AGENCY IN ACTION

STUDIES IN COGNITIVE SYSTEMS

VOLUME 11

The titles published in this series are listed at the end of this volume.

AGENCY IN ACTION

The Practical Rational Agency Machine

by

S. C. COVAL

and

P.G. CAMPBELL

Department of Philosophy and the Faculty of Law
Artificial Intelligence Research Project,
The University of British Columbia,
Vancouver, Canada

SPRINGER-SCIENCE+BUSINESS MEDIA, B.V.

Library of Congress Cataloging-in-Publication Data

Coval, S. C. (Samuel C.),
Agency in action : the practical rational agency machine / by S.C. Coval and P.G. Campbell.
p. cm. -- (Studies in cognitive systems ; v. 11)
Includes bibliographical references and index.
ISBN 978-94-010-5068-5 ISBN 978-94-011-2422-5 (eBook)
DOI 10.1007/978-94-011-2422-5
1. Act (Philosophy) 2. Agent (Philosophy) 3. PRAGMA (Computer file) I. Campbell, P. G. (Peter G.), 1950- . II. Title.
III. Series.
B105.A35C68 1992
128'.4--dc20 92-3901

ISBN 978-94-010-5068-5

To
Meredith, Margery, Simon and Becca Coval

and

To
Marge and Rob Roy Campbell, and to
Leslie, Marylouise and Jennifer.

CONTENTS

SERIES PREFACE

This series will include monographs and collections of studies devoted to the investigation and exploration of knowledge, information, and data-processing systems of all kinds, no matter whether human, (other) animal, or machine. Its scope is intended to span the full range of interests from classical problems in the philosophy of mind and philosophical psychology through issues in cognitive psychology and sociobiology (concerning the mental capabilities of other species) to ideas related to artificial intelligence and to computer science. While primary emphasis will be placed upon theoretical, conceptual, and epistemological aspects of these problems and domains, empirical, experimental, and methodological studies will also appear from time to time.

Sam Coval and Peter Campbell provide a painstaking and distinctive analysis of the nature of action and agency. They introduce a conception of acts which encompasses the purposes that motivate them, the beliefs on the basis of which they are undertaken, and the effects that they bring about. They compare and contrast their account with ones advanced by Davidson, Brand, Searle, Danto, and other, while elaborating its consequences for understanding the nature of alibis, mistakes, accidents, inadvertence, and the like. The valuable diagrams and the discussion of the software program they have developed, which implements their theory, amply displays the potential of combining philosophy and AI with law and other disciplines focused upon agency.

J.H.F.

LIST OF TABLES AND FIGURES

PREFACE

This is first an essay in philosophy of action. And since we believe that an action is an object explanatory of its agency, our work took us into what ascriptions of action are revelatory of, namely, the practical rational states which causally precede action. As we dug into action and the process which explains it, we encountered a group of fundamental terms and relations — conceptual archaeology, it seemed to be — which lay beneath the main concepts of action and which created a temptation we did not resist. We wanted to realize and test the performance of these terms and relations in a computer program. This made a two year project into a five year project. The discipline of working back and forth between the exacting clarification needed by good computer analysts and programmers and the development of a philosophical theory took time. But the *Practical Rational Agency Machine* got written and *PRAGMA*, the computer program which is based upon it, got built.

PRAGMA will accept agency cases from users, prompting them for relevant and complete information which it then places in the appropriate syntactical slots of the analyzer allowing *PRAGMA* to infer and/or to correct the user's inferences about what the proper ascription to the agent is: whether what was done was intentional or unintentional; if intentional, what its essential features were; if unintentional, whether it was mistake, accident or inadvertence and what their essential features were; what the area of responsibility consists in; and so on. Chapter 5 gives the reader an idea of what *PRAGMA* can do. At this stage in *PRAGMA*'s development some limited use of canonical action discourse is required of users but syntax plays a dominant role in the analysis of cases of agency. Greater use of canonical forms for user interface could be made later. *PRAGMA*'s own ability to analyze given cases without further help from the user could also be expanded.

PRAGMA and its manual will be available separately from the book through Kluwer Academic Publishers and could be used on its own. Because we believe that *PRAGMA* and this book are based on fundamentals, we think other theories of action and agency will find themselves, with not much rearrangement, placeable within our theory and its program. So while the present use of these fundamentals in *PRAGMA* is as an agency analyzer of interest to all students of action theory, whether philosophers or not, we see these same fundamentals playing a role in subsequent *PRAGMA*'s. Some preliminary work has already begun on the analysis of tort and criminal law as it might be organized by *PRAGMA*'s fundamentals. We believe further areas which depend upon distinctions and evaluations of actions and agency will find uses for this book and for *PRAGMA*.

The fundamentals of our model include two separable information-passing functions which do the work of *belief* and *desire*, which, however, pass their information to different types of object: respectively, to other mental states and to the world. The information passed is a representation of the world which itself consists of events, their causal relations and properties. These features are arrangeable into a process which together constitute an analysis of the main syntactical terms of practical rational agency, *viz., the event of agency, means* and *goal.* These features also offer an analysis of the main nodes in practical reasoning such as *desire-to, intention* and *action* each of which has internally the syntax of practical rationality. Our theory treats action terms, such as 'intentional', 'unintentional', 'mistake', 'accident', 'inadvertence', etc., as making reference to objects whose intentionalized (and unintentionalized) properties constitute explanations of why the agent's effect in the world had the properties and relations it did.

We believe these fundamental terms to be the basis of a theory which is powerful in its ability to identify and arrange the logical spaces of agency and action and provides a framework upon which normativity may be hung. Practical areas which presuppose, make

use of, or evaluate action and agency will find the fundamentals organizing and clarifying.

Since, for example, criminal and tortious acts (and omissions) are cases of agency (or required agency) set within two normative schemes — respectively, the criminally forbidden and the negligent — we believe that the application of these schemes to the precise aspect of the practical rational agency their normativity addresses will yield a model which has an excellent prospect of attaining pedagogical perspicuity and direct practical usefulness in these two areas of law.

We further expect that the idea of the corporation as a "legal person" can be rid completely of what remains of its metaphorical flavour by modelling the corporation functions as agency functions. This would allow descriptions and normative appraisals of precise functions of corporate agency and action within a structure like *PRAGMA*'s. Finally we expect that there is a connection to be made through *PRAGMA* between the Philosophy of Action and Agency theory in Economics.

We would like to thank the following people for their help: our philosopher-colleagues, Leonard Angel, Hector Castañeda, Danny Daniels, Mark Glouberman, Andrew Irvine, Howard Jackson, Michael Philips, Simon Coval, Greg Boothroyd, Euan Carnie, Deborah Graham, Greg Hagen, John McGuire, Leo Paquin, Mike Rostad and Joe Naylor; our colleagues at the UBC Faculty of Law Artificial Intelligence Research Project (*FLAIR*), J.C. Smith, Daphne Gelbart, Deborah Graham, Keith MacCrimmon, Doug Arnold and Max Krause; Timothy Bult, Steve Furr, Randy Roesler, Don Johnson and Eric Mitchell of Diamond Software Associates; and Donna Toews of UBC Media Services.

We also thank the Social Sciences and Humanities Research Council of Canada and the UBC President's Research Fund for their financial assistance.

1

ACTION AS EXPLANATION

1. COMMON GROUND AND BROAD THEORIES

There is broad common ground in the philosophy of action occupied by all those theorists who believe that action theory addresses concepts which organize our experience in useful ways. No such action theorist questions that, relevant to the analysis of the concept, there is a process of practical rationality which culminates in an intentional state, I, of the agent, which is causal,[1] C, of a certain effect or set of effects, E, in the world: I-C-E.[2] That is, it is not disputed that agency stands in a causal relation to the world, nor that an analysis of action must be drawn from this broad tri-partition of conceptual phenomena. Set within this tri-partition, or otherwise related to it, is, of course, an entire panoply of action concepts, among which are *desire*, *belief* and *intention*; *action*, itself, and *mistake*, *accident* and *inadvertence*; *means*, *goal* and *consequence*; *intentionality* and *unintentionality*; *responsibility*, *culpability* and *excuse*. Accounts of these close conceptual neighbours, among which we intend to venture, constitute action theory in its refinements.

As one might expect, action theory has seen the emergence of competing theories about what sort of object an action is, even as expressed in terms of the broad common ground of I-C-E which comprises merely two terms and a relation. It has been held that action is identical to E, if E has been caused in the right way, that is, proximately caused by an I which "contains" at least one true

description of E. It has also been held that 'action' refers to I when I has caused E. And action has been taken to be the causal sequence itself, of I causing E. The first of these two views we shall call the Effect Theory of Action:[3] an action is what is caused through the relation with I.[4] Its converse is the Cause Theory of Action: an action is the initial mental event, I, of a specified sort which begins the associated causal sequence.[5] Where the causing itself, the relational complex of Intentional State—Causal Relation—Effect, I-C-E altogether, is held to be the semantic ground of action discourse, it may be referred to as the Relational Theory of Action.[6] The object of all three causal theories is of a semantic type which is compatible with causation and its terms: for the Cause Theory, the Relational Theory, and the Effect Theory, an action is either an event or a causal sequence of events.

Agreement around action semantics as coming out of the terms of I-C-E hides a deeper agreement. It is that the concept of action must contain information which is explanatory of a certain class of behaviour, so that in ascribing an action to an agent we therewith give certain information about the agent's relevant intention. Theories will differ on the nature of the information imparted and we deal with these differences in the next section.

There is other disagreement, however, concerning the less broad concepts which appear integral to the notion of action, such as *desire*, *belief*, *intention* and *rationality*. Some[7] hold that a desire-belief psychology is necessary and fundamental to the explanation of practical reason and action. Some[8] hold that intention is also necessary, while it is argued by others[9] that intention is not an additional consideration but only a stage in practical reasoning which itself is constituted only of desires and beliefs and their relations. More radically yet, still others[10] argue that desires are actually beliefs, although this view may be held independently of the preceding one. Further, some writers[11] claim that there is no meaning relationship, certainly not that of synonymy, between 'intention' as it qualifies a mental event and as it qualifies acts or non-mental events. Others[12] assert that there is more than homophony here.

Despite these differences, there again seems to be common ground. No action theorist, no theorist who believes that action and its family of concepts are in the main defensible and useful, will want to deny that two functions must be assumed for theory to proceed in this area. Any explanation of action will presuppose a causal function for agency and a causal function for reason or belief. We may name these functions as we like, or as they are already named: desire and/or intention and belief.[13] Both of these differentiable functions involve, however, the further feature of being functions with respect to information. That is, each is a function with *cognitive* (or representational) *efficacy*. These two functions and the notion of content, however named, must be in place for the concept of *agency*, and therefore of *action*, to be alive. And we accept this common ground as a beginning.

If the common ground consists of desires and beliefs as causal or functional antecedents of an action,[14] then if an event was intentional, desires and beliefs were in the causal ancestry of the event. This is merely to assert the commonplace that only events caused by reasons are actions. This commonplace contains the assumption that of necessity action will be explicable in cognitive terms — that what explains an action will only be mental states which are representational and efficacious. If we want access to the concept of action we must respect or put up with this connection. We cannot unlock the concept of action if we treat an act as an object caused as might any other event in the world be caused. Even, however, if we accept as necessary this connection between those objects we deem actions and those cognitive events which are their causes, we offer no constraints, other than viability, upon what a theory of cognitive events may be.

Accepting, for the moment, the common ground of I-C-E and the relevance of "desires" and "beliefs" to I, we return to theories about the nature of action.

2. CONDITIONS ON A THEORY OF ACTION

It would seem that natural action discourse should be at least a starting point for theorizing about action. This condition appears

attractive because the concept of action is essentially a creature of natural discourse. In any event, if one works, as we do, with the assumption that the family of action concepts found in natural language will be useful to any theory of action, one cannot begin theorizing without the guidance of natural discourse on action.

A powerful condition, then, on a theory of action and agency is that it give an account, or else justify the exclusion of, relevant parts of our natural language. Let the natural language concepts and connections determine, at least in the first instance, the nature of action concepts. We can imagine the theorizing of humans as consisting in the continuous improvement of their language and the consequent duty to be accountable to it and necessity to begin with it. In following this condition one would want any theory of action to make clear its manner of disposition of such natural language distinctions as follows. The sentences which we all know how to apply to agents would include:[15]

(1) *A* desires/wants (that) *P*.
(2) *A* desires/wants to *F*.
(3) *A* wishes/hopes that *P*.
(4) *A* wishes/hopes to *F*.
(5) *A* likes, values, appreciates, is pro, is for, approves of/that *P*. That is, *A* has a desire for *P*.
(6) *A* likes, values, appreciates, is pro, is for, approves of/that *F*.
(7) *A* is/would be satisfied by *P*.
(8) *A* intends to *F*.
(9) *A* *F*-ed (intentionally).
(10) *A* mistakenly *F*-ed.
(11) *A* accidentally *F*-ed.
(12) *A* inadvertently *F*-ed.
(13) *A* negligently *F*-ed.
(14) *A* tried (un)successfully to *F*.
(15) *A* *F*-ed (un)luckily.
(16) *A* "choked".
(17) *A* compulsively *F*-ed.
(18) *A* unwillingly *F*-ed.
(19) *A* was coerced to *F*.

