern that process supplied by the evaluative function of belief.

The question ‘Why be rational?’ appears to have an obvious answer: whatever goals for his cognitive processes an agent has, rationality is the best means to achieve them. Notice, too, what this account of rationality does for the question ‘Why be moral?’. We may presume that the question would arise only for someone already embarked on the project of living among other agents, and questioning the value of morality towards that project. Morality may be seen functionally, as that theoretical normative which is meant to govern agents so as to produce the sort of conduct which is best conducive to agents living together. A particular moral theory will be a set of ideals of action around which belief may function so as to produce the sort of behaviour which is judged to be the most conducive to agents living together. Morality, then, just is an exercise of rationality.

We now want to return to the systemic state dysfunctions described in chapter 3 to consider what rationality may require of a thickened agent who (truly) believes that he is not immune to those state dysfunctions. Given how powerfully non-goal-functional the state dysfunctions may be, agents with the goal to be agents and the means to ameliorate or avoid those non-goal-functionailities, will be required by rationality to exercise those means to minimize the effects of the state dysfunctions on agency.
4. STATE DYSFUNCTION IN THE PROCESS: A NOSOLOGY OF AGENCY, PART 2

In the preceding chapter we analyzed the process of intention formation in terms of the individual causal and informational functions of the agency states, the active and the epistemic attitudes. Each of those functions is essential to the construction of a state which may play the roles we have assigned to intention, and the account of the process was an idealized description in which each of the functions occurred as it must for the overall process to be goal-functional. However, each of the agency states is susceptible to the fundamental failures we identified and described in chapter 3. These are failures which may occur at any point in the process of intention formation, and should any one of them interrupt that practical process, it would be sufficient, if left unameliorated, to prevent the goal of that process, namely the production of a state which is cognitively and causally sufficient for action. Each of those dysfunctions is a dysfunction precisely because it would amount to non-goal-functionality of the agency process. An agent whose agency states were dispositionally dysfunctional in any of the ways identified above would thus be an agent faced with the added burden of attending to those dysfunctions if his agency process is to culminate in a state which is sufficient for action. Therefore, much more in the way of his own resources will be required of an agent whose beliefs are not always reliable indicators of facts, and whose agency states may be dysfunctional, if he is to be goal-functional. Part of the project of this and the next section is to identify the means that agents have
available to overcome their own non-goal-functionalities, and to distinguish those which are ideal means from those which are something less than ideal.

It will be obvious and perhaps familiar to the reader how many of the agency failures from chapter 3 would disrupt agency. False or mal-formed content in the process is obviously non-goal-functional: one cannot get one's goal with false means beliefs. Thus we may expect to find within the theory of what it takes to be an agent, means which specifically address the possibility of false belief, such as a set of believability conditions on content. An agent whose intentions were recalcitrant, and so resistive of revision, would be an agent much less likely to be goal-functional than an agent whose intentions were always open to amendment by belief. And so on, for each of the other dysfunctions. Each such agency failure would pose a problem it is the job of an agent's theory of how to be an agent to solve. That theory or set of beliefs will address how the agent may ameliorate or avoid the sorts of failures to which he may be susceptible. Answers to such questions will be means to goal-functionality and thus become a component of the agent's theory of rationality, a theory which comprises his ideal for agency.

\[13\] Particularly if one holds a pragmatist, i.e., goal-functional, view of truth itself.
Below are some salient examples of how intention formation may be rendered non-goal-functional by the dysfunctionality of its states. It would be the task of rationality to address these dysfunctions, where possible.

### 4.1 Recalcitrance

Given the importance to agency of information transactions, both epistemic and motivational, recalcitrance to such transactions would be strongly non-goal-functional. Consider the agent who has, say, intentionally washed his hands, and so has caused his goal in just the way he intended. Imagine that the conditions which are standardly sufficient for an agent to come to believe that he has so acted are present. But suppose that our agent suffers from belief recalcitrance, the result of which is that the belief that his goal has occurred is not formed. The belief that a goal has been achieved through action is essential to cognitive intention satisfaction. Its rôle is to render the intention non-motivational by cancelling certain content necessary for motivation: specifically, the belief that causation is necessary to make the goal actual. An agent who suffered from belief recalcitrance, and so failed to form this crucial belief would be an agent left in a motivational state with respect to an event he has already caused, and thus an agent in a state which could not be satisfied.

In the alternative, suppose that our agent has again acted intentionally, but that in this case, the occurrence of the goal does cause the belief that the goal has occurred. Ideally, the belief that one's goal has occurred would, because it entails
the falsity of the belief that the causing of the goal is necessary for its actuality, cancel that content of the intention, and, in effect, reduce the motivational strength of the intention to zero. But suppose that, in this example, the agent’s intention to wash his hands is dysfunctional in being impervious to belief, perhaps impervious to a particularly important and relevant belief, such as the belief that he has just washed his hands and so need not do so again. If so, that belief could not have the effect on this intention that, in rationality, it ought to have; namely, the extinguishment of that intention, a normal or standard effect of successful action and one of belief’s functions in intention-satisfaction. If we have not mis-described the case, so that it is truly a dysfunction of intention and not one of belief; if, that is, the relevant belief has done its work, but without the practical effect it ought to have had on the intention, then the failure is due to the dysfunction of the intention. The intention remains unaffected by the belief that it has caused what it represents, and the agent is left, irrationally, in the non-goal-functional state which is causal with respect of an event which is actual: a state incapable of satisfaction.

Recalcitrance at this stage of agency, whether it be recalcitrance of belief or of intention, may be implicated in some forms of obsessive behaviour, or of so-called obsessive-compulsive disorder. Some such cases may be explicable as those in which an intention (or desire), despite having been successful, retains its causal or motivational strength because of the agent’s failure to uptake information, due
either to the recalcitrance of belief or of intention. Recalcitrance may also explain
the ideologue’s beliefs that survive the strongest evidence to the contrary.

4.2 Incontinence

‘Incontinence’ is the name we gave to the state dysfunction of the inability to
contain content. For an agent who instantiates a process of intention construction
which proceeds by adding content to an intention which concatenates it,
incontinence could mean the premature termination of a process which would
otherwise have terminated in a fully-formed intention and action. Thus, suppose
that intention formation has all but completed the construction of an intention: all
that remains is the “planning” consideration of when the intention ought to be
efficacious. Now, suppose that the agent believes that the occasion for action has
arrived. He would thus have some description of the event of agency which would
particularize it, and this description, say, ‘\(E_4\)’, ought, in practicality, to replace ‘\(E_3\)’
in his intention. This is the event of agency he ought to cause. Suppose that the
identificatory belief is incontinent, and that before it can pass its content to the
waiting intention, it loses its content. Since that last part of the identification of
the event of agency is necessary for action, the intention could not be causal.

As a one-shot dysfunction, belief incontinence at the point penultimate to
action may mean a missed opportunity. But an agent with dispositionally
incontinent beliefs would be handicapped in the extreme.
We should notice that this belief dysfunction should be distinguished from another dysfunction which may make its appearance at just this place in intention-formation. We may imagine an agent for whom a belief in the imminence of action causes enough interference in the process to prevent that action. A dispositionally timid agent, afraid to act at all, may be describable in this way: the belief in impending action may produce enough fear to disrupt the practical process altogether.

Since content concatenation is essential to intention formation, an intention which could not contain its content, could not proceed through the process to action. So, suppose a fully formed intention, one with all the identificatory and evaluative content required for intentional action. Since it is complete, it would qualify as an immediate intention, as distinct from a future intention, since completeness entails that it is to be causal here and now. Now suppose that, at the point of causation, this incontinent intention has content “leak” away: suppose, that is, that it simply loses identificatory or evaluative content. In an ideal system, such a loss of content would prevent efficacy, since a condition of efficacy in an ideal system is cognitive completeness. Should this dysfunctional intention be efficacious nevertheless, its lack of sufficient content increases the

\[\text{It is possible that cases like this have been mistaken for akrasia or weakness of will. If everything necessary and sufficient for action is believed to be in place, but no action is forthcoming, the only alternative explanation would seem to be failure of the causal function of the intention.}\]
likelihood of failure: if the intention loses content which identifies the causal route to the goal-event, the probability that it will cause the event of agency which will initiate that causal route diminishes.