Theorists should also be alert for ingredients presupposed and functioning beneath the practical level at which natural language tends to function.[16]

This first condition is perhaps best seen as methodological. What follows is a list of intuitive and highly plausible properties of an action which theories should have to include and which the major theories seem to accept.

2.1 Properties of Action

Property 1: Supports singular reference and predication[17]

If there is singularity of reference to actions, then there are items or objects which support such reference, and which take predication. Theorists seem to agree on this but disagree, as we have seen, on which properties, and therefore which predicates, if any, are essential to these singular items. In order to respect this disagreement, while making use of the agreement, we shall henceforth use the neutral term 'object' for the item referred to in action discourse. The objects of the competing causal theories all qualify as possible supporters of singular reference: the object which consists only of the causal intentional state of the agent (Cause Theory); the object which consists of the event directly caused by the intentional state of the agent (Single Effect Theory); the object which is the chain of events initiated by the intentional state of the agent (Chain Effect Theory); and the object which is the chain of events which begins with the causal intentional state of the agent and ends with the goal event (Relational Theory).

Property 2: Informational

True action ascriptions, such as 'He intentionally embarrassed the government by the release of the secret papers', make reference to certain information about practical agency. For a Davidsonian Effect Theorist,[18] the information referred to is that the agent held the event he directly caused under some true description. That is, the informational content is that the action has at least one description

held true of it.[19] Thus, any true description under which the event was intended will satisfy this informational requirement. For a Cause Theorist, such as J. Hornsby, however, the information given or referred to is what is contained in the effective intention of the agent — the state which, on this theory, is the action. That information would include the means and the goal of the agent. Thus, on Hornsby's view, the action is "revelatory" of the agent.[20] For a Relational Theorist, such as J. Searle,[21] the incorporation of the intention with what was caused into the relational object which for them comprises an action makes the informational content of the intention a proper part of that object. So, for at least these three major theories of action, some information about the relevant agency is an essential property of that object. These theories vary in the amount and sort of information about agency they deem is essentially carried with an action attribution; but they each seem to find information-carrying about agency to be an essential feature of the object attributed. How much information is meant to be communicated with an action ascription will be decidable in the light of the uses to which these objects called 'actions' may be put.

Property 3: Causation

Since causation is the relation of change, an action, being a change, must be compatible with causation. If activity exists, then the relation of causation must hold between certain internal states of an agent and other external states of the agent.[22] An action, then, must be an object capable of sustaining this relation, either as cause, as effect, or as cause together with its effect. It would have to be an object which had the property of being either the cause of a bodily movement, or a bodily movement which had been caused by a certain mental state of an agent, or both. It would have to be an object capable of engaging in the relation of change.

2.2 Purposes of the Concept of Action

In addition to these three central properties of action, there are uncontested uses to which the object is put which bear upon the

nature and amount of essential information which the act object must carry. There are two generic purposes which a true action ascription allows.

Purpose 1: Explanation

With the information essentially imparted with a true action ascription we are able to construct an explanation of why change or causation occurred. If Holmes intentionally kicked the dog, then we can legitimately infer that his state of agency had certain corresponding properties — that kicking the dog was his goal, for example — which are causally explanatory of the object ascribed. This explanatory inference is possible only if the object attributed to him with such discourse gives us that sort of information. An act is an object which makes use of the possible correspondences between states of agency and states of the world and of the capacity of agents to represent and with fidelity cause what they cause. These relations allow reference to be made to objects which are not only explanatory, but therefore self-explanatory.[23] Moreover, if a true action ascription ascribes an object which contains explanatory information about the state of agency which was its cause, rather than merely implicating such a cause, then the essential informational content of the object called an action is greater than Davidsonian Single Effect Theorists have held.

Purpose 2: Moral Appraisal

Clearly there is a distinction between agency causings and non-agency causings. Although both sorts of causing may be normatively appraised or judged, for example, for their aesthetic or prudential values, only agency causings may be morally judged. Of the two, only actions are morally and otherwise normatively appraisable, while mere objects, events and their causings are otherwise normatively, but not morally, appraisable. If an action is an object which is in itself morally appraisable, then its nature must incorporate sufficient information about the agency which was its cause in order to give sense to that appraisal. In moral appraisal of an action, one appraises its agent. Only if, then, actions themselves contain morally relevant information about agents is moral appraisal of actions

coherent. Without the transparent informational connection[24] between intentional actions and their agents, only non-moral normative judgements of such objects would be possible. As the Law has it, a crime is not committed without *mens rea*. In its absence, we may have before us an *actus reus*, a proscribed state of affairs, but not yet a criminal item,[25] nor a moral one. Theories which take proscribed acts to be objects akin to the *actus* (*reus*) have objects which are only necessary for moral appraisal.

So, if we had only the information that the act-object of the Davidsonian Single Effect Theory provides, namely that there was agency involved in the causation of an event, we could, on that basis, make no moral judgement of the object caused. We would know that one perhaps could be made, or perhaps that it would be immoral to do that; but we could not yet make a moral judgement of the object being considered without further information about the agency involved. For an action, therefore, to be morally appraisable, for it to be, in itself, a morally reprehensible or admirable case of agency, it must essentially have certain properties which are transparently informational about morally relevant properties of its agent. The fact then that direct holistic moral appraisal of actions seems possible argues for more specific informational content in that object than what most Single Effect Theorists have admitted.[26]

In sum, then, the question of the nature and extent of the informational content of an action seems to be answered by the uses of actions. Where an action is attributed, there is attributed a set of properties which is explanatory of its causation and the moral appraisal of which is a moral appraisal of its agent.

3. GRADING THE THEORIES

The table which follows[27] shows how the objects posited as actions by the major theories fare on their accommodation of the fundamental properties and uses claimable for these objects. We shall describe again the theories and then take the reader through the reasoning used to justify the grades.

All the theories assume as a minimum the context of I-C-E; that a true action ascription may be made only when an intentional state,

understood broadly enough to include volitions, tryings and other such mental states capable of bodily causation, has been causal of the effect therein represented. We should not expect the members of a set of theories so clearly in agreement on the context they seek to organize to be separated by very much. One should be wary here of the philosophical instinct for the conceptual knockout. The theories before us are survivors. Within their mutual context, as we have seen, however, the theories separate themselves as follows.

(1) The Cause Theory says the act is the effective intention, or other conduct-causing rational mental state, I, of I-C-E.[28]
(2) The Single Effect Theory says the act is the event, E, which is the proximate effect of the causal mental state of I-C-E.[29]
(3) The Chain Effect Theory claims that the proximate effect, E, of the causal mental state, I, as well as the further events which are causally downstream from it, constitute the act.[30]
(4) The Relational Theory says that the full trio of I-C-E itself constitutes the action. The causal mental state, I, when it is effective, taken together with its effect, E, is the thing done.[31]

Consider each of these objects, first with regard to singular reference and predication.

3.1 Singular Reference and Predication

The Cause Theory has the following difficulty with this property. If we assume the likelihood that internal states such as intentions are theoretical, or if we foresee some other difficulty with direct (and/or autonomous) reference to mental states, and therefore accept that for unproblematical singular reference to actions, an objective or public referent is needed, then the Cause Theory of Action cannot score strongly on the item before us. Legal systems reflect this emphasis on objectivity when they insist on the *actus reus*, the *conduct*, as one necessary component of what is referred to as *the crime* in legal discourse. Although the Relational Theory includes the referentially troublesome mental term within the act-object, its reference is theoretically anchored, as it is not in the Cause Theory, by the effect it causes, presumably in the public world. This accords with legal

theory at least in the area of criminality: a *criminal act* requires not only the *actus reus*, or inadmissible conduct, or effect in the world, but an inadmissible causal mental state itself: the *mens rea*. So the Relational Theory may do better on matters of reference than the Cause Theory. Nevertheless, if, as the Relational Theory supposes, singular reference to an action were actually reference to a compound of two particulars, then referential appearances are somewhat deceiving and need explanation. In the absence of such explanation, a theory which would leave the apparent particularity of actions in place and could deliver on the other items would be superior.

The Chain Effect Theory has even greater problems with singular reference. The extremes through which this theory has taken its treatment of singular reference are instructive. One could hold, as J. Feinberg did,[32] that there are as many actions performed by the agent as there are events in the causal chain intentionally initiated. Or, a Chain Effect Theorist could hold, as J.J. Thomson did,[33] that there is nevertheless a single action since for her a string of causally related events is a single event. The first option for the Chain Effect Theory shows us the result of not taking singularity of actions seriously when singularity of agency exists and so letting the number of actions performed by an intentionally causal agent be determined by a one-to-one correspondence with the events in the ensuing chain. Contrariwise, the second strategy accepts the singularity of action where there is singularity of agency and collapses the number of events in the ensuing causal chain into a single event in conformity with that recognition.

The Single Effect Theory, however, is strikingly strong on singular reference and predication. It takes singular reference seriously as Davidson's program always has. Event theory, despite certain problems,[34] provides a sound basis for the requirements of singular reference and most theorists accept events, including mental events, as the particulars which support it.

3.2 Information

Each theory accepts that some information is given, integrally, with its version of the object attributed to an agent with a true action

ascription. They differ, however, as we have seen, on the nature of that information. We therefore withhold appraisal on their informational components until we question the adequacy of that information for explanation and moral appraisal, two of the major uses for which the information incorporate to an action must be sufficient. We turn first, however, to how the requisite fit with causation may sort out the theories before us.

3.3 Causation

Single Effect Theorists of Action, perhaps, tend to take singular reference and causation more exclusively than do others as determinants of the nature of action. Perhaps this is so because they come to action theory informed by certain views on both.[35] The requirement that an action have causal properties is taken by them to indicate that it is in the category of events. Moreover, they tend to take causation to be a relation between single events. As its dominance over the last three decades shows, these indications give the Single Effect Theory a strong case, not only for actions being events, but for their being the single event, often a bodily event, with which the nature of causation permits a case of agency to be associated.

With the exception of the Cause Theory, our other theories do not fare well on the causation requirement. The object posited by the Cause Theory may be an event and a singular event. That it is a mental event causes it, as we saw, to fare less well when judged by prevailing theories of the semantics of singular reference. But on adequacy for causation, it matches the Single Effect Theory.

If, then, one's best theory of causation says it is a relation which sits only between singular events as its terms, then a set of, say, causally related events will not be an allowable term in a causal relation (although, of course, one of its members may be). And if action is to be an item in a causal relation, then causally related strings of events cannot be actions. This would require our two remaining theories to either take lower grades on the causation property or offer alternative theories of causation as well as the

semantics of its terms, or show why the consequences of what an agent causes need be included with what he causes so as to form a single object. There have been just such attempts by some action theorists.[36] But no action theory as well grounded in causal and semantic theory as is Davidson's Single Effect Theory has yet appeared. Neither does there seem to be a reason to wipe out the distinction between act and consequence and create objects which include consequences as proper parts.

3.4 Explanation

Let us assume that the attribution of an action is meant to give us an explanation of the relevant behaviour when agency was exercised. The quartet of theories seems to accept this since in each case, although different in other respects, the information seems to be directed to that end. The Cause Theory wants an action to contain information which makes them "revelations of the human mind"[37] and therefore of the cause of the agent's behaviour. The Single Effect Theory takes an action to be just that intended effect of an agent's immediate or effective intention and therefore an object, an act, whose unique explanation in agency distinguishes it from objects, such as events, not so caused.

But whereas the Cause Theory gives us with its object a great deal of information about the agency whose behaviour is to be thereby explained, the Single Effect Theory tells us only that the behaviour was caused by an intention in which it was truly represented. This account would deliver no information about the agent's motivation or other relevant practical states with a true action ascription. In contrast, the Cause Theory, with its identification of the action with the effective intention, gives us explanatory information about the agent's behaviour, namely information about the relevant causal state of agency.

In addition, if accidents, mistakes and other non-standard actions are exculpatory, they must contrast in some way with intentional action. This contrast must consist in how the properties of the conduct of an agent are explained. Accidents, mistakes and intentional actions

have type-distinct causal and rational histories. And this explanatory information is part of each type-distinct object.

Similarly, therefore, an action attribution would, on the Relational Theorist's account, supply adequate explanatory information of the agent's behaviour. If the Chain Effect Theory includes the events which flow from the active agent, then its ascription of an action does not contain explanatory information unless intentional descriptions of the events were involved; but that would likely be a different theory.

3.5 Moral Appraisal

If actions are up for moral appraisal, then the practical information about the agent, which on both the Cause Theory and the Relational Theory comes along with a true action ascription, is adequate for such appraisal. Moral appraisal of an action then amounts, as it should, to moral appraisal of the agent himself. The Single Effect Theory and Chain Effect Theory would, however, have to add information, perhaps available, but on their accounts inessential to an action and not necessarily therewith ascribed, in order for moral appraisal of the act to be possible. These views, one might then say, *under-implicate* agency with their theories of action. The act-object attributed to an agent on the Single Effect Theory constitutes an under-ascription to the agent in that the object itself is devoid of information about the practical content — the motive and means — of his effective state. The under-implication of agency which results from the Single Effect Theory and the Chain Effect Theory makes what is ascribed inadequate both as an explanation of the event(s) caused by agency and therefore as a basis for moral appraisal of an action *per se*.