It is obvious that incontinence of either function, whether episodic or more global or dispositional, would be a serious impairment of agency.

4.3 Garrulousness

One of the control functions we have assigned to belief in rational agency is to determine whether what is believed is in fact believable, and whether what is merely desired is in fact desirable. That is, one of the functions of rationality in agency is to set justificatory conditions on the content of agency states which must be satisfied before states with that content perform their functions as beliefs, desires or intentions. These evaluations of content by belief mark the junctures between states which are prevented from functioning in practical processes and those states which are allowed to function. It is plausible to argue that one of the properties which distinguishes the rational from the non-rational agent is the ability to control under what conditions his states will function as agency states. This is to place the expression function of a state under the control of belief.

Suppose, then, that positive evaluation of content is a necessary condition for states to function on that content.\(^{15}\) If so, the presence of doubt about the content

\(^{15}\)It is a lamentable fact about agents like us that we form beliefs with false content and desires that are unsatisfiable, etc. For such an agent to be goal-
of a belief ought\textsuperscript{16} to be sufficient to prevent the belief from performing its informing rôle. Otherwise doubt would serve no practical purpose: an agent whose doubtful belief functioned just as his non-doubtful beliefs would be an agent much less likely to be goal-functional than is the agent in which doubtful beliefs were prevented from functioning as beliefs by the filter of believability. In trying, which we have argued\textsuperscript{17} differs from doing *simpliciter* due to the presence in trying of a functioning doubtful belief, doubtful beliefs are allowed to function, despite their doubtfulness, by virtue of the additional judgement that the doubtful belief is, in the circumstances, the best belief available.

Similarly, we have argued that a condition on the conversion of a desire-to to an intention ought to be that what is desired is judged to be desirable. Without that restriction on intention formation, agents would form intentions which may have them act in ways that frustrated more goals than they achieved. We may think of these conditions as amounting to *commitment* to the content of one's states, a condition which is necessary for them to function in agency. It would be a dysfunction, then, for a state to express its content without the *imprimatur* of believability or desirability. For a state to so function is for its functionality to be functional, or perhaps maximally goal-functional, requires that one's states be vetted, be held to some standard to ameliorate their potential non-goal-functionality. This is one of the means to being an agent, for agents contingently like us.

\textsuperscript{16}The goal-functional 'ought'.

\textsuperscript{17}See *Agency In Action*, chapter 1.
unamenable to belief. We have called this state dysfunction *garrulousness*. Our folk-psychological term for a garrulous desire is *compulsion*—a motivational state which is strong enough to resist control by belief, and which therefore may be causal without having been judged to be desirable, or in spite of having been judged to be undesirable. The folk-psychological term for a belief which enters action but is strong enough to resist control by other beliefs, is *obsession*.

Suppose an agent, then, who, on the strength alone of his garrulous belief that there are people out there who are trying to harm him, acts in those ways we associate with paranoid disorder. For the belief that there are plotters everywhere to be garrulous is for it to function on its strength alone, and without being subject to the believability test. As the old joke would have it, just because this belief of his is dysfunctional in this way does not mean that it is false. But an agent with beliefs which function as beliefs without being subject to the filter of believability would be an agent much less likely to be goal-functional than the agent who acted only on content which had satisfied the tests of believability.

Suppose an agent whose desire to wash his hands is garrulous, and so has effects in agency that standardly only intentions have: namely, the prompting of practical reasoning about how, and action itself. If positive evaluation of one’s desires is a condition of *intentional* action—one of our action ideals—then behaviour caused by garrulous or compulsive desires, in circumventing the desirability test, would be something less than intentional actions: compulsive
actions are not fully rational and would therefore not be intentional. An agent with compulsive desires is an agent much more likely to act so as to frustrate more of his goals than he may satisfy than is the agent who intends to do only what he judges to be desirable.

4.4 Reticence

If, as we have argued, belief supplies content which serves rôles in agency other than the handmaid of desire—for example, belief’s evaluation rôle in the selection of goals—then reticence could result in an agent failing to intend to do what he believes is desirable. If it is a condition on a desire’s going forward in practical reasoning that it be positively evaluated, that it be judged to be desirable, then should that judgement be reticent, the desire which otherwise would go forward would not. Furthermore, imagine the case of the reticent judgement that what is desired is undesirable. Reticence may render the judgement incapable of preventing a (strong) desire from advancing toward action. This case is to be distinguished from the case of the garrulous or compulsive desire which proceeds toward action due to its causal strength alone, and irrespective of any judgements as to its suitability for action. When the culprit of non-goal-functionality is reticence of belief rather than garrulousness of desire, what explains the “progress” in the reasoning process, and what may explain the eventual action, is

---

18 It’s not real progress, but movement through the process despite insufficient evaluative content.
not that the desire was too strong, but that the belief or judgement was too weak to perform its vetting rôle. This, as distinct from compulsiveness, is what we would call incontinent agency: it is a control deficit, not an excess of causal strength, and is the inability of belief to "amend" those states it ought to. Reticence of belief may be implicated in incontinent action, cases in which the belief function is too weak to control desires which cause "actions." If impulsiveness is the condition which leads to action because of the lack of, rather than the weakness of, belief's vetting rôle, then reticence of belief may be implicated here too. Impulsive action may be a case in which the relevant vetting beliefs were so reticent as to fail to function altogether: the occurrence of a desire or "impulse" is not followed immediately by an application of the desire-vetting function, and action without evaluation—impulsive action—occurs.

Reticence may also affect desires and intentions. A crux of intention formation is the point at which the information supplied by belief comes to constitute cognitive sufficiency for action. The effect of cognitive sufficiency is to "trigger" or let loose the inherent causal power of the intention to cause what it

---

19 The infelicitousness of having chosen the term incontinence for the state-dysfunction described as the inability to contain content, when "incontinence" is used to describe lack of control of reason over process, requires that I indicate, in some way, which of these different non-goal-functionalities I am referring to. Incontinent agency or reasoning may, in fact, be due to the dysfunction we have called belief reticence, not belief incontinence. Thus, when referring to a process failure, such as the failure of belief, in its rôle as governor of a practical process, to "contain" or control what it ought to govern, I will use such terms as 'incontinent agency' or 'incontinent reasoning.'
represents. In action, the intention causes an event of agency with the intended causal relations to the goal. Suppose, then, that an intention to kick the cat is otherwise fully-formed and standard; that it has been informed by belief to the effect that it is both causally and cognitively sufficient for action. However, when the time to act arrives, the intention simply and utterly fails to be causal of the kicking: it is a state which fails to perform its expression function. If we have supposed that up until the time to act the intention was standard, or functional, or normal, that would be to have supposed that those properties which enable active attitudes to cause what they represent—those properties which explain that function—were present, but when the time for action arrived, the causal strength of the intention simply vanished: this is, presumably, a state which had, but no longer has, causal sufficiency for action. If this sort of failure is plausible, it counts as a pure functional failure, and a case of irrationality due to a dysfunction of the state. If we have not misdescribed the case, so that there remains in place the judgement that the kicking is the thing to do here and now, that there has been no change of mind, no cognitive change, no reason, that is, for the lack of action, then the agency malfunction may be attributed to a dysfunction of the intention, clearly not now one of recalcitrance, as above, but a pure causal failure, a pure diminishment of motivation, perhaps a pure weakening of will,\(^\text{20}\) due to the dysfunction of reticence.

\(^{20}\)If will may be identified with \(\rightarrow\).
A word on the plausibility of this sort of intention dysfunction is in order. What could explain the waning of motivation, the pure causal diminishment of an intention? This is a failure that may occur at any point in the process. It is also a failure of a type that may affect desires and beliefs as well. If it is a pure functional failure, then, *ex hypothesi*, the loss of causal strength was not due to some cognitive change in the agent; not due, for example, to a change in his judgement about the desirability of the kicking. We want to distinguish the pure dysfunction from the cognitive change: distinguish, that is, a case in which fear, for example, of an envisaged reprimand changed the agent’s mind—a case of belief, or perhaps some active attitude, functioning so as to change the desirability of the action. Therefore, we must presume that the cognitive state of the agent remains constant. Now, if active attitudes are natural causal states, then, since flux is a feature of the natural causal world, there seems to be nothing to prevent the idea that a desire or intention may be caused to weaken or lose its causal strength altogether, not for any reason, but due simply to some brute causal matter of fact. If the causal powers of agency states are, like other causal powers, subject to the influence of brute causation, one effect of which may be to diminish the causal strength of an intention to zero, then we can suppose that such an

---

21 This is to generalize *akrasia*—the lack of power—so that it may afflict all agency states.

22 This hypothesis runs counter to one tenets of (some versions of) the causal theory of action, namely, that the motivational strength of a desire is a function of a judgement about its desirability: the more desirable, the stronger the desire.
occurrence is possible. Perhaps lack or diminishment of motivation is sometimes explainable in this way. If our case is to be an example of reticence of intention, and not some dysfunction of belief, then we must presume that belief has done its job: that is, the belief function has supplied what amounts to cognitive sufficiency for action. The event of agency—the kicking—has been sufficiently described and identified and judged to be the thing to do here and now. If this is not to be another case of recalcitrance of an active attitude, as above, then the dysfunction cannot be one of the intention failing to uptake the information from belief. A plausible alternative is that, despite having been sufficiently informed by belief, despite, that is, having satisfied what amounts to a set of cognitive conditions on efficacy, so that the state ought now to be both cognitively and causally sufficient for action, no action occurs. So despite belief having performed its function, which ideally should have “triggered” efficacy of the intention, nothing happens. Since we have described the case as one in which cognitive sufficiency exists, the causal diminishment is, ex hypothesi, unattributable to the information supplied by belief. When an intention is reticent, the only form its dysfunction can take is plain lack of causation, since causation is its expression function. This form of non-goal-functionality fits the model of pure functional failures and counts as a case of irrationality. I will argue below that just this sort of functional failure is
implicated in some cases which have been described, variously, as incontinence, weakness of will, and _akrasia_.

Reticence of desire or intention may occur, we may suppose, at any stage in the agency process. If _will_ is the causal strength of one's motivational states, then _weakness of will_ may be explained as the reticence of the active attitude at the point of contact between an intention and an action. The general lack of motivation wherein one's active attitudes generally lack sufficient causal strength for action, a condition perhaps describable as a form of depression, again may be characterized as global reticence of the agent's active attitudes. Thus is raised the philosophically troublesome problems of _akrasia_, incontinence and weakness of will, dysfunctions which we have found reason to distinguish. I address these problems below.

These, then, are some examples of the sorts of dysfunction with which agents may be required to contend. We saw above that rationality, the control of content and function by beliefs about how to be goal-functional, is one means by which agents may overcome such problems as false belief and incompatible desires. For such an agent to be rational requires that content and function be

---

23 Our taxonomy of dysfunction allows us to distinguish these types of agency failure. See below for more on incontinence and _akrasia_.

24 Is it true that we reserve the epithet _weak willed_ for failures to act, and therefore, to describe the failure at the point of contact between an intention and an action. We may. But why not extend the notion to include reticence of active attitudes in general, wherever in the process they may occur?
vetted for goal-functionality. What more rationality requires will depend upon what other goal-functional processes the agent engages in, what non-goal-functionalities he is susceptible to, and what means are available to him to ameliorate or avoid those failures. In the next section, then, we turn our attention to expanding the rôle of the ideal of cognitive control over agency to include the means agents may have to repair or ameliorate certain non-goal-functionalities, and in contrast, the means they may have to avoid or circumvent those non-goal-functionalities.

5. IDEAL AND NON-IDEAL CONTROL, REPAIR AND AVOIDANCE OF NON-GOAL-FUNCTIONALITY

5.1 Rationality and Control

The ideal of agency expressed in our concept of rationality is the control of content and function by belief. Belief is the perfect cognitive means to goal-functionality, and our ideal of agency would have belief reign. We have described the mechanism by which belief performs this rôle: in its capacity of governor of practical processes, desirability and believability beliefs grant or withhold assent to the content and/or function of states that would enter those processes, and the presence or absence of positive evaluation determines which content and what functions are allowed into the process and which are not. Thus, the ideal for agency includes that one intends to do only what one judges to be desirable
because one so judges: an ideal agent is moved to the state of intending to A, for example, because he desires to A and believes that to A would be desirable. An agent for whom the desirability belief fails and who must therefore resort to some alternative means to forming an intention would be diminished in this respect. Those alternative means of repair of an agency failure may themselves be measured against our ideal of agency. The ideal for repair is the same as the ideal of agency: amelioration for or because of the right reason. Non-ideal repairs employ other, less direct means to overcome a dysfunctionality, and an agent may be forced to resort to such measures when reason alone proves insufficient.

In this section we will consider the measures agents may take to ameliorate or avoid the state dysfunctions that have disrupted the practical process in the examples in the previous section. The ideal for repair of a dysfunction-induced non-goal-functionality will be some direct exercise of cognitivity on the dysfunction. The precise form of the cognitive correction will be, in part, a function of the sort of dysfunction it is. Some failures may be correctable by belief itself, in much the same way we saw belief act to prevent false beliefs and undesirable desires from entering a practical process. Other failures may require more of agency than the exercise of belief: some repairs may require mental acts. In addition to ideal corrections, there are non-ideal skills agents may develop to address non-goal-functionality. Alfred Mele, in his important and recent
contribution to the study of irrationality and self-control, discusses a variety of strategies, cognitive and otherwise, that an agent may follow to, as he would put it, bring one's motivational strength back in line with one's judgement about what one ought to do, should the two diverge. What we have been calling the ideal means of control, and what Mele disparagingly calls "brute resistance," is but one way to correct a motivational imbalance. In addition, an agent's control repertoire may include techniques of "skilled resistance" such as those techniques used by behavioural therapists. Skilled resistance includes techniques to distract and/or focus the agent's attention in ways designed to tip the motivational balance in favour of what the agent judges he ought to do and away from what he most strongly desires to do. For example, an agent may employ skills to distract his attention away from the strongly desired aspects of what he judges to be undesirable. Alternatively, he may focus his attention on the desirable aspects of what he judges he ought to do so as not to be distracted by the strongly desired aspects of what he judges to be undesirable.

These are valuable skills for the irresolute agent. But the distinction between ideal and non-ideal modes of control plays no part in Mele's account of self-control and intentional action. The truly self-controlled agent, we should say, would intend and do what he judges he ought to do, because he so judges, and not

because he has managed to so avoid or circumvent a motivational state his judgement cannot control, useful though that skill may be. It is Mele’s failure to distinguish the ideals from the non-ideals of control that make his views on self-control and *akratic* action deficient.

If action discourse is to remain a device we use to explain ourselves to each other, then whether the causal history of an action is ideal or not ought to make a difference in what we ascribe to agents. Should an action be preceded by a dysfunctional state, both the fact of a dysfunction, and the nature of the means used to address it, ought to find expression in the terms of our action discourse. Action discourse, like other of our classificatory schemes, recognizes preferred forms of the phenomena it classifies, and distinguishes these from the various non-preferred forms. Therefore, if one’s theories of agency and self-control fail to distinguish ideals from non-ideals, that classificatory scheme may class as ideal phenomena which are not. This, I argue, is true of both Mele’s and Donald Davidson’s views of *akratic* action.