TABLE 1. GRADING THE THEORIES

		THEORIES			
		CAUSE THEORY	SINGLE EFFECT THEORY	CHAIN EFFECT THEORY	RELATIONAL THEORY
PROPERTIES	SINGULAR REFERENCE AND PREDICATION	NO	YES	NO	NO
	INFORMATIONAL	YES	NO?	NO?	YES
	CAUSATION	YES?	YES	NO?	YES
PURPOSES	EXPLANATION	YES	NO?	NO?	YES
	MORAL APPRAISAL	YES	NO	NO	YES

It is possible in the light of the above exercise to posit an object which exhibits the properties and uses of an action better than those considered and allows it to be the practical object it needs to be. Such an object would be, for these reasons, an attractive candidate for the semantics of action discourse.

What singular object would be causable by agency and contain certain adequate explanatory and moral information about its cause? We suggest that such an object consists of those and only those properties intended of the event caused by the intentional state of the agent. Such an object has the following characteristics:

(1) It would support singular reference: it is *the* set of intended properties of *the* proximate effect of the intention;
(2) It would be appropriately informational: its properties give exact information about the content of the practical state of the agent;
(3) It is compatible with prevailing and other theories of causation: its properties are those of an event which is caused by an agent;

(4) The informational content of the object ascribed allows for explanation of the agent's behaviour, giving, as it does, the reasons and therefore the cause of that behaviour;
(5) The informational content of the intentional act-object allows a moral appraisal of the act, *per se*, to make sense since it is just that same information which is necessary for a moral appraisal to be made.

We may now say more about an aspect of actions already alluded to. If with an intentional act ascription we get an explanation of what was caused by the agent, we also thereby get an explanation of the action itself since the action is just the explanatory part of what was caused, *viz.*, the part that was intended. Actions are thus not only explanatory objects, they are self-explanatory. The properties which constitute actions are just those which explain them. Such self-explanation may be what is behind teleological explanation. More, however, of this and the nature of the act-objects we posit, in chapter 4, 'The Semantics of Action'.

4. IS THE INTENTIONAL INTRINSIC OR EXTRINSIC TO ACTION?

The Single Effect Theory holds that an action is the event intentionally and proximately caused: an act is an object which consists of an event with this additional intrinsic property. It is what we shall call 'the event of agency'.[38] It is what Danto called 'the basic action', as if there were others, a view to which Davidson objected that it is not basic since it is the action. This object of the Single Effect Theory may be considered "under (many) descriptions." It has *variable polyadicity*:[39] thus an action may be fast or slow, may be done grumpily and in the morning. Saliently, however, it may be considered under the descriptions 'intentional' and 'unintentional'. Although an action, on this view, may be considered in the light of these two properties, i.e., under these two descriptions, neither of them is essential, and the object, when so considered, is not being considered *qua* act; rather, an act is being considered *qua* intentional. Actually, on this view, for an act to be "under a description" is for it to be considered in terms of certain of its inessentials.[40] This is what

fundamentally separates our view from the Davidsonian Single Effect Theory: we take the intentional to be the *essence* of action; Davidson finds it, when he finds it, to be an *accidental* part of action.

In order to address whether the intentional properties are merely incidental or central to actions, we need some independent clues to the conceptual relations between the intentional and action. Consider, therefore, the following.

There will exist in any natural language or other representation of our understanding of the world a primary set of concepts which classifies favoured phenomena according to a rather stable set of criteria and a secondary set of concepts which classifies the well-known recurrent aberrations of these favoured cases. Thus we have coins and counterfeit coins, ducks and decoy ducks, Burgundy wines and Burgundy-like wines, orange and mock-orange, coffee and ersatz coffee, osprey and near-osprey, acts and mistaken, accidental and inadvertent acts. To understand this standard-adjuster mechanism it helps to keep its probable purposes before us. Chief among these are:

(1) to facilitate exchange of information and
(2) to allow our language means for dealing with the world's unpredictability and novelty while
(3) preserving the language's stability and minimizing the growth particularly of its substantive vocabulary.

Consider the first. If, say, a predicate is to be productively usable by a speaker for a hearer, there will have to be a standard use pre-fixed between the two to allow the term to function effectively for the hearer when invoked by the speaker.[41] Otherwise the hearer will find himself constantly in need of adverting to the speaker's intentions, and/or the speaker will find himself constantly in need of explaining his words. But merely to have a standard case to which the convention would be fixed would not be enough. If we are to describe the world with a public language, there will be two competing factors at work. There is our need as speakers and hearers for fixity in the language so that the intentions of the speakers may be readily revealed to the hearer. There is, secondly, our need, if our language is to be adequate to the world it is meant to describe, for devices which

allow us to deal not just in its uniformities, but with its novelties, and multiformities as well. We need two contrapuntal sets of devices: devices of fixity and devices of flexibility. Our standard-adjuster mechanism gives us both. Naturally, and thirdly, we also want these needs served with an economy and perspicuity of vocabulary.

With the standard-adjuster mechanism we are armed with a "flexibility-device by whose aid, in spite of the limited scope of our vocabulary, we can always avoid being left completely speechless."[42] We do not then need to strike a new term when faced with a new turn in the world, for with the aid of adjusters such as 'like' we can bring to bear all of the riches of our already incorporated stock of terms. This allows us to defer the question of whether a new term ought to be struck or not. Without the deferment, allowed by 'like'-plus-substantive, our vocabulary would be needlessly inflated and often inadequate.

Thus, standard and adjuster terms are common phenomena in natural language. Two broad classes of adjusters are action adjusters and epistemological adjusters: the first marks well-known and recurrent deviations from a standard in our claims about actions; the second marks well-known and recurrent deviations from a standard among our knowledge claims.

Consider the following list of action adjusters:

(1) Accidentally
(2) *Mis*takenly
(3) *In*advertently
(4) *Un*intentionally

Each action adjuster cancels a feature or set of features of a standard case of action, thus marking an important way in which an action deviates from the standard. Notice that each of these terms except 'accidentally' contains an overt negative, a fact in keeping with its cancelling function. 'Accidentally', however, contains an overt negative in its definition: 'An act due to an *un*foreseen event, etc.' The attenuated cases of action are worthy of treatment as standard deviants because each marks an important way in which an action may fail. Each such well-known aberration, marked by an adjuster

term, is retained within the radius of the concept of an action. The deviant cases are kept within the ambit of the standard because the standard case is constructed out of what is of first importance to our linguistic community in this context.

The following things about these adjusters must be remembered:

(1) They mark standard or typical deviations from a standard.
(2) They are kept within the conceptual ambit of the standard concept for various good policy reasons.
(3) They each cancel a set of features of the standard case.
(4) They assert (or imply) specific ways in which the deviations are like the standard case.
(5) They assert (or imply) specific ways in which the deviations positively differ from the standard case.
(6) Since each adjuster functions by cancellation upon some feature of the standard concept to which it is attached, we can, by conversion, uncover these features affirmatively and originally set into the standard case itself.

Recall that our objective in this section is to show the relations between the ideas of the intentional and action. Most Davidsonian Effect Theorists[43] hold that they are totally distinct ideas. Our purpose here is to give reasons to reject this. The negative function of action adjuster terms such as, for example, 'mistakenly', allows, we claim, a certain set of features of the standard case to be denied while leaving the others in place. The claim that an act was unintentional asserts a generic cancellation of aspects of the standard case while 'mistake', 'accident' and 'inadvertence' are more particular deviations.

To rebut the implication or assertion that a case before us is deviant in either a generic or particular fashion and assert that it is, rather, a standard case is a linguistic move whose use is created by the presence of standards and adjusters. Thus there exists actually a triad of terms here: (1) a standard term; (2) its set of adjusters; and (3) its set of readjusters. The standard terms refer to favoured types of cases, to cases which we have invested with a standard or central function and which consist of packages of information pragmatically selected. The adjusters recognize certain important (and therefore standard)

deviations from the standard case as well as the need to accommodate unforeseen deviations. The readjusters are terms of rebuttal of the adjusters: they cancel the implications or assertions of the adjusters. The cancellation has the effect of reasserting the standard since what adjusters do is cancel a property of the standard (and insert the replacement property which typifies the deviation). A negation of the cancellation of a property of a standard case puts us back to the standard. In action discourse the main readjuster or placement-back-to-standard term is 'intentional'. Its function is to rebut any particular or generic cancellation asserted or implied from the standard case of action and reaffirm the presence of the properties cancelled. 'Unintentional' is a generic adjuster term which denies the generic property of intentionality of the standard case of action. It offers, however, no replacement property for the one it cancels. 'Mistakenly', 'accidentally' and 'inadvertently' are adjusters which not only cancel certain features of the standard case but offer typical replacement properties. Thus an accident is not only unintentional but its deviance is due to a particular history which is distinct from the deviations of mistake and inadvertence.

On this account of the interrelationship of these terms an intentional act is a pleonasm since it does not differ from an act *simpliciter* — just as real coins and real ducks do not differ from coins and ducks *simpliciter*, nor real tomatoes from tomatoes. The difference marked by readjuster terms is contextual. They presuppose that deviance from a standard was asserted or threatened and issue a denial. This denial as we have seen constitutes an affirmation that the case in question conforms to the conceptual standard. The intentional would thus be an intrinsic part or whole of an action, rather than merely an inessential property which may be referred to by the action being put under the appropriate inessential description.

What then finally separates our theory from the Davidsonian Effect Theory are different views on what uses these terms before us have developed. We connect their use to an important aspect of linguistic theory about the formation of complex substantives and the role of their satellites. The chief rival theory here treats these terms as predicates unconnected except as accidents to the substantives of action.

To recognize that actions have *variable polyadicity* is to have identified a half-truth. Jones may have buttered the toast, slowly, at midnight, etc., but his action has, we argue, a core of essential properties which is itself monadic, although that core may be described in "poly" ways by the use of its inessential properties.

A possible object of reference for singular action discourse is an event an agent with an intention proximately caused. Then, other matters of interest may be introduced by the addition of relevant descriptions of that event. We thus begin with a reference to a small object, as it were, and add to it as interest dictates.

But another possible object of reference for singular action discourse is an event, as above, but with and only with the properties intended of it. We claim this object is of greater interest and use and fits the discourse of action better than the other. If such an object did not already exist it would have to be invented. On this view we find our language with objects composed of packets of information whose content is in accord sometimes with both our interests in discriminating standard from non-standard occurrences of objects and our interests in communication with language as well. These are the objects we are interested in, and where, in natural language, conceptual investment takes place. Where our substantives are connected to such objects, pragmatics determines semantics. Is singular reference in this area connected, then, to the informationally larger objects, as we claim, or the smaller ones of the Davidsonian Single Effect Theory? Are the larger objects of interest buildable only by users of the language through their additional description, or the addition of properties, or are such objects at least often already built and ready for use? We claim the latter semantics is pragmatically superior, not just for linguistic purposes, but for the other interests of users of language as well. The real difference between our theory and the Davidsonian Single Effect Theory, and probably the others as well, is that, while Davidsonians can, by adding certain practical descriptions/properties to their object, *construct* the object of our theory, that object, because of its power, has already, we claim, been adopted and hardened into natural language by its community of users.

Since we claim identity between the intentional — what was intended — and action, some further description of the intentional is useful here.

5. THE INTENTIONAL

5.1 Against the Matching Hypothesis

Action is the intentional and the intentional is informational of its agency. If an action is an object revelatory of its cause, it is implied that intending and the intentional are at bottom the same notion. Since the synonymy of these two terms is not universally accepted, and its denial is indeed a canon of the Davidsonian Single Effect Theory, we must look at this claim.[44] Michael Bratman, a proponent, we believe, of the Davidsonian view, has recently argued[45] incisively against the matching of the items and content of an intentional action with those of an efficacious intention.[46] He holds, rather, that "although to A intentionally, I must intend to do something [i.e., there must be some intention which is efficacious], I need not intend to do A."[47] Bratman's ingenious argument is based on the following case:

> Let us . . . suppose that . . . two [video] games are known to me to be so linked that it is impossible to hit both targets. If I hit one of the targets, both games are over. If both targets are about to be hit simultaneously, the machines just shut down and I hit neither target. Both targets remain visible to me; so I can see which target I hit if I hit either one. And there is a reward for hitting either target. But I know that although I can hit each target, I cannot hit both targets. Still, I know it is difficult to hit either target, so I . . . decide to play both games simultaneously; I see the risk of shutting down the machines as outweighed by the increase in my chances of hitting a target. I proceed to try to hit target 1 and also to try to hit target 2. I give each game a try.
>
> Suppose I do hit target 1 in just the way I was trying to hit it, and in a way which depends heavily on my considerable skill at such games. It seems . . . that I hit target 1 *intentionally*.[48]

The argument which Bratman depends upon for this conclusion is as follows. First, we may assume that intentions have a *strong consistency requirement.*[49] We cannot efficaciously intend both x and y when we also believe that x and y are not capable of conjunction. Second, Bratman feels, one must admit that since the player tried to hit both targets, and the one he did hit was hit in just the way he was trying, " . . . and in a way which depended heavily on my considerable skill at such games,"[50] that the player hit the target intentionally. These two admissions would be fatal to the Matching Hypothesis: that what is intentional always matches what was intended. If the player intentionally hit target 1, then since matching requires that there be a fully correspondent intention, the player should have efficaciously intended to hit target 1. But since the antecedents for the intentional hitting of target 1 and the missing of target 2 are by hypothesis identical, the player must equally have efficaciously intended to hit target 2. But this conjunction of intentions breaks the consistency requirement and thus violates rationality. The Matching Hypothesis can on occasion, perhaps not all that rare, require irrationality.