Suppose, then, that a dysfunctional state—a reticent means belief, for example—has made its way into a practical process. That it is *in* the process is clearly non-goal-functional. That it was *allowed into* the process is clearly a departure from the ideal of *control* of that process: it would show that the controller had, contrary to its rôle in rational agency, allowed a dysfunctional state into that practical process. It would be irrational to allow a *reticent* means belief
into intention formation, since once there, it could not perform its practical rôle. In the following section we will consider a set of cases, not of dysfunctional states being allowed into the practical enterprise, but cases in which dysfunctionality occurs within the practical process, and what agents may do about it.

5.2 Ideal Repair of Dysfunctionality

Let us take as our example the failure to form an intention, given what ought to be its sufficient conditions. Suppose, that is, that our agent desires to A, believes that to A would be desirable, yet does not as a result intend to A. To simplify the case, let us assume that both desire and belief are standard, non-dysfunctional states. If there is a threshold of causal strength which states must have to qualify as “real” desires and “real” beliefs, then being sufficiently strong to contribute to the formation of an intention, when conjoined in the appropriate manner, is a likely candidate for that degree of strength. So, we will assume that, prior to co-function, neither state suffers from reticence or incontinence.

5.2.1 Emergent Dysfunctions

First consider the case in which there are no competing desires and/or intentions which may have interfered with intention formation. We want to consider an emergent weakness which is due simply to the fact of co-function of the desire to A and the belief that to A would be desirable, and not some external cause of the failure to come to intend to A. An emergent weakness due to co-function is to be
distinguished from the case in which the belief that an action is imminent causes fear, for example, or some other reason for not A-ing, which disrupts the intention formation. The simple fact of the imminence of action terminating intention formation may be a version of the phenomenon sometimes referred to as "choking." Should this emergent dysfunction be an end of the intention to A, we would attribute that to some weakness: either the weakness which explains the failure of the intention, or the weakness which explains the agent's omitting to repair the failure.

We are assuming that our agent is a complex, self-aware, rational agent who is apprised of the value of rationality, and so intends to be rational. This is to assume that he has the "standing" intention that his agency processes function according to the ideal expressed in his theory of rational agency. It would follow that he has the standing intention that every rational step in the process of intention formation occur, including the production of an intention to A from a desire to A and the belief that to A would be desirable. Thus, if he encounters the failure to produce the intention to A, he would, on pain of irrationality, intend to do what he could to repair it. What he can do will depend on the sort of failure it is.

One possible explanation is that the intention failed because of an emergent weakness of the causal function of the active attitude: the desire to A was not reticent, but reticence emerged upon co-functionality and so diminished the causal
strength of the desire to A that the resultant state fails to qualify as in intention to A. What it lacks is sufficient causal strength. What it needs, therefore, by way of repair, is an “injection” of causation. That would be available only from another active attitude, a desire or an intention, and one way to effect a repair would be to produce it intentionally. That would be to perform the mental act of intending to A: that is, intentionally intending to A—that mental act.

If our agent makes the repair in this way and goes on to A, how ought we to describe his A-ing? Were we to describe it as a mere intentional A-ing, that would not do justice to the agent, given the extraordinary means he employed. Intentionality is one of our ideals of action, and just what is required for intentionality is a matter of some dispute. But quite apart from those disputes, to classify this case as intentional simpliciter would be to fail to distinguish actions whose ancestry contains no departure from an ideal of agency from those actions which do. A-ing, in our example, required more of our agent than would be captured by the description ‘intentionally A-ed.’ Although that ascription may not be false, given the rôle in the A-ing played by the intention to A, it would be an incomplete description of the case, and possibly misleading. This was no ordinary A-ing, and that fact ought to be expressed in our ascriptions so long as action discourse remains an important means of describing agents. Since it was an A-ing due to an intentional addition of motivational strength, and if will just is rational motivation, we might describe the action as a strong-willed intentional A-ing: a
description that indicates that the A-ing was non-ideal and indicates the nature of the means used to rectify the failure. Strength of will being part of the ideal of agency, the strong-willed repair is an ideal repair.

If these means prove insufficient to overcome the emergent reticence of the intention to A, we would describe the agent’s failure to produce the intention to A as a weak willed failure, since it was the active attitude intention to intend to A which proved incapable of producing the intention to A. Alternatively, if the means prove insufficient due, not merely to weakness of will, but to a misjudgement of the amount of will or causal force required to overcome the original failure, we might describe the case as a mistakenly weak-willed failure.

The state dysfunction we have called incontinence is another possible cause of the failure of the intention to A, given what would otherwise be its sufficient conditions. Thus suppose the desire to A and the belief that to A would be desirable combine, as we have described, to produce the intention to A which, because it is incontinent, immediately upon formation loses content necessary for intention, such as the positive evaluation provided by the belief that A-ing would be desirable. There may be several possible means to repair such a failure. One ideal means of repair would be for the desirability belief to “reiterate” the desirability of A-ing so as to add positive evaluation to the desire to A. If successful, we might indicate the presence of this repair by ascribing to the agent, not strong-will, but strong reason, or strong beliefs, an aspect of the ideal of
agency. We might describe such an agent as having the *courage of his convictions*, since he overcame the incontinent intention by *courageous* or strong beliefs. Alternatively, we may describe his agency as *continent*, to indicate the success of extraordinary means of control. If these repair functions fail to overcome the incontinent active attitude, we might explain the agent’s failure to intend to A as due to an *incontinent* or *vaporous* intention.

The above repairs qualify as ideal because they employ just those functions integral to the ideal of agency itself. An agent who is able to repair a dysfunctional state would be irrational for failing to do so for the same reason that an agent with the capacity to vet his intention formation for goal-functionality would be irrational for failing to do so: both are failures to employ available means to a goal. Naturally, we should like an explanation of a failure to repair a reparable dysfunction, and that explanation may include a dysfunctional state at the level of the repair mechanism: nevertheless, the failure to repair would count as an irrationality.

### 5.2.2 Competing Desires: Incontinence and Compulsion

In the two cases of failure above, the intention failed for what we might call internal reasons: the *intention* was dysfunctional, either reticent or incontinent. We now want to consider cases of intention failure in the face of competition from other agency states, and what agents may do to repair such failures.
As above, suppose that our agent desires to A and believes that A-ing would be desirable. Ideally, that would be sufficient to move the agent to a state of intending to A. But suppose the intention to A isn’t formed because of the presence of the desire to B. The precise details of how the presence of a competing desire to B has disrupted intention formation will determine how we classify this failure. There is a distinction to made at precisely this point that has often been overlooked in the literature on irrationality with the resulting failure to distinguish between weak self-control and weak motivation. The cause of the failure to intend to A may either be due to an overly-strong desire to B, or to an overly-weak judgement that to A would be desirable. Put in our terms, the failure to intend to A may have been caused by a garrulous or compulsive desire to B, or it may have been caused by a reticent belief that to A would be desirable. Much hinges on this distinction, including our assessment of the agent and what he may do to ameliorate the failure.

Suppose, first, that the failure to intend to A was due to reticence of the judgement that to A would be desirable. Such a failure of belief to control intention formation and action is clearly a breach of our ideal of agency, and, since it is a failure of the control function—it is a control deficit—we may describe it as a case of incontinence or weakness of the control function. Should the agent go on to B, rather than A, his B-ing would be incontinent. Furthermore, the motivational state responsible for the B-ing was a mere desire to B, and not an intention to B:
B-ing has not been judged to be desirable and so the state which was causal of the B-ing lacked the evaluative content essential to intention. This is as good a ground as any for denying that the B-ing was intentional. We ought to say that the B-ing was incontinent and unintentional.\(^{26}\)

The other explanation we have available for the failure to intend to A is that it was due to a garrulous desire to B which usurped the desire to A's rightful place in the practical process as the \textit{ceteris paribus} cause of action. Since \textit{ex hypothesi} this would not be a control deficit but an excess of causal strength, it ought to be distinguished from the previous case: we should describe the B-ing, should it occur, as a \textit{compulsive}, but not an incontinent, B-ing. Furthermore, as above, since B-ing, \textit{ex hypothesi}, has not been positively evaluated, it would be a B-ing that was not caused by an intention to B. Thus, it would be a compulsive and unintentional B-ing.

If incontinence of judgement has been the culprit, one possible means of repair of this failure would be, as we have seen, for the desirability belief to reiterate the desirability of A-ing so as to produce the intention to A. If successful, we would again describe the agent as having strong reason or beliefs, or as having the courage of his convictions so as to indicate that the failure due to weak belief was overcome by strong belief. Should the agent now A, it would be an intentional

\(^{26}\)If compulsive actions are unintentional, then so should incontinent actions be classed as unintentional, and for the same reason.
A-ing, given the role of his intention to A. And we could indicate the presence of the extraordinary measures of control by describing the action as an intentional and continent A-ing. If these ideal means to ameliorate the failed intention themselves fail, he would be an agent who lacks the conviction of his beliefs, specifically, the belief that to A would be desirable. That belief, together with the desire to A, should have produced the intention to A, but did not because it lacked conviction. Should the agent B, since ex hypothesi he did not intend to B, it would not be an intentional B-ing: it would be, rather, an incontinent and unintentional B-ing. Now, our agent may have skills he may fall back upon in order to overcome his self-control deficit, and rationality may require that he use them. But if he were successfully to use back-up strategies, we would, as we shall see, describe the case rather differently.

In contrast to the incontinent agent, suppose that our agent fails to come to intend to A because of an overly-strong desire to B. Despite what would otherwise be sufficient conditions for an intention to A, what the agent finds in the practical process is the desire, but not the intention,²⁷ to B, rather than the intention to A, and this aberration of the process is due, not to an overly weak judgement that to A would be desirable, but to the overly strong desire to B which has, in its garrulousness, overridden the practical and rational restrictions on practical

---

²⁷Since it lacks the evaluative content essential to an intention, it remains a mere desire to B.
reasoning, and placed itself where the intention to A ought rightfully to be. That
is, we are assuming, in this case, that the judgement in favour of A-ing is not the
culprit: it is, by our standards, sufficiently strong to perform its rôle. The agent
does not, therefore, have a control deficit. Rather, he suffers from an excessively
strong or garrulous desire: he has a compulsion to B.

The rôle that the desirability belief plays in our theory of intention
formation is to determine which desires are transformed into intentions quite
independent of the desire's causal strength. Thus what an agent is committed to
doing need not be what he most desires to do. Indeed, rational agents may never
act on their strongest desires should those desires conflict with what they judge
they ought to do. This seems undeniably true of rational agents. Therefore, what
ought to occur for the agent who finds himself with a compulsive desire to B
occupying a place in practical reasoning that he has judged should be occupied by
the intention to A, is the reiteration of the judgement that to A would be desirable.
If this occurs with the result that the intention to A displaces the garrulous desire
to B in practical reasoning, we should again credit our agent with the courage of
his convictions in overcoming the overly-strong desire to B. Should this strategy
fail, however, it would not be appropriate to credit our agent with lacking the
courage of his convictions, on the assumption that it is the garrulousness of the
desire to B, and not the reticence or weakness of the belief that to A would be
desirable, that explains this agency aberration. He has the conviction, but it is
ineffectual in the face of the garrulous desire to B. If the agent goes on to B, it would not be an intentional B-ing, lacking, as it does, the evaluative component essential to intention. It would be an unintentional and compulsive B-ing, a description which indicates the presence in its explanation of this specific non-ideal cause of behaviour.

Another means an agent may employ to overcome his compulsions is to intentionally form the intention to do as he judges he ought. Given the standing intention to intend as he judges he ought, together with the belief that one means to so intending is to intentionally intend to A, an agent may perform the mental act of intentionally intending to A, the intentional production of that state of intending. If successful, we would credit the agent with strong will, to indicate the extraordinary rôle will, in the form of the intention to intend to A, played in intention formation. Should the agent go on to A, it ought to qualify as an intentional A-ing, given the rôle of the intention to A, and it ought to count as a strong willed A-ing, given will's extraordinary rôle: it would be a strong-willed, intentional A-ing. Should the agent fail to subdue his compulsion to B by intending to intend to A, ought we to charge him with weak will? That would be appropriate only on the assumption that will, or intention, ought to have been strong enough to subdue the desire to B, an assumption contradicted by our assertion that the desire to B was compulsive. So he was not weak-willed. Should he go on to B, it would be a compulsive and unintentional B-ing.
These example of various means to ameliorate dysfunctionality ought to qualify as ideals of control since each has been either an exercise of belief or an exercise of intention, and when successful the dysfunction was ameliorated because of that exercise. In the next section we will compare these ideals of control with other useful but non-ideal measures agents may take to overcome dysfunctionality in the practical process. The difference in means of control will be marked in our ascriptions to agency.

5.3 Non-Ideal Repair of Dysfunctionality

In chapter 2 of his *Irrationality: An Essay on Akrasia, Self-Deception and Self-Control*, Alfred Mele, concerned to clear the way for his account of strict akratic action—free, intentional action contrary to the agent’s judgement of what he ought to do—argues against several attempts to show that what others have thought to be strict akratic actions are not. Of special interest to us here is Mele’s

---

28 See, especially, pages 22-30. Mele is criticising the views of Wright Neely, David Pugmire, and Gary Watson. If my argument against Mele is successful, I will not have vindicated the views of these theorists. However, my argument turns on a distinction which may have been behind their arguments, although it is not mentioned. It is possible to interpret their arguments as resting on a distinction between the ideals of self-control and non-ideal means to control oneself, although neither of them puts matters in these terms. They may be sympathetic to the view that the failure of the right reasons to control action and the resultant need to resort to skills requires a qualification to the action worth noting in our ascriptions to agents. See W. Neely, “Freedom and Reason,” *Philosophical Review* 83: 32-54; D. Pugmire, “Motivated Irrationality,” *Proceedings of the Aristotelian Society* 56: 179-196; G. Watson, “Skepticism About Weakness of Will,” *Philosophical Review* 86: 316-339.

29 I will address the problem of *akrasia* in the next section.
discussion of the "freedom" condition of strict *akratic* action. Mele argues that, whereas theorists such as Neely, Pugmire and Watson have denied that a B-ing, for example, is free when it was caused by a compulsion to B contrary to the agent's judgement that he ought to A, the mere fact of the presence and function of a compulsion to B is not enough to show that the B-ing was not free. If every act contrary to an agent's judgement of what he ought to do were unfree, then strict *akratic* action would be impossible. Mele argues that what more needs to be shown in order to show that a compulsive B-ing was an unfree B-ing, is that the agent *could not have done otherwise*, where this he interprets as meaning that there was nothing he could have done, of an ideal nature or otherwise, in order to subdue the compulsion. It is in this context that Mele discusses what we have been calling the skilled means of control. Thus, should a compulsion to B have supplanted the intention to A in an agent's practical process, as we have described above, and the agent as a result compulsively B, that would not entail that the B-ing was unfree: there may have been some measures the agent could have used to defeat the compulsion to B. If so, he could have done otherwise than B, and his B-ing was free. Among the additional means of control which Mele discusses, all of which count as means to self-control,\(^3^0\) for Mele, and all of which need to be shown to have been ineffectual or unavailable before we can count the B-ing as unfree,

\(^3^0\) Again, we should distinguish self-control in the ideal—that is control *for the right reasons*—from control brought about in support of the failure of the right reasons.
are such techniques as the following: distracting one's attention away from the
desired aspects of B; focusing one's attention on the desired and desirable aspects
of A; promising to reward oneself for resisting the desire to B and A-ing in its
stead; refusing to entertain second thoughts about the desirability of A-ing over B-
ing; and, picturing the desired B as something decidedly unattractive.