We believe, however, that there is a defect in the second premise of Bratman's argument. The player claims that he is in the process of trying to hit target 1.[51] It is that intention which is efficacious for him. We must remember that the contents of this second premise are what Bratman believes allow him to infer that the player hit target 1 intentionally.

What this *reductio* by Bratman of the Matching Hypothesis needs in order to be fully presentable is a view of what intending to try, or trying, involves. We submit the following. Bratman, we believe, is correct in the view that doubt is involved. The reason that doubt exists, we suggest, is that whereas the player may have had some success in the past and, when successful, the commission of the relevant events may have been attributable to his skill, nevertheless, there exists a relevant epistemic gap in the player's intention formation and agency with respect to these events. Something relevant to a full, unconditional, agency claim is unknown and the agent is aware of this. We may of course progress from intending to

try to A to intending to A by means of learning or by otherwise adding the absent true causal beliefs needed for the unqualified intention and the ascription of full agency. In the case before us, the player does not know the covering laws which relate successfully to the desired events. He may know that he was successful and therefore retrospectively attribute causation and some agency to himself. It was his skill, he "retrodicts", that caused the success. But he does not have relevant beliefs about the causal laws which allow him to believe that the particular behaviour he employs will be successful. Therefore, he essays a particular bit of behaviour. He tries. Unless now, every case of intending is a case of trying, the agent, when he does not have such a relevant epistemic gap as described above will, therefore, armed with what he believes are causal beliefs which span the gap between his agency and his goal, report that he is intending to A. Since these two states of agency are contraries, intending to try to A is incompatible with simultaneously intending to A.

Bratman admits that the inference from the contents of the second premise to the intentionality of the hitting of target 1 is not solid. The player recognizes the game to be difficult and is doubtful of success and for that reason intends to try to hit target 1 (as well as target 2). It is his intending to try to hit the target which, together with its background, is efficacious of target 1 being hit.

Now if it is true of a player that he is *intending to try* to dribble the ball the length of the court, then it simultaneously cannot be true of him that he is intending to dribble the ball the length of the court since he cannot simultaneously have doubts about whether he can intentionally cause that series of events to take place and not have these doubts. The agent cannot both have positive beliefs about the causal relations between his agency and the event(s) he desires and not have them. With this incompatibility in hand, the Matching Hypothesis will then block the inference that the player intentionally dribbled the ball the length of the court by denying that the requisite antecedent matching state of efficaciously intending to do so (*simpliciter*) was present. It could not be, since the intention to try was. According to the Matching Hypothesis, the supposed presence of that intending would be necessary in order to say of the player that

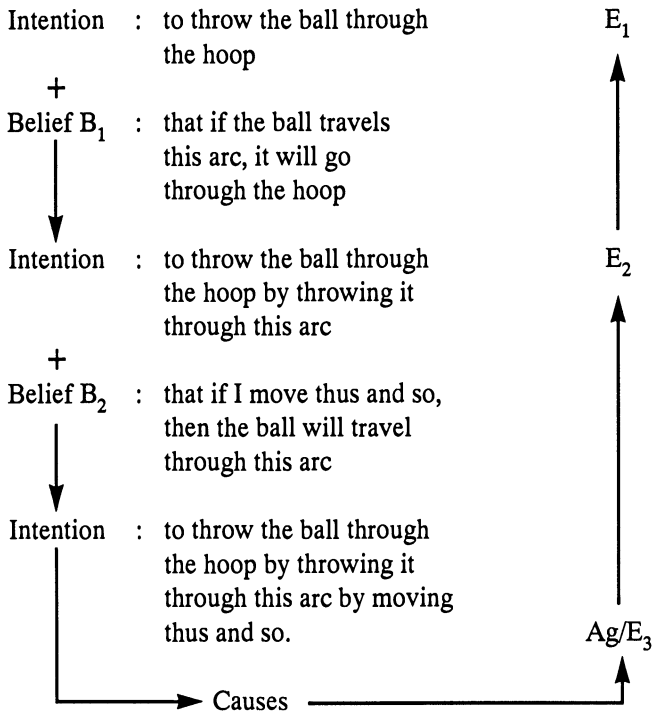
he dribbled the ball that distance intentionally. So if the player is intending to try to A, neither is he intending to A, nor would he intentionally A should the try — the experiment — be successful. To be allowed to assert of a player that he was both intending to try to A and intending to A (trying and doing?) would be to extinguish an important distinction in action attribution, namely, that there are relevant epistemic states of the agent which are important to our notion of rational agency and which are absent in one case and present in the other.

The Matching Hypothesis insists on preserving the distinctions in the mental causation between desires, beliefs, their justification and other qualities of the process of practical reasoning. It claims that to preserve these distinctions on the mental side just is to mirror them on the action side since that side is merely the diagnostic hypothesis usefully held about cases of intentional causation.

What must the Matching Hypothesis say, then, when the intending to try to A results in A, given that it there denies both that the player was intending to A and that he intentionally A-ed? The Matching Hypothesis claims that when an intending to try to A is efficacious of A, then it was a successful try or attempt. A successful attempt to A differs schematically from intentionally A-ing in the following way.

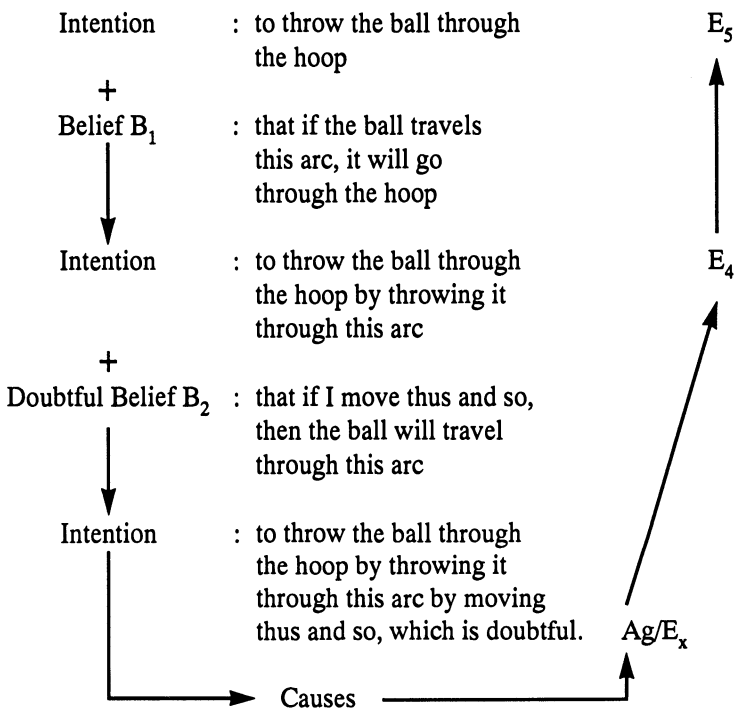
Intentionally making the free throw, or better, making the free throw *simpliciter*, is ascribed when the following state of affairs is the case: a string of causally related events in the world symbolized by 'E's and their causal arrows, and an implicated chain of cognitively causal mental states culminating in the efficacious intention, I, which causes the basic non-mental event of agency, symbolized as 'Ag/E_3', which has the intended causal relations to the goal event, E_1, the ball going through the hoop.

FIGURE 1. INTENTIONALLY MAKING THE FREE-THROW



Event E_1 is the ball going through the hoop which was caused by event E_2, the ball's travelling through the arc which was caused by event Ag/E_3, the agent's body moving thus and so.

FIGURE 2. AN UNSUCCESSFUL ATTEMPT TO MAKE THE FREE-THROW

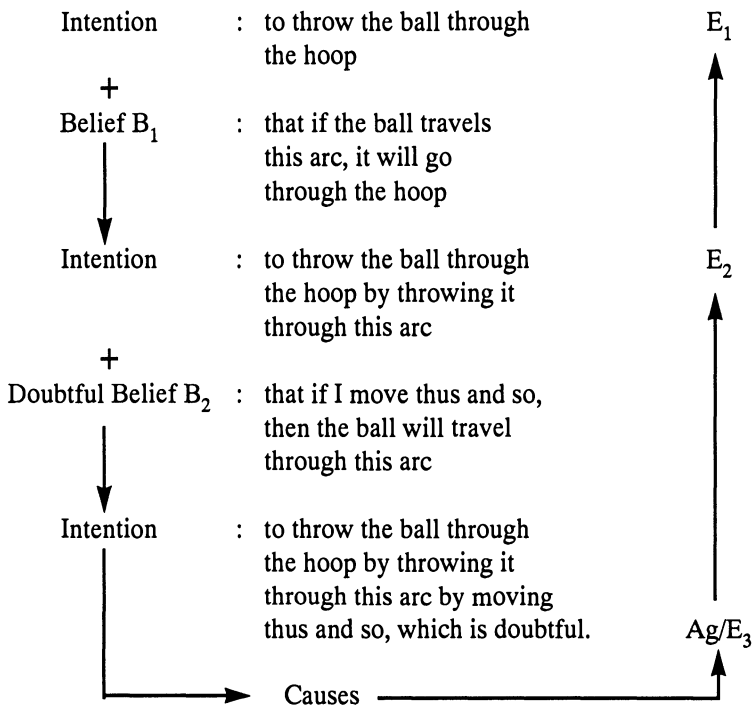


Let event E_5 be the ball missing the hoop which was caused by the event E_4 of the ball travelling through some arc other than the one intended which was caused by the event Ag/E_x of the agent's body moving in some way other than the way intended.

This representation of *unsuccessfully trying* shows us that the backup beliefs of B_2 are believed by the agent to have some gaps relevant to their justification and that in this case the doubt was warranted by the failure of Ag/E_x to cause E_2. (This would still be a case of trying, of course, had it been B_1 which lacked justifying backup beliefs.)

In a *successful attempt*, although B_2 remains unjustified, it is true and so Ag/E_3 does cause E_1 through E_2.

FIGURE 3. SUCCESSFULLY TRYING TO MAKE THE FREE-THROW



Event E_1 is the ball going through the hoop which was caused by the event E_2, the ball travelling through the arc, which was caused by the event Ag/E_3, the agent's body moving thus and so which he doubted he could cause.

With a successful attempt, we may credit the agent with cognitive causation of the event but not with the full cognitivity with which we are able to credit him when his causation was informed by justified

causally functional beliefs. It is only when the intention attains functionality through such information that we call the case intentional.

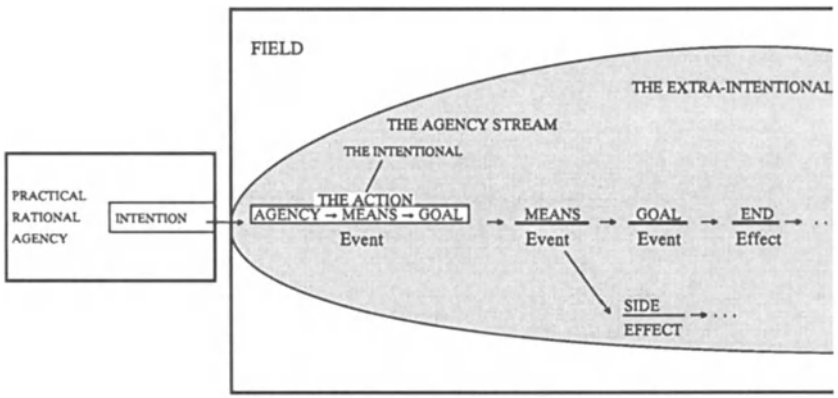
Choking is a complicated and interesting case combining certain aspects of intentionally A-ing, trying successfully to A and failing to A. An agent who chokes is one who in the past has been able to intentionally A but on this occasion has uncertainty about whether he has the means, e.g., to make the shot. But his uncertainty is *ex hypothesi* unjustified. What a choker is incapable of doing is accepting the unjustified doubt as doubt and proceeding to try. The unjustified doubt tends to cause either his beliefs or his intentions to fail him.

5.2 The Intentional

The intentional is best understood when set alongside its contrasts: the unintentional and the extra-intentional. One should therefore look at the entire logical space subsequent to a case of agency. The single event (directly) caused by an agent — the event of agency — is distinguishable from those events which are causally downstream. These *agency stream* events would not have occurred but for the event of agency together, perhaps, with certain contributory events. There are the non-agency stream events — *field events* — which are either causally contributory or non-contributory to the agency stream. The logical space subsequent to agency will consist of properties as well as events and their relations. Thus among the stream events will be a goal event, an event with an intentional property the desire for which was motivating for the agent. The stream will also contain means events which are markers of the causal route intended between the event of agency and the goal. The intended goal-functionality of the event of agency — the causal properties of that event for the goal — comprise the intentional within the agency stream of events.