It is clear that there is a sense to 'irresistible desire' such that should an
agent have been able to successfully resist his compulsive desire to B by employing
one of the above techniques, that would be enough to show that the desire was not
irresistible. It is equally clear that to fail to employ available means to subdue a
compulsive desire to B does not entail that the agent was unfree with respect to a
compulsive B-ing. On the contrary, that would be a compulsive B-ing that the
agent could have resisted. And, it is also clear that to successfully resist the
garrulous or compulsive desire to B in either of these skilled ways, ought to qualify
as a form of control. What separates the view of self-control that I have been
defending from Mele's view, and what will explain our differences over how such
cases as these ought to be described, is that he makes no distinction between what
I have been calling ideal self-control and skilled means to control oneself. Should
we accept that the determination of which motivational states are efficacious of
behaviour by the use of any of Mele's techniques amounts to self-control
*simpliciter? Should we allow, that is, that skilled manipulation of motivation is as
good as the sort of control of agency states that we have found at work in our
analysis of the ideals of agency? Isn't being forced to resort to such a skill by the failure of one's judgement about what one ought to do, indicative of a diminishment of agency? Isn't skilled control of oneself clearly second best control? I am arguing that this difference is important to our assessment of agents, and ought to be marked in our action discourse. We should resist the temptation, therefore, to describe the skilled self-controller in the identical terms in which we would describe the agent who is self-controlled *simpliciter*: that is, the agent whose beliefs control his agency. What difference would this distinction make to the this case?

I am not arguing that Mele is wrong in his claim that an agent is free with respect to a compulsion he is able to control with skilful modes of resistance. What I am trying to show is that if we distinguish between an intention which is efficacious *because* the agent judges that it ought to be, and an intention which is efficacious only because skilful self-manipulation has been brought to bear to subdue another motivational state which otherwise would have been efficacious, and not, therefore, an intention which was efficacious because the agent judged that it ought to be, then we have solid grounds for resisting the invitation to treat the action which results from skilled manipulation of agency as an ideal case of intentional action. I am not suggesting that, should an agent successfully subdue a garrulous or compulsive desire to B through skill rather than through "brute resistance", there would be no intention to A playing its rôle in the cause of the A-
ing. Rather, I am arguing that, since what explains the efficacy of the intention to A is not the agent’s judgement in favour of A-ing, but his use of skilled resistance, we ought to mark that fact in what we say about the A-ing. We should deny that such an A-ing is a *strictly* intentional action. What we ought to say, just so long as action ascription remains an explanatory device by which we explain ourselves *qua* agents to one another, is that, while perhaps intentional, because it was an A-ing due to the agent’s resourcefulness and not due simply to his reasons, that it is not a *strictly* intentional action. Neither is it a *strictly continent* action, if we reserve ‘continence’ for ideal control of agency: that is, for self-control by strong reason or belief itself. For Mele, the fact that an agent’s beliefs were insufficient to determine which of his motivational states become efficacious is not to be marked in action ascription. Nor is the presence of the agent’s use of skilled resistance to be so marked. For Mele, the presence of the compulsion to B requires neither that we note this irregularity in our ascriptions to the agent, nor that we indicate the nature of the route the agent travelled in order to wrest control away from the compulsion to B so as to bring his motivations back in line with his judgement. Mele would have us describe the A-ing as a free, intentional action.

To further show just how Mele and I differ on this matter, consider the case in which our agent, while judging that he ought to A, nevertheless B’s because the desire to B, in being garrulous or compulsive, overrides the agent’s judgement in favour of A-ing, and causes the B-ing. Suppose further that our agent could have
employed Mele-like skills to resist the desire to B, but did not. What would the tools of agency analysis that we have been examining have us ascribe to this agent? If we reserve the term 'compulsively' to describe agency processes and actions that occur because a desire has been excessively strong and so able to override an agent’s reasons, then we may say of this agent, quite irrespective of the fact that he could have resisted the desire to B in other ways, that he compulsively B-ed. Since he could have resisted through the use of skills, but did not, the beliefs which would have had him so resist may have been incontinent. Or, he may have been negligent in his rational duty to monitor his agency for failures. Without a more complete account of the case, we cannot say which we ought to ascribe to the agent.

In the next section, I apply this theory of agency to the so-called problem of akrasia. In order to appreciate more fully the nature of the problem and to see exactly where my view differs from both Donald Davidson’s and Mele’s, we will begin with a brief overview of the problem.
6. THE IDEALS OF AGENCY AND THE PROBLEM OF AKRASIA

6.1 Introduction

In treating of this issue we are immediately met with a terminological problem: what should we call that aberration of agency which has been variously described as weakness of will, akrasia and incontinence? The problem is further complicated here by the fact that we have used the term ‘incontinence’ for a certain weakness of agency states—namely, their inability to retain content—which may have nothing to do with some cases of akrasia.\(^\text{31}\) What is typically intended by the use of either of these terms is that a certain phenomenon—akratic or incontinent action—is due to a weakness in agency. Since ‘akrasia’ means weakness or lack of power, I will apply that general term to the phenomena which are responsible for such diverse things as akratic action and akratic belief.

Akrasia, with few exceptions,\(^\text{32}\) has been thought to be conceptually problematic as it applies to action. Thus, there are competing views about the nature and possibility of so called strict akratic or incontinent action, which is defined as free, intentional action contrary to the agent’s judgement of what he ought to do. The problem of akrasia, as we would cast it, is whether an ideal of

\(^{31}\) Although it may. Perhaps some cases of akratic omission are due to an intention’s having lost essential content.

rational agency—free, intentional action—is compatible with action contrary to the agent’s decisive better judgement of what he ought to do. Sceptics argue that cases of proposed strict *akratic* action are either really not free[^33] or really not intentional, or both. Our question will be whether the compromises to the ideals of agency which so-called strict *akratic* action require are such that we should resist the invitation to classify such action as “strictly” intentional.

### 6.2 The Possibility of Strict Akratic Action

The phenomenon of *akratic* action may be described as the failure of an intention to cause its intentional object despite the presence therein of informational content which ought to be sufficient for practical rational agency. The problem of *akrasia* may thus be seen as really a challenge addressed to any theory which claims to have an account of the generic conditions necessary and sufficient for action. Those who assert the possibility of *akratic* action attack any such offered conditions with the argument that their satisfaction is compatible with inaction, or, as frequently, with some action other than the one for which sufficiency, *ex hypothesi*, exists. Defenders of a thesis which claims to have specified the sufficient conditions for action, thus attacked, counter with the charge that to deny action under their conditions is conceptually incoherent. But if action is a viable concept then we are entitled to believe that there are *bona fide* sufficiency conditions for action.

[^33]: See section 5.3 above for Mele’s views on the freedom condition of strict *akratic* action, and my response to those views.
conditions for rational action; that some cognitive and causal states must be sufficient for some actions. As we have demonstrated, we can separate the causal sufficiency or the active causal function of such states from their information or content, and view the latter not as contributory to the causal sufficiency of the active state but as a condition upon its sufficiency for action. Thus when an agent reports that he intends to A, he reports that he is in a state of sufficiency for the causation of A, *ceteris paribus*. The process of practical reasoning, which consists of the identification and evaluation of the event to be caused, is a set of *ceteris paribus* conditions which such active states must satisfy before their rational release into activity. Thus when states sufficient, *ceteris paribus*, for action have accreted all the content which practical reasoning adds to them as they pass through the process of intention formation, they must achieve at some point sufficiency *simpliciter* for action.34 If we accept this picture of causation’s presence at the beginning of practical reason, and of practical reason being the control of rationality conditions over causation, then we have the following options of how to see the problem of *akrasia*. Where the agent reports that (1) he intends to A, and that (2) he has removed all the *ceteris paribus* conditions—has satisfied practical reason—and so reports that he is in a state which is causally and cognitively sufficient for action, yet has not acted, then either:

34 This view of the control by belief of causal sufficiency present throughout the practical process, is at odds with Davidson.
a) Report (2) was false and he has not satisfied practical reason; or (2) was true and (1) was false and the agent was not in a state of sufficiency, *ceteris paribus*; or,

b) Reports (1) and (2) were true in which case the rest of the case is incoherent since he is said to be in a state of cognitive and causal sufficiency but not to have acted; or,

c) Reports (1) and (2) were true but report (1) has since become false due either to a decrease in, or other failure of, the causal strength of his intention,\(^{35}\) or to its causal override by some other active state\(^ {36} \) of which the agent may be unaware.

### 6.3 The Problem of Strict Akratic Action

Strict *akratic* action has taken several forms in the literature. On the narrow view, it is intentional action inconsistent with one's occurrent judgement about what one ought morally to do, where that preferred course of action is available to the agent. Others\(^ {37} \) have seen this narrow view as but one type of a broader phenomenon. They have noticed that there need be nothing essentially moral (or immoral) about the overridden judgement constitutive of *akratic* action, so that acting against one's moral judgement is but one way to act *akratically* against

---

\(^{35}\) Weakness of will.

\(^{36}\) A compulsion.

\(^{37}\) Notably Davidson.
one's judgement about what one ought to do. There are, or may be, depending upon where you stand on the question of the possibility of strict akratic action, akratic acts which are moral, immoral, prudent, imprudent, preference-optimizing or otherwise. The essential ingredient of strict akratic action is action inconsistent with one's judgement of what one ought to do.

What makes strict akratic action conceptually problematic are certain widely accepted features of rational agency: in particular, the relation between judging something good and wanting or desiring it, and the relation of judging and wanting to intentional action. It is widely held that judging something good must be reflected in desire for that thing, and desire for a thing must be reflected in action addressed to attaining it. Whether you find the possibility of strict akratic action problematic will depend on how tight you take these dependency relations to be.

It is a widely shared intuition about ourselves, and about the function of judgement in particular, that weakness of will is, quite apart from posing any pressing philosophical problem, a common occurrence. E.J. Lemmon has held that any model of rational agency according to which akratic action would be problematic would be thereby shown to be flawed. On the contrary, the early Plato, among others, claimed that since judgement determines desire and strength of desire, and that since strength of desire determines action, strict akratic action

38 E.J. Lemmon, 'Moral Dilemmas,' Philosophical Review, 71, 139-158.
is impossible. The later Plato is said to have come to hold, against the Socrates of
the Meno, that while an agent may be incontinent with respect to his desires—
may desire what he or his Calculative Faculty thinks bad—so-called clear-eyed
akratic action was impossible\(^{39}\) and cases of unjust action which might appear to
be due to akrasia are in fact either cases of ignorance or compulsion.

The Judaeo-Christian model of rational agency incorporates the faculty of
the will, the function of which is to arbitrate between competing desires, and
whose decisions, on pain of irrationality, are obeyed. But disobedience is easily
conceivable and supplies the pattern for akratic action on the Judaeo-Christian
model of the psychology of action.

We should note that the Judaeo-Christian conception of will is not what we
have been calling will above. Our hypothesis has been that will is the causal
function of the active attitude. One possible interpretation of the Judeo-Christian
will is that it is reason-informed desire, perhaps what we have been calling
intention: an amalgam of cognitivity and causal efficacy. On this interpretation,
weakness of will could be due to weakness of either of the two component
functions. The model of agency set out above, and the dysfunctions which agency
states may suffer from, make available a set of alternative diagnoses of akratic
action. Akrasia, then, may manifest itself in either of the following ways. An
intention which suffered from what we have called reticence, or weakness of the

\(^{39}\)See W. Charlton, Weakness of Will, esp. chapters 1 and 2.
causal or expression function, would be unable to cause what it represents, so that the intended action simply does not occur. Alternatively, the judgement in favour of what one would otherwise intend to do may itself be reticent, the effect of which would be that a desire to A, say, remains merely that, rather than be converted to an intention to A, despite the presence of what ought, ideally, to be its sufficient conditions. Finally, an intention may be what we have called incontinent and so unable to contain its content, despite the presence of its sufficient conditions. In either case, the failure to act as the agent has judged he ought to act would be due to a weakness, to lack of power of one or another sort, and the omission deserves to be described as due to *akrasia*.

It should be noted that these state dysfunctions may occur at any point in the process toward action. While some theorists have cast the problem of incontinence in terms of action alone, it would be a shortcoming of a model, and therefore also of a proposed solution to, or dissolution of, the problem, if the model could not accommodate incontinent omission. The case ought to be as familiar to us as incontinent commission: the agent believes that to A is what he ought to do, yet he fails to A, without intentionally doing anything in the place of A; that is, where no desire or judgement overrides the dictates of Reason to become efficacious of an action. Assuming that cases of incontinent commission and omission have something of significance in common by virtue of which they fall under the same concept, extending the model of incontinence in action to
accommodate incontinent omission ought to shed light on problem no matter what form it takes.

The models of *akratic* action imply a theory of *rational* agency from whose standards *akrasia* is a departure. That theory gives reason, in the form of judgement about what one ought to do, a vetting or arbitrating rôle in agency. That is, the job of reason in rational agency is to "select" from among the agent's motivating desires those which the agent, or his reason, judges would initiate the best course of action. The effect of selection ought, in rationality, to be that that motivating desire and not another, nor nothing, is efficacious of an action. When the corresponding action is not forthcoming, *some* deviation from the standard case of agency has occurred. *Akrasia* is one such deviation, compulsion another, and so on.

6.4 Davidson on Akrasia

Donald Davidson, who is credited with reviving the issue of *akrasia,* presents us with a *prima facie* logical inconsistency between the Greek conception of the psychology of action, on the one hand, and the claim that there are *akratic* or incontinent actions on the other. The *prima facie* inconsistent triad, each member of which Davidson encourages us to embrace, is:

P1. If an agent wants to do *x* more that he wants to do *y* and he believes himself free to do either *x* or *y*, then he will intentionally do *x* if he does either *x* or *y* intentionally.

40 Ibid., 10.
P2. If an agent judges that it would be better to do $x$ than to do $y$, then he wants to do $x$ more than he wants to do $y$.

P3. There are incontinent actions.\footnote{Donald Davidson, “How is Weakness of the Will Possible?”, in \textit{Essays on Actions and Events.} (Oxford: Oxford University Press, 1980), 23.}

What makes \textit{akratic} action a problem is that on this model of agency, the relation between judging something good and both desiring it and acting upon that desire is meant to be so close that both desiring and acting against that judgement appears to be a logical impossibility. Davidson’s way out of the problem involves a minor refinement of the Greek model.

Davidson argues that incontinent action only appears to be problematic if we fail to distinguish between two kinds of judgement with distinct rôles in action. One essential ingredient of agency is what Davidson calls the \textit{conditional, or all-things-considered} judgement in favour of an action, according to which an action is judged favourably relative to one’s reasons so far considered. That such judgements are only conditional means, argues Davidson, that they may not play the rôle required of the judgements which explain intentional actions. What explains why an agent does what he does rather than something else, or nothing at all, cannot be merely that he judges the action favourably relative to the reasons so far considered. Intentional action requires more than that a proposed course of action is judged favourably \textit{so far}. What more is needed is what would
amount to commitment to the action, and that must come from an unconditional judgement in favour of the action, what Davidson calls an "all-out judgement." In standard or ideal cases of intentional action, one's reasoning takes one from a conditional, all things considered judgement in favour of A-ing to an unconditional all-out judgement in favour of A-ing—an intention to A.