In addition to the intentional, the stream contains consequences which we shall claim are not. Such events may be the result of the means chosen, or of the goal: side-effects and end-effects. The agency stream thus consists of the intentional and what we may call the extra-intentional. The area of the unintentional is reserved for streams

where, although agency has occurred, the intentional does not exist. These are the cases of accident and mistake and are covered in chapter 4. The extra-intentional and the unintentional comprise the domain of Negligence, a normative category of events and their properties which are in an agency stream but should not have been.



2

AGENCY: ITS TERMS AND RELATIONS

1. THE NEED FOR AGENCY THEORY

In the previous chapter, we argued that actions are objects which explain agency-causings. In order to fulfill this purpose the informational content of the object ascribed with a true action ascription must be determined by the practical and motivational content of its intentional cause. Where the agent's intention fails to be realized, the object ascribed must contain additional *extra-intentional* information which explains why the (mis-)intention had the effect it did.[1] All of the distinctions which we find useful to make among the objects ascribed with action discourse will have their source, therefore, in the articulations of the relevant representations. Thus, in order to specify the structures of the various act-objects, one needs an account of Practical Rational Agency. The act-object types *action, mistake, accident*, etc., will mirror, in their structures, generic structural features of agency and certain corresponding standard deviations therefrom. In this chapter we begin our analysis of the processes of agency, which we complete in the sequel. Ours is a causal *cum* functional analysis of natural agency[2] in which will occur states of a process which have the logical and functional properties of our psychological terms such as 'desire', 'belief' and 'intention', as well as other states which our basic terms and relations allow. We begin with the fundamental terms required for understanding agency.

2. THE PRESUPPOSITIONAL TERMS

2.1 Causation and its Relata

Action needs the relation of *causation* and causation needs singular terms as *relata*. The particulars which we nominate for the role of causal *relata* in action are events. We assume, that is, that semantic items such as *Holmes kicking the dog* are *events*. We would be open, however, as we are on other relevant metaphysical questions, to another type of particular so long as it was capable of entering into causal relations. We want our theory of intention formation only to maintain relations of consistency with all such relevant promising theory. For us, nothing really hangs on which basic semantic items are designated in the theory so long as the concept of action and its large *corpus* of concepts and relations survives as worthy of distinction. Even if certain terms within the *corpus* do not survive due to their incompatibility with other more powerful areas of theory, so long as the *corpus* largely survives, our concern, as action theorists, would be met.

Events, then, at the least, must be mentally representable and capable of standing in the causal relation. Such objects or particulars would have, by virtue of their causal properties, spatial and temporal properties. Events appear to satisfy these requirements which action discourse entails.

There is a tightrope here to be walked, however. We want, on the one hand, to make use of the idea of events as at least one essential semantic component of action discourse. On the other hand, we want to keep the notion we need as simple and unembroiled as possible in the controversies of event theory. We want our represented, and representing, objects to be at least datable, locatable, causal particulars, each with indefinitely many other properties or true descriptions. And we believe these characteristics are defensible in event theory.[3] This list of event characteristics is minimal and there may be others needed of these basic objects, or already implicit. Nevertheless, we begin with these objects and their properties and defer questions about the remainder until and if our theory of action discourse forces further properties upon these posited basic particulars.

One could disagree, as Davidson and Thomson apparently do, about the nature of the metaphysical notion of event which supports singular reference to actions. Davidson, we saw, maintains that an action is an event of a certain sort — a bodily event caused by a reason which rationalizes it.[4] Thomson, who argues that chains of causally related events can themselves be events, maintains, on one possible reading, that an action is the event which consists of the caused bodily event together with its intentional cause and the causal relation which binds them. Both accept events as the particulars needed for a metaphysics of action but disagree somewhat about their nature as well as about what an action consists in. We can accept either view of the nature of events. But we deny what they both assert: that either such a simple or a complex metaphysical entity or particular is adequate as an analysis of the object which is an action.

2.2 The Represented and the Representing

Causation, we have assumed, requires that our basic particulars be events. But we cannot make sense of *agency causation*[5] without particulars which are capable of representation as well as causation, and which can stand in the cause or effect relation to what they represent. We need, therefore, a notion whose instances serve as the representational and causal particulars of the theory, and a notion for the particular which is, or may be, represented and caused by instances of the first.

It is not necessary that the semantic objects of action discourse be of a unitary class. The mental states which are integral to action may be of a class separable from the equally integral objects caused through agency. We here assume, however, that both objects are of the same class, that there are mental as well as physical *events*, and so assume that the propositional attitudes and their intentional objects are events. Nothing irreversible will hang on that rarefied monism either since the theory we offer could be stated in basic dualist as well as monist terms.[6]

In summary then, the minimalist metaphysical or fundamental world-constructing items necessary for action are as follows:

(1) Causation;
(2) Events, or alternative particulars, which are independent of and representable by states of the agent and which are capable of causal relations;
(3) Events which are cognitive or representational of other events and which may cause events both of their own type and events of the type in (2) above, and which themselves fall into distinguishable sub-types by virtue of their distinct causal functions.

2.3 Distinctions Among the Attitudes

It is a commonplace of natural discourse about action to distinguish unique types from among those cognitive and causal mental events which are integral to agency. In our model of Practical Rational Agency, the familiar, if sometimes controversial, propositional attitudes are treated as states of an agent with cognitive content and a causal function. The first broad division among the agency attitudes is between the *epistemic attitudes* and the *active attitudes*. The class of active attitudes includes desires, intentions, wishes, hopes and compulsions. Each of these types will be distinguishable from the others by its unique causal function within the processes of intention formation and action. The active attitudes differ from the *epistemic* propositional attitudes by their causal role in the process and the nature of their cognitive content. The active attitudes alone represent what they cause.[7] The epistemic attitudes represent what caused them[8] and have effects in practical reasoning which we are about to describe.

In our model of agency, the cognitive and causal mental events fall into distinguishable sub-types at least by virtue of their unique causal functions. Thus, while beliefs and desires may be distinguishable by their cognitive content, they will also be distinguishable by their causal roles.[9] Beliefs, and the other epistemic attitudes, will have one set of causal roles in practical reasoning while desires, intentions and the other active attitudes[10] will have another. We leave open, for the

moment, the question of how the contents and causal roles of the cognitive and causal mental states are related, and return to it in our discussion of the generic functions of practical rationality in the following sections.

3. THE STATES OF PRACTICAL RATIONAL AGENCY AND THEIR FUNCTIONS

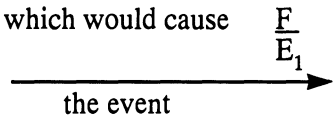
First, a note on our notation.[11] The events represented in desires and beliefs we symbolize with 'E'. In intensional contexts, such objects are representable by means of that 'E' together with some description, F, under which it is known, apprehended or cognized, desired or intended. So, we may represent the (E_1) embarrassment of the government (F) as the event, E_1, under the description, F: F/E_1.[12] That will constitute part, as we shall see, of the content of the desire for the government's embarrassment. We assume, subject to confirmation within the fuller theory of agency, that the *attitudinal* aspect of the attitude *desire* is causation and we employ '$\rightarrow$' as its symbol.[13]

3.1 The Structure of the First Fully Practical State

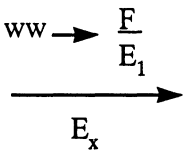
As action theorists we need in our explanatory arsenal the notion of a state of efficacy, $\rightarrow$, with respect to a representation or description, F, of a part of the world, E_1: thus, the simple state, $-F/E_1\rightarrow$,[14] a state of cognitive efficacy which the coherence of the concept of action presupposes as one of its antecedents. However, if the object of a desire *to embarrass* the government, for example, is the *coming to be* of that embarrassment, it would be wrong, using our notation, to render that desire simply as $-F/E_1\rightarrow$. The simple state so represented is a causal state with respect to the embarrassment of the government *tout court*,[15] rather than for its causing. The intentional object[16] of the desire to embarrass the government is the event(s) which would bring that state of affairs about.

A desire, in contrast to $-F/E_1\rightarrow$, is a state whose structure is already fit for rational agency, by design already a perfect antecedent instrument for action. The structure of a desire-that, we claim, is a causal state, $\rightarrow$, with respect to an event, E_x, which is the one or sort which is believed would cause the motivating event, F/E_1 (in our

example, the embarrassment of the government). A desire that a motivating event come to be the case is a state with the structure:



which we further symbolize as:



It is a state of causation, $\rightarrow$, with respect to the event, E_x, which would cause, ww$\rightarrow$, the goal event, E_1, which is F.

A desire, then, is an attitudinal or functional state, informed in its structure by the belief that causation is necessary to make actual a non-actual and causable state of affairs. A desire state contains a representation (F) of its goal event (E_1) which is used to commence the identification of the event (E_x) that the agent will select as the cause or change in the world believed necessary and/or sufficient to produce the goal. This identification of E_x by the accumulation of sufficient identifying descriptions or properties is part of the role of practical reason. Only with this form and function of the state of desire in place is the language of means-end and/or teleology introducible. Only an active attitude which has a representation of a causing which is incomplete in this way, and completable by belief, poses the question which it is the function of 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.

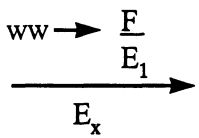
3.2 The Fundamental Pre-Practical Syllogism

A desire, then, is not the simple state it appeared to be, but a conclusory[17] one which is already the result of a *fundamental pre-practical syllogism* necessary for action. This fundamental syllogism is composed of two premises — two causal or functional states in a rational system — whose conclusion provides the framework for the practical task of getting a sufficient description of the event whose causal properties it is believed would make actual the goal event, F/E_1, represented in the motivating mental state, $-F/E_1\rightarrow$. The states which are premises to desire may be read off, then, from the complexity of the description which rationality has there placed. The object of a desire is the event, E_x, which it is believed would cause, $ww\rightarrow$, the goal event F/E_1. The goal event, F/E_1, is the intentional object of the *proto-desire* $-F/E_1\rightarrow$. When this state is combined in pre-practical reasoning with the belief in the necessity of causation, causation shifts from the object, F/E_1, of the antecedent proto-desire to the object which would cause it, the object E_x, which it is believed would have causal relations to the goal event, F/E_1. The intentional object of a desire-that has this form. The minimal description of the object of a desire-that is: 'that event with the causal relations to the actuality of the object of the proto-desire':

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

We may, then, excogitate two states which are causally and rationally prior to the desire that the government be embarrassed. The first is a state of causation, an attitude, to the embarrassment of the government: the proto-attitude $-F/E_1\rightarrow$. Such a cognitive or representational state uninformed by belief in causation might occur, as we have seen, but a system which remained in this state would be in a state of efficacy with respect to an object but without the conception of means: an agent without any way of representing *how*.

The second state, the belief in causation,[18] is necessary for the question of means to come forward. But a belief in the non-actuality of the object of the proto-desire is also necessary in order for the attitude attached to the representation, 'F/E_1', to shift to the cause of that object, *the coming to be of* F/E_1. This new object of the attitude is the event, E_x, which is believed to have the causal properties needed for F/E_1. The subsequent state is, again, the desire that F/E_1 come to be:



3.3 The Pre-Practical Active Attitude

The presence of such a simple causal and representational state as $-F/E_1\rightarrow$ is excogitated from what the notion of an action requires as antecedent. The state of being causal with respect to some event F/E_1, thus the state $-F/E_1\rightarrow$, is a possible state but it is devoid of the content which the basic belief in the necessity of causation would add. $-F/E_1\rightarrow$ is a mental state which does not represent the world as in need of change (causation) in order for F/E_1 to come to be. Its content is not then an adequate representation of the sort of world to which agency must connect since $-F/E_1\rightarrow$ is comprised only of the attitude of causation and a representation of a certain possible state of the world. Such a simple representation lacks the capacity to accommodate an essential feature of agency: that the means to a desired possible state of the world be addressed. Only beliefs about the causal structure of the world and mental states which are open to such content allow that connection between goal and means to be made. $-F/E_1\rightarrow$ is a state whose representation is blind to the causal structure of the world, as yet incapable of accepting causal beliefs. If an agent were to affect the world on the basis of that state alone and were to cause F/E_1, it would do so as motivationally as one billiard ball affects another: that is, where the representation of F/E_1 plays no role in the causation. In order for 'F/E_1' to play a role in the causation

of F/E_1, this representation itself must be part of a causal representation.

The simple state $-F/E_1\rightarrow$, being a state of pure nascency, is not, therefore, capable of entering the process of practical reasoning which searches for means to its end. It is a state not yet structured to accept the effects of practical beliefs about how to cause F/E_1. It would be a state, as we shall see, also incapable of satisfaction: not now because it lacks the representation which would be supplied by the causal belief, but because it lacks the representation to which the belief that F/E_1 was the case could be relevant. It is most of all a state which, although a prelude to action, is not yet involved in the rationality of that enterprise. None of the above is meant to assert that the mental state $-F/E_1\rightarrow$ cannot have its structure added to, so that practical beliefs are relevant to the resultant state, but rather that it is not as yet a state capable of such rationalization. It is a *pre-practical* state.