In defending the possibility of strict akratic action, Davidson exploits this inference from the conditional judgement to the unconditional judgement in favour of an action. On his model, akratic action is the result of an inconsistency between judgements of these two different sorts. An agent may judge, all things considered, that it would be best to A, yet find himself committed to B-ing: may find himself, that is, with an all-out judgement in favour of B-ing—an intention to B.

According to Davidson's view of standard cases of agency, an all things considered, conditional judgement that to A is better than to B should cause the agent to want to A more than to B, and should, barring change in the relevant circumstances, cause the all out, unconditional judgement that to A is better than to B which will in turn cause the A-ing. Akratic action, on this model, occurs when an all out, unconditional judgement that to B is best occurs despite the antecedent conditional judgement that to A is best and without any relevant changes in the circumstances which would justify such a change of mind. That is, given the circumstances—i.e., relevant to the conditional judgement and the context—the
unconditional judgement that to B is best is \textit{irrational}, and the action which eventuates is irrational, incontinent, but \textit{intentional}, given that it was caused by the intention to B—the all-out judgement in favour of B-ing. \textit{Akratic} action is thus irrational but intentional action due to a judgement inconsistent with the agent's judgement about what he ought to do, all things considered.

On this conception of \textit{akratic} action, the agent's judgement that \textit{all things considered} to A is better than to B, which should have caused him to desire to A more than to B, and therefore to A intentionally; that is which should have had appropriate effects in both what he wants to do and his actions, is instead preempted by the unconditional judgement that to B is best, \textit{tout court}, which is efficacious of his intentional B-ing.

\textbf{6.5 Against Davidson}

As we would describe the case, an incontinent action is the effect of a state which, while being causally sufficient for its effect, lacks an ingredient essential to intentional action. That ingredient cannot be desirability, since \textit{ex hypothesi}, the agent judges B-ing most favourably and is thus committed to B-ing. What the causally sufficient state lacks is cognitive sufficiency: it has been detached from the agent's reasons, all of which \textit{ex hypothesi} count in favour of his A-ing. The agent's final causal state in the practical process has not carried with it the agent's reasons for acting. It is for this reason an aberrant intention, and it is that fact which allows Davidson to call the resulting action irrational. But an intention is
not simply the causally strongest motivational state: it is a motivational state for an action which has been judged to be desirable in the light of the agent's reasons. In making the move required to show that the B-ing is irrational, Davidson has detached the causal state from the rational process necessary to make the action intentional. If so, incontinent actions are not strictly intentional actions, contrary to his claims.

6.6 Against Mele

I have argued above that Mele's treatment of the freedom condition for strict akratic action is a significant compromise of our ideals of agency and for that reason we ought to resist this attempt of his to make room for "strict" akratic action. While there is a sense in which an agent with Mele's techniques of skilled resistance is free with respect to a desire to do what he judges he ought not do, there is an ideal of freedom with respect to one's motivants that an agent who is forced to resort to such skills lacks. Mele has weakened the freedom condition on strict akratic action, and one may, for this reason, deny that such actions would be "strictly" free.

As for the intentionality condition on strict akratic action, Mele accepts the Davidsonian move which would permit an action to be incontinent and intentional. However, he attempts to move the point at which akrasia may occur "closer" to the

---

42See section 5.3.
action than even Davidson had managed. That is, Mele argues that an agent may, on the basis of his conditional judgement that to A would be best, unconditionally judge that to A would be best, and thus be in a state which would otherwise be both causally and cognitively sufficient for intentional action, yet fail to A due to the intention to A being displaced by an all-out judgement in favour of B-ing, even at this late stage in the practical process. For Mele, what makes the B-ing irrational is just what would make it so for Davidson: the detachment of commitment from reasons. And for this reason, his view is subject to the same objection we made against Davidson: the detachment renders the intention non-standard. It is no longer a state which is causal with respect to a doing which has been judged to be desirable in the light of the agent’s relevant reasons. For Mele, as for Davidson, the “intention” in “strict akratic action,” is not a “strict” intention. If not, the action in not a strictly intentional action. If not, it cannot be a strict akratic action, contrary to Mele’s claims.

The price of showing the akratic action to be irrational—namely, severing it from the agent’s reasons—is that the action cannot be ideally or “strictly” intentional. The ideal of intentionality carries with it the supposition that the action is judged by the agent to be desirable in the light of his reasons, that is, relative to his relevant beliefs and desires. The fact that akrasia has severed this connection gives us one good reason to deny that the action is intentional.
6.7 Differences with Davidson and Mele over Desire and Motivation

In our theory of agency, 'desire' marks the presence in the system of causal sufficiency, *ceteris paribus*, for doing what is represented in the content of that state. This feature of our account marks a significant departure from those of other action theorists, most notably, perhaps, Donald Davidson, who finds it self-evident that "If an agent judges that it would be better to do x than to do y, then he wants to do x more than he wants to do y." This feature of their theory allows an account of *akratic* action according to which the desire to A, together with the judgement that it would be best to A, comprise insufficient motivation for A-ing. An explanation available to them is that the desire was insufficiently motivational, and that insufficiency remained, despite the increase in causal strength supplied by the judgement that A-ing was what one ought to do. In cases of *akratic action*, this insufficiency would allow a desire to B, which is sufficient for action, to itself be efficacious, despite the presence of the judgement that one ought to A, and despite the lack of judgement that one ought to B.

Our theory recognizes differences in strengths of desire, but insists that in order to qualify as a desire, a state must have a certain minimum causal strength: that minimum is sufficiency for action. Our theory also recognizes that the effect of positive evaluation of a doing can increase the motivation for the doing:

---

however, on our theory that does not equal an increase in causal strength of a desire. An increase in motivation for a doing is not an increase in causal strength of the desire: it is, rather, the effect upon the likelihood of the doing, of the belief that the doing is better by some margin than other available doings. It means that perhaps more ceteris paribus conditions on the doing have been removed; that less, in the way of judgement, stands in the way of the doing. Motivation is not a purely causal matter on our view: it is a composite of causal power and judgement. To say that one is motivated to A is not to report that one is merely causal with respect to A-ing, but that one is causal with respect to an A-ing one judges is desirable. Thus motivation and intention share a common core. Intention is positively evaluated causation and comprises commitment to a doing. Motivation is that concept which, like intention, indicates positive evaluation of a doing, but unlike intention, admits of degree. An agent may be more motivated to A than to B and this means, on our view that, while he desires to both A and B, and while the desire to A and the desire to B may be equally strong desires, he judges A to be more desirable than B.

Alternatively, since desires may vary in causal strength, motivation may be a product of causal strength and judgement. The desire to A and the judgement that A-ing would be better than B-ing may result in the agent being more motivated to A than to B. However, what if, due to the causal strength of the desire to B, he is more motivated to B than he is to A, which he intends and is
committed to doing? Suppose now that he B's. If so, it is not a case of the judgement in favour of A-ing being insufficient to boost the causal strength of the desire to A to the point where the agent A's. The fact that he was more motivated to B is irrelevant. What explains the B-ing is a failure of belief to prevent the doing the agent judges he ought not do. This may be because the belief is too weak or the desire too strong: because, that is, of incontinence or compulsion.

It is possible, however, for the active states of agency to become effective at any stage of practical reason without the benefit of full rationalization. That such impulsivity, or transgression against belief, may occur at any point in the process is some confirmation of the view that the active component of agency is capable of being present throughout. Breaches of rationality by impulsivity are sometimes called compulsions. Whereas compulsion is the paradigm case of the failure of practical reason through its override by a too powerful active attitude, rashness or recklessness is the paradigm of the failure of the application of practical rationality to an active attitude. Psychosis and poor judgement mark the poles of this defect as compulsion and depression mark the other.
SELECTED BIBLIOGRAPHY