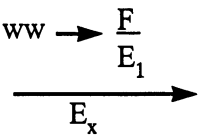
4. CONCATENATION OF CONTENT AND THE FUNCTION OF BELIEF

4.1 The Belief Function in Content Concatenation

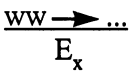
Practical reasoning is a process by which a desire's incomplete representation of its intentional object is completed.[19] It is a process whereby an insufficient description of the event with causal relations to the goal event becomes sufficient. In order for the identificatory content of a desire to be sufficiently completable, desires must be capable of taking on information or content about such causal relations from beliefs. The function of taking on such information defines the relationship of desires to practical reasoning: they can accept and store practically relevant information from beliefs. Desires are thus transformable by beliefs. The impressibility of the content of desires by the content of relevant beliefs is the essence of rationality in agency.

As we saw earlier, the belief that causation is necessary to make actual any non-actual but causable event would impress that content,

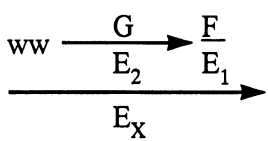
where the system is rational, upon an active attitude such as $-F/E_1\rightarrow$. The attitude becomes:



where the content

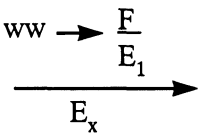


which has been added to $-F/E_1\rightarrow$ is the content of the belief in the necessity of causation for the non-actual and causable goal event. Suppose, then, that an agency system is in such a desire-that state. The occurrence of such a state sets the task for practical reason of providing a further identifying description of the event, E_x, which is as yet described only as an (the) event with causal relations to the motivating event, F/E_1. Suppose, further, that the agent believes that the release of the secret papers, G/E_2, would be sufficient and/or necessary to cause the embarrassment of the government. In practical rationality, the effect of this belief is to transform the antecedent desire into the desire for an event under a new and more particular but perhaps not yet sufficiently identifying description: it becomes the desire whose intentional object is the event, E_x, with causal relations, ww$\rightarrow$, to the, E_2, release of the secret papers, G, which in turn has causal relations, ww$\rightarrow$, to the, E_1, embarrassment of the government, F. The structure of this resultant desire state is:



the desire, $\rightarrow$, for the event, E_x, which would cause the event, E_2, of the release of the secret papers, G, which would cause the motivating event, E_1, of the embarrassment of the government, F: thus the *concatenation* of the structure of desire, and, as we shall see, ultimately of the structure of intention as well.

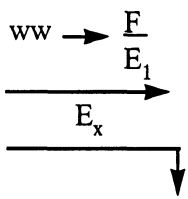
So, one of the functions of desire is to accept and concatenate the content of the beliefs which inform or affect them. Another desire function in a rational system is the causing of certain beliefs. The desire for the coming to be of F/E_1 could or would, in a rational system, cause the belief that the agent so desires. The representational content of the belief would be of that desire state. It is:



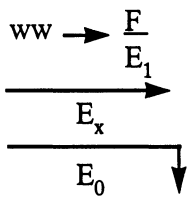
This structure forms only a part of the representation of the belief state itself, lacking, as it does, representations of the belief attitude and the object of the belief attitude which we have yet to identify.

In our model of agency there is a function, the belief function, which is limited to affecting the content of desires and other beliefs by transferring information to them, and (with Hume) which lacks the function of affecting the (rest of the) body directly. Beliefs, then, 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 its functional difference with an arrow, ⌐↓, directed *downward*, as it were, upon states within the system and not at anything external to it. We mean this to indicate that the causal role of belief is entirely the transmission of content to internal states of the system.

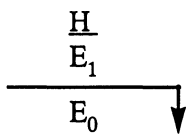
To further complete our representation of the belief that the agent desires the coming to be of F/E_1, we therefore add the symbol for belief function, ⌐↓, at the place in the above structure of that belief's content analogous to the place for the symbol for desire function in desire:



What remains to be added to complete the representation of this belief state is a symbol for the intentional object of the belief: the object upon which it will function. If beliefs are causal and transfer the representation they contain to their objects, then their objects, as distinguished from their contents, are those states to which they may cause their content to be passed. Let us, for these purposes, suppose that the object of this belief state is some other state of the system, E_0. Therefore, the complete structure of the belief that the system desires the coming to be of F/E_1 is:



The belief that the, E_1, cat is on the mat, H, would be represented as:



or, in "natural" language, the causal state with respect to that (type of) event, E_0, to which the content or representation H/E_1 is to be passed, this being the function of such a belief state. The event upon which the representation is impressed is the causal object of the belief but not its "grammatical" object which is, however, how its content is identified. The belief state itself is that object which bears the entire above structure: the object which is causal with respect to the transferral of some representation, H/E_1, to some object, E_0. If truth and causation are to be co-preserved, then E_0, the causal object of the belief, must itself be an object capable of and susceptible to the amalgam of efficacy and cognitivity. The objects of beliefs will then be other beliefs and other intentional or attitudinal states. The representational content of beliefs is derivable from both non-

cognitive and cognitive states of affairs: the physical as well as the mental. Although the content of beliefs 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 then that we have cognitively efficacious states with quite different causal capacities. One set, beliefs, is capable of causation which consists in affecting the content of the other main set, desires. A desire is capable of causing change in the physical world according, startlingly but presumably, to its representational content which is affectable by beliefs. These capacities are clear and close to the surface. As we proceed with our account of the antecedents of action, subtleties will emerge which are not visible on the surface.

4.2 The Belief Function in Content Decatenation

In our account of the belief function, we have so far identified the cognitive function of *informing* an active attitude state by supplying content relevant to causation of the object of that active attitude. We also mentioned that belief is implicated in the cognitive satisfaction of an active attitude. That this is so requires that belief states have another function besides the addition of content. We suppose an agency system with the desire that F/E_1 come to be. Suppose further that F/E_1 comes to be believed actual. The content of this newly acquired belief contradicts certain content already embedded in the desire that F/E_1 come to be, namely, the need for a cause of F/E_1. The belief that F/E_1 is actual entails the falsity of the belief that its causing is necessary for its actuality. The representation which, in this reversal of the process of practical reasoning, is now *decatenated* from the content of the desire that F/E_1 come to be is:

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

That content is the event, E_x, under that description. The resultant active attitude thus lacks a representation of the event, E_x, which was believed causally necessary and sufficient for F/E_1. That representation had been derived from the antecedent belief in nonactuality and causability and its addition to the desire for F/E_1

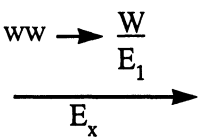
transformed that state into the desire that F/E_1 come to be, the causal object of which was the event, E_x, which would cause F/E_1. With that belief content now decatenated from the desire that F/E_1 come to be, the attitude may stay with the content F/E_1 so that the resultant state is $-F/E_1\rightarrow$, the proto-desire for F/E_1, or perhaps the *liking* of F/E_1, where that cognitive attitude is unimpressed by the belief in the necessity of causation.

Thus, in decatenation of this form, the belief that the goal event of an active attitude is actual deactivates that attitude, renders it non-causal, by causing the removal of content which is essential to the process which leads to action. Such "de-effication" is an essential ingredient in satisfaction of a desire. Where, as we shall see, the belief that F/E_1 is actual does not shut off the pursuit of F/E_1 and the procession of the desire that F/E_1 come to be goes through into action, we find various forms of irrationality which categorize various aberrant relations between desires and beliefs, and therefore also between cognitive causal states and the rest of the world.

4.3 The Practical Effects of Certain Beliefs

A belief, then, is an item which may play several roles, each of which may be seen as a distinguishable but related causal function in the process which may or may not culminate in action. First, a belief may interact with a desire in practical reasoning where the efficacy of the belief is to cause its content to be added to that of the desire thereby further identifying the desire's object in what amounts to an increase in rationality. If the process is rational then only relevant content will be added. In practical reasoning, at least, there is an underlying relational question at work which determines which beliefs are relevant. The first question posed by the desire state is about which event or events (which will subsequently be evaluated and thus chosen among) are believed by the agent to be those which have causal relations to the goal event. As a result, only beliefs about which events stand in the above relation to the goal event are relevant to this stage of practical reasoning. Beliefs thus perform an enabling or, as we have seen, a disabling function in the action process, desires a motivational one.

Second, certain beliefs can stop or disable a desire state from going through to action. Imagine the agent in a proto-desire state for the end of the war, $-W/E_1\rightarrow$. In addition, he believes that causation is necessary for all events. This belief, we have seen, puts the agent into an active attitude with respect to that event, E_x, which would cause the end of the war, W/E_1. Thus the state:



The agent may then have four typical modalities of belief about the likelihood of W/E_1 being caused, or, equivalently, about the likelihood of there being an E_x with causal relations to W/E_1. He may believe: (1) that it is impossible nomologically or logically; (2) that it is not likely; (3) that it is already actual; (4) that it is nomologically possible of causation and non-actual. Suppose that the agent does desire that the war come to an end, and then has each of the above beliefs in turn.

1. If we believe that desire states concatenate or take into their structure the information from relevant belief states, then, when impressed by the belief in the impossibility of W/E_1 coming to be, the above desire that W/E_1 come to be becomes one whose content is a desire for the coming to be of an event, W/E_1, and, through concatenation, a coming to be which is believed impossible. That structure describes a *wish*. But in rational systems, wishes do not function as motives to action. Wishes have structures which do not allow the right questions to be asked, viz., "What is a further specification of the coming to be of W/E_1?". The structure of a wish shows that the system holds a belief, namely, the impossibility of the causation of W/E_1, which is incompatible with beliefs about how W/E_1 could be caused. A rational system could not be in a state of assent to both beliefs. This means that since it holds the first belief the system has no beliefs which could satisfy the existence demands which a desire would put upon practical

reason. A state with the structure of a wish would, in a rational system, be shunted aside, off the practical reasoning rails. It is not irrational to be in such a state and it is a non-practical rather than an impractical state since, being "parked", it does not exercise practical reason in a rational system the way desire does.

2. The agent may believe that the war's ending is not likely. In this case we have a *hope* and perhaps some practicality for the system to remain in that state since beliefs about likelihood are sometimes revised.
3. The system might contain the information that W/E_1 is actual. This makes practical reasoning otiose rather than empty of relevant beliefs as did the first alternative above. The present case is actually similar to the satisfaction of a desire-that, although how W/E_1 was caused may not be due to the agent. With the belief that W/E_1 is actual we get reversion to the proto-desire $-W/E_1\rightarrow$ or its extinction, unless our desire attitudes, which include our appetites, are obsessive in which case the belief in actuality may not be sufficient for extinction or for more than momentary reversion.
4. Finally if our desire that W/E_1 is accompanied by the belief that W/E_1 is possible but non-actual, the resultant state, again, in a rational system, engages further the process of identifying and causing the event, E_x, which has the causal properties believed relevant to causing W/E_1.

We may thus see the causal role or function of belief in action as one which adds or subtracts content to those other states — the motivational ones — which drive us through practical reasoning and ultimately may affect the world. This addition of content to the motivational states allows them to be identified as states which the system will either shunt aside, cause to revert or allow to progress. The belief functions of picking up and passing on information which are therefore restricted to interaction between cognitive items are functionally fundamentally distinct from that of being motivational, which is to say, causal of non-cognitive as well as cognitive items, where the causation does not consist in the passing on of information.[20] That functional bifurcation is what separates beliefs

from desires. A rational system, then, consists at least of a function, the causal function shared by proto-attitudes, desires and intentions, to whose role information is essential, and a set of gates through which the causal function must pass. This gate function, shared by identificatory and evaluative beliefs, provides information which either further determines the first function or shuts it down. Only if we find that bifurcation unnecessary because we find, for example, the functional roles non-distinct[21] in a theory of action and agency, may we denigrate belief-desire psychology.

Here one is reminded of the sentiment expressed in J.L. Austin's words about the dismissibility of our common language distinctions especially in this area:

> ... our common stock of words embodies all the distinctions men have found worth drawing, and the connexions they have found worth marking, in the lifetimes of many generations: these surely are likely to be more numerous, more sound, since they have stood up to the long test of the survival of the fittest, and more subtle, at least in all ordinary and reasonably practical matters, than any that you or I are likely to think up in our arm-chairs of an afternoon — the most favoured alternative.[22]

4.4 The Structure and Function of Belief

If it were true that each type of practical attitudinal state had typically distinguishable content, then we might have simplicity of theory as a reason to take those differences as explanatory of the differing causal roles of the attitudes, or, equivalently, explanatory of their unique positions in the process of practical rational agency. Each attitude-type's role would 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 role 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 rationality and causes the sequent state in the process. Suppose, for example, that a desire state harbours the following 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 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 an 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 event (whatever in the further development of cognitive theory that will turn out to be). Desires and the other *active attitudes* will 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 only sufficient completion of its incomplete definite description. For such causation to occur, a description is needed which is sufficiently identificatory. The function of belief, however, is to add further descriptions to the definite description structure which is in the process of sufficient completion by the rational practical process. 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. 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 role of the belief's predicate place to add its information to, and thereby cause, the next rational 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 epistemic attitudes are properly named 'propositional' attitudes. The others should be called 'practical' rather than 'propositional attitudes'.[23]

4.5 Content Concatenation and Practical Reason

The representational content of the proto-desire $-F/E_1\rightarrow$ 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 role 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 'motive', 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 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 allowed it to both concatenate and decatenate content from beliefs put 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 $-F/E_1\rightarrow$. 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 $-F/E_1\rightarrow$ has been the occasion of an action

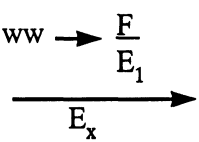
schema or process, then the representation, F/E_1, will occur in all subsequent states leading to action including desires, hopes, wishes, beliefs and intentions. The inheritance of the first term by all subsequent mental states which lead to action is the basis of our notion of concatenation of content in that each segment in the process leading to action concatenates this content from its antecedent as well as all other content accreted in the process. 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 role 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. Wanting some movement of your body would be such a case. The motivational role 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.

Together with the initial desire state, we saw there was an acceptance of the basic causal belief and a resultant structure which has as object the event, E_x, which would stand at the head of a casual chain leading to F/E_1. This states the problem for practical reason: identify that event which would best head such a causal chain; such a chain must essentially make use of the description, F, which must therefore be retained in all pre-action intensional states of which $-F/E_1\rightarrow$ is the motive. Concatenation of descriptions of the agency event which would head the chain leading to F/E_1 continues until that event is satisfactorily identified for the agent.

As we have claimed, the attribution of action itself cannot be made properly useful without this intensional element of event identification being an essential part of the attribution.

4.6 Full Decatenation of Content

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 $-F/E_1\rightarrow$ is further dissociable. Consider the meditative process, 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 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 $-F/E_1\rightarrow$'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, $-F/E_1\rightarrow$, 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,



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[24] or agency-unrelated content. We should notice the similarities between this

process of content removal and that effected 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 the active causal function so that it is denudable of content leaves the total 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 lies 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. We address the relation between full decatenation and autonomy somewhat more in the section on autonomy.[25]

5. DISTINCTIONS BETWEEN THE PRACTICAL STATES IN THE PROCESS

It begins to emerge, then, that the process of rational agency will consist of states with the following properties:

(1) type-distinct structures or contents, which determine a state's
(2) 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.[26] That this possibly typical

difference of structure is a determinant of the distinct[27] causal roles of the attitudes themselves is also debatable. We shall argue for the truth of the first and offer a reason to accept the second.

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 as two distinct attitudes, 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. We cannot therefore treat a desire-to and a desire-that as completable by the same, and therefore interchangeable, content.

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 causal with respect to an event of one's agency which would cause the sinking, that action, and which relational property 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.

With the basic terms and relations of practical rationality before us, we turn to an account of the practical rational processes within which the basic terms function.

3

THE RATIONAL PROCESSES OF AGENCY

1. PRE-PRACTICAL REASONING: FROM PROTO-DESIRE TO DESIRE-TO

1.1 From Proto-desire to Desire-that

Our theory of agency asks us to remain within the realm of the cognitive: it is a theory meant to describe the nature of the relations of the cognitive states required by a theory of agency. It is a theory which takes us from those cognitive states which are emanations of the physical at one end to those which emanate in the physical at the other. An action, then, being the expression of cognitive causation at the end which emanates in the physical, implicates a state like desire at the other. This state must have the properties of cognitivity and efficacy in order for what emanates in action to have a mental aetiology.

Assume a proto-desire, or some state, however named, the content of which is the government's embarrassment. As the reader recalls,[1] we render this state as causation, '→', with respect to the event (E_1) under the representation, 'the embarrassment of the government, (F). This causal state, or proto-desire,[2] for the embarrassment of the government, we then represent as:

$$\frac{\text{(which is) the embarrassment of the government}}{\text{the event}} \longrightarrow$$

or, $-F/E_1\rightarrow$. The embarrassment of the government itself we represent as F/E_1.

As we recall, a "later" more rationalized attitudinal state with the content *that the government be, or become, embarrassed* is conclusory in that we may excogitate two earlier cognitive states which may be seen as premises to it, or as states which, in a rational system, cause the above conclusory state. A causal state with the content *that the government become embarrassed* may have, as a cognitive and causal ancestor, a causal state, the proto-desire D_1, $-F/E_1\rightarrow$, with the content *the embarrassment of the government*, which is uninformed by beliefs about actuality and causability. It is, unlike its descendants, consistent with beliefs in the actuality or non-actuality and the causability or non-causability of that state of affairs. One can, therefore, be in this proto-desire state and hold either belief without irrationality. But if a rational system has the proto-desire for the embarrassment of the government and the general belief in causation, then, given the belief in the non-actuality and causability of that motivating event, it will be caused to be in a causal state toward a new object, one with articulation derived from the above complex of beliefs, B_1.[3] It will be an attitude whose intentional object is the event with causal relations required to actualize the embarrassment of the government. Or, as it is more commonly referred to, it will be the desire, D_2, that the government be embarrassed, the desire for that embarrassing: a causal state with respect to its representational content *the causing of the government's embarrassment.*

This state, the active attitude whose object is the causing of the motivating event, is the *first conclusory state* in the genesis of an intention and is the effect of a proto-desire premise and premises consisting of the belief in the non-actuality and the causability of the object therein represented. Thus it may be seen as the *proto-practical state*,[4] one which, in its form, is open to the practical questions of how the object, F/E_1, of the proto-desire may be caused; of how the causing of F/E_1 may be identified. Proto-desires are not yet states with that practical capacity.

Although a proto-practical state, such as the desire for the causing of the embarrassment of the government, may function as a premise to reasoning about the identity of possible causal pathways to that motivating event, such reasoning would not yet be *practical reasoning* proper which we reserve for that part of the process of intention formation which identifies the agent's causal pathway to his goal which begins with an event of his (or some other's) agency.

Consider now some possible futures of the proto-practical state, the desire for the causing of the government's embarrassment, D_2, when combined, in rationality, with the following beliefs. Suppose, first, that the causing which is represented in the active attitude is believed, B_2, impossible. Since it would be irrational to devote the resources of agency to searching for the causal pathway to an event believed not to have one, the effect of this belief is to remove the desire from the practical process so that it is "parked" as D_3 or otherwise set aside. This is not to say that the belief in non-causability necessarily expunges the efficacy of the desire: that state itself may survive, or it may survive as its progenitor, the proto-desire for the embarrassment, $-F/E_1\rightarrow$. While the agent, as a result, is no longer in the action mode with respect to F/E_1, he can be returned to the line of connection toward action by the change of belief from the impossibility to the possibility of causing F/E_1.[5]

Suppose, then, that our agent, *PRAGMA*, believes, B_3, that a causing of the embarrassment of the government is possible. The effect of this belief is to allow the efficacy of the desire to proceed through to the next stage of intention formation rather than have its path blocked as it was by the belief in the impossibility of causation. We may think of the belief in the possibility of causation as opening a gate to the next stage in this process and the belief in the impossibility as opening a gate to merely *hoping* or *wishing*.[6] Thus, the proto-practical state, the desire for the causing of the government's embarrassment, becomes, by virtue of the belief in the possibility of causation, a premise or input to the enquiry about whether that causation is possible for *PRAGMA* itself: i.e., whether there is a causal pathway to the motivating event which begins with *PRAGMA*

itself. (Figures 1 and 2 represent the process of pre-practical reasoning: from proto-desire to desire-that.)

The order we describe of the states in this process exhibits a portion of a theory of rationality. But it should be recognized that some plasticity in the order is possible. For example, if the desire, D_2, for the causing of the embarrassment of the government is to have a rational history or itself be a rational state, then the belief, B_2, that this causation is possible[7] must be held in the system. It is not rational to proceed in the practical vein without this belief. So the order is not plastic on this point. But it is plastic on what results from the denial of this belief. If the system holds the belief, B_2, that the causation of the embarrassment of the government is impossible, then several alternative states — either D_1, D_2 or D_3 — may result and rationality be preserved.[8]

1.2 From Desire-that to Desire-to

The belief, B_4, that *PRAGMA* has at its disposal an *entry point*, as it were, into the world — an event it can cause — which would have the causal relations sufficient for the embarrassment of the government causes, *ceteris paribus*, the desire, D_4, for a possible embarrassing of the government to be transformed into the desire, D_5, *to embarrass* the government. This is the desire that *PRAGMA* itself be the agent of the change which is sufficient for the goal.[9] The belief that there is an entry point for *PRAGMA* is an important further articulation of the causal object, E_x, of *PRAGMA*'s desire in that the search for the identity of that event has now been greatly narrowed. *PRAGMA* now desires, not merely some event which would be sufficient and necessary for the embarrassment, but an event of its agency, Ag, an event it can cause, which would be sufficient and necessary for the embarrassment. Heretofore, in merely desiring that the embarrassment occur, there was too little direction or representation of the event E_x, which was the object of the previous active attitude, the desire-that, to invoke practical reasoning. Only desires-to are proper inputs to practical reasoning since only then is the agent attempting to identify and evaluate an object of its own agency.

FIGURE 1a. PRE-PRACTICAL REASONING: FROM PROTO-DESIRE TO DESIRE-THAT

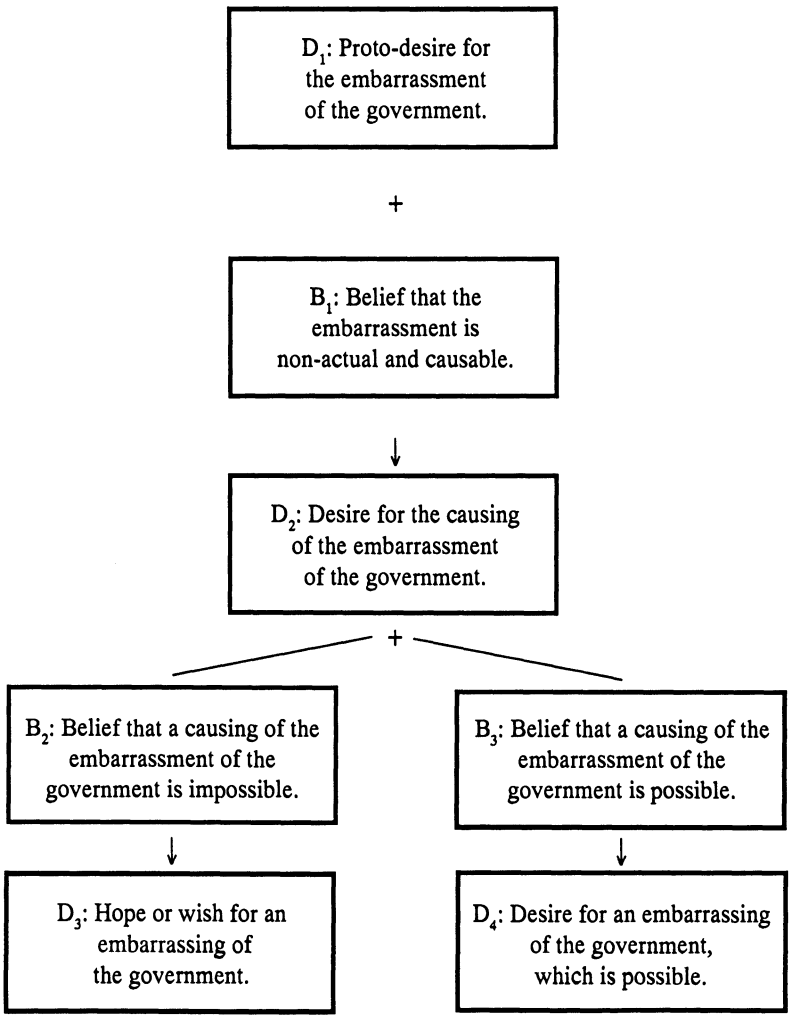
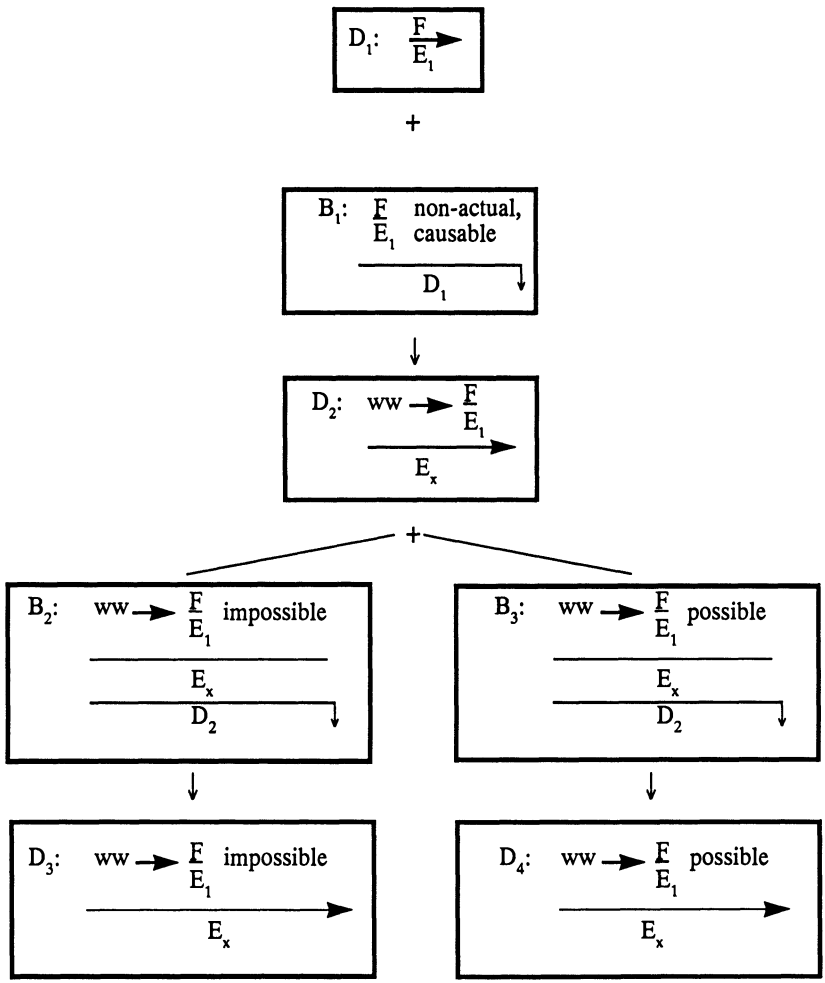


FIGURE 1b. PRE-PRACTICAL REASONING: FROM PROTO-DESIRE TO DESIRE-THAT

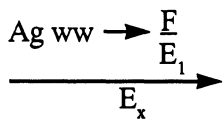


What follows is an alternative treatment of the transition from desire-that to desire-to.

The desire, D_4, for a possible embarrassing of the government raises the question for *PRAGMA* of whether or not there is an entry point for it into the world which would have causal relations to the embarrassment. If *PRAGMA* believes, B_4, that some event of its agency, such as the moving of its arm, could be a sufficient condition of a string of events leading to F/E_1, then it will be caused to move to the attitude of *desiring*, D_5, *to embarrass* the government. With this development a crucial change in the object of the attitude has occurred and the state of intention is being approached. This is the first point at which "agency causation" is directed at a member of that subclass of the class of events, namely an event of agency, Ag, which would have the required causal relations to F/E_1.

The object of the desire-to is an event which *PRAGMA* itself can (directly) cause, and that generic description, although in need of completion by practical reason, remains constant until action has occurred. The essential and steady question of practical rationality is: What event causable by the agent will satisfy the description of being the desirable sufficient condition of the goal? The desire to embarrass the government, D_5, is the desire for that agency-event with causal relations to F/E_1, and we represent it as follows:

D_5:



With the introduction of a desire-to, practical reason has its final direction and specification of the description of the agency event. The necessary further identification of the object of the attitude will proceed with means-end and evaluative reasoning as beliefs about possible causal entry points to strings of events *PRAGMA* believes would lead to F/E_1 are surveyed and evaluated.

The causal function, →, with which we describe the performance of both the desire-that and the desire-to, is, despite differences in their objects, the identical function. Both states are causal, *ceteris paribus*, with respect to the events they represent. If, however, *PRAGMA* believes that the event it desires is not causable by its agency, then, if rationality is to be preserved, *PRAGMA*'s desire-to must disengage from practical rationality as did *hopes* and *wishes*: *PRAGMA* must cease desiring-to, and with that, the processes of agency with respect to that goal will cease. Though *PRAGMA* may remain in a state of desire — it may continue to desire the government's embarrassment, for example — that desire will not proceed into the practical reasoning process. Only those states which are impressed by the belief that there is an event, E_x, which is an event of agency, Ag, and which has the properties to cause *PRAGMA*'s goal may proceed into practical reasoning. The set of necessary conditions for agency will also include the satisfactory identification of the event of agency and the evaluation of that event and certain of its properties. We describe these processes of practical rationality in the following sections.

2. PRACTICAL REASONING: FROM DESIRE-TO TO EFFICACIOUS INTENTION

There are three major features at work in practical rational agency. First is the function of causation; second, the identification or representation of the object of causation; and third, its evaluation. The second and third features contain the logic of the rationality of agency. They comprise the process of rationality and are themselves seen as the causal or functional roles of certain types of beliefs. Thus, the entire process of rational agency will consist of two functional roles — one for the accumulation of certain information within causal states (desire and intention), the other for the transfer of information (belief) — and a logic of the interaction of the two in virtue of which the process is rational.

To return to our account of practical rationality, given the desire, D_5, to embarrass the government, the rational next steps are for *PRAGMA* to "vet" the consequences of embarrassing the government. This is a complex process and may be plastic in its order and

structure. It might begin with how the embarrassment of the government would affect *PRAGMA*'s world. We may limit ourselves to *PRAGMA*'s point of view on the consequences as well as other relevant matters and make its scope as narrow or broad, as selfless or self-interested, as we like. Since our purpose is to describe the types of content and their relations which constitute the eventful history of a rational agent, and later that of an autonomous rational agent,[10] and to leave these distinguished from a moral agent, we are able at this stage to limit *PRAGMA*'s vetting considerations to whatever set of interests (or attitudes) would give rationality a function in the process.

The content which is concatenated by means of the belief function as the rational agent passes from state to state toward action is of two sorts. The first instruction is to identify for the agent the object of the desire and the second instruction is to evaluate that object. Identification and evaluation of the object of the intensional states which lead to action will proceed throughout the process until the agent acts. With action, the two main functions must come together. All the content concatenated through that process which defines rationality is added to the function which effects[11] the external world but only then is that external-facing function, the attitude embedded throughout in desire and intention, active: only with the confluence of content or information and active attitude is agency possible. That final state, or efficacious intention, is an amalgam of two functions: that supplied by the external efficacy or attitude of the state and that supplied by the content itself. Agency is that principle which can be informed by rationality, can have content affect it, and when so informed, and only so, lead to action. Later, we shall attempt to show that the relationship between the active attitude and the role which beliefs perform within practical reason is to be a *condition-remover* on the causal sufficiency of that active principle. If the agency attitudes of desiring, wishing, hoping, intending and unconditional causal sufficiency, or efficaciously intending, are states which are stages in the process toward action, and if the above conceptions of the functions are constitutive of rational agency, then the process can be cast in terms of these functions.

Suppose that as a result of the vetting of his desire-to, D_5, *PRAGMA* comes to believe, B_5, that his embarrassing of the government is desirable, all relevant things considered. If so, the desire-to, D_6, would be caused which would be in content similar to D_5 but with the significant addition that the agency-causing is now believed to be desirable.[12] The content of the belief that the embarrassing is desirable is thus added to the content of the desire, producing, with this first subjection to, and affirmation by, rational scrutiny of a state within the process, some *commitment*, namely, a state of *intention*. D_6 is the intention, I_1, to embarrass the government:

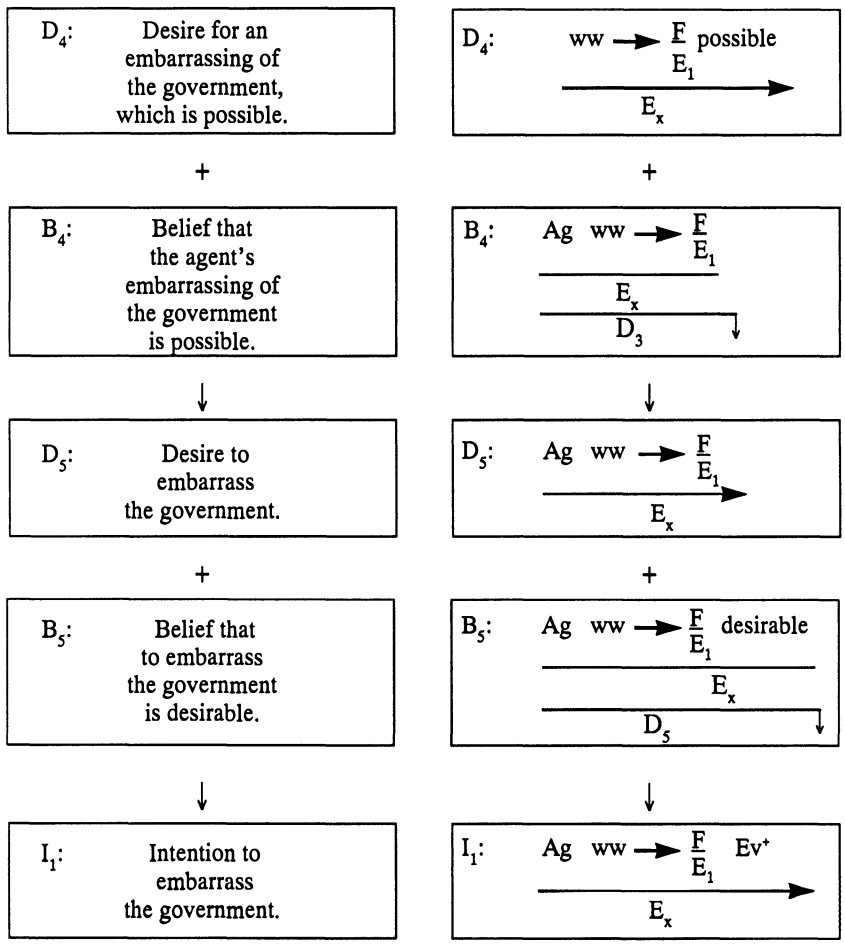
I_1:

$$\text{Ag ww} \xrightarrow[E_1]{F} \text{is desirable}$$

$$\xrightarrow[E_x]{\qquad\qquad}$$

Rationality and commitment are related here, as we shall see in some detail in the next section, through the fact that the rational procedure here just is to examine the desires believed to be related as consequence to a goal and to proceed to the next stage toward action only when the agent's desires so far identified are believed by him to be related to the enterprise so that they are sufficiently positive: this constitutes the first sign of commitment[13] and therefore intention.

FIGURE 2. PRACTICAL REASONING: FROM DESIRE-THAT TO INTENTION



With the occurrence of an intention, I_1, we enter the iterative phase of practical reason. I_1 takes us further into practical reason and through to planning and action and marks an important junction in the process toward action. It is what is distinguished as the attitude of intention: that point in the process of practical rationality at which the agent's causal state with respect to an event of his agency with causal relations to his goal passed a threshold worth marking. The belief in the desirability of what is represented in his active state may derive from a theory he holds about which desires are worth acting upon and could be as simple as the belief that this desire has a positive value relative to his other attitudes or values. Or its source could be as complex as an objective moral theory. Each of these normative beliefs could, as a matter of fact, move the agent into a state of intending; but some of these beliefs will be more rational, perhaps, than others. Heretofore, no evaluation had occurred, only the facts of the agent's attitudes and his beliefs about actuality and causability. It is the combination of a desire-to with the belief in its desirability which yields the state of intending. Now *PRAGMA* will cause the event of agency which he believes is sufficient for the embarrassment, his goal, so long as he believes that desirability is maintained. With the addition of those causal properties, *PRAGMA*, being what it is, will, *ceteris paribus*, embarrass the government. Before this point in the process, that could not be said if we believe that evaluation, or the exercise of normativity, is an essential part of rationality. Thus the importance of the belief in the desirability of the goal, creating, as it does, this difference in practical states.

It has been noted that the content of each state of *PRAGMA* concatenates the information of the states which effect or inform it in the process of practical reasoning. Thus, the final intentional state, that attitudinal state which is efficacious of the event intended, has as its content the *full practical history* of the determination of the properties of the event intended. That event is the one believed to have causal relations which are sufficient and necessary for the embarrassment of the government; it is the event with the causal relations to the release of the secret papers and is also the moving of *PRAGMA*'s own arm. With concatenation of the agent's description of

the event he intends to cause, questions about how certain parts of this complex of descriptions function are natural.

There are three things to be distinguished in an efficacious intentional state: the *means*, the *goal*, and the *event of agency* — the represented causal object of the intention. Ag/E_x is the *event of agency*. Suppose that G/E_2 is a means event[14] which therefore stands in a causal relation to both the goal, F/E_1, and the event of agency, Ag/E_x. Reference to that causal relation of Ag/E_x may serve to identify it as the represented causal object of the active attitude. For *means*, only the relationality of the event under description is essential and while some description is necessary to identify an event as one with the means relation, any description of that causal *relatum* may do for the identification of its functional role. In contrast, a goal-event is motivational only under its goal description.

As we proceed with *PRAGMA* the content of the intention to embarrass the government, I_1, becomes modified by beliefs about which causal strings of events containing F/E_1, the goal-event, have entry points for the agent and which entry-points for agency would therefore be appropriate. Suppose, as does our example, that *PRAGMA* believes, B_6, that the release of the secret papers is the best way to achieve the embarrassment of the government. That belief (which of course presupposes a desire-vetting loop) causes its content to concatenate with the content of I_1. The resultant state, I_2, is describable as the intention to cause that agency event which will result in the embarrassment of the government through being the cause of the release of the secret papers.

As a momentary aside, a modest test of the theory would be to see how helpful it is in dealing with such seemingly innocent but refractory natural language expressions as 'You did it'. What is referred to by 'it' in 'You did it'? Is it the goal-event, the F/E_1, or the action, the F-ing? If the 'it' refers to the goal-event and if 'You did it' is therefore a relational sentence, then it is ascribing a doing to an agent where that doing is being identified by its relation, which is causal, to the goal event, which is F/E_1. It is saying that your doing, your event of agency, was an F-ing: it is identifying the action by

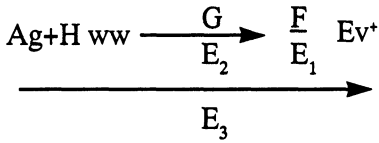
means of its relation to F/E_1. In the second alternative the 'it' refers to the action itself in which case you did a doing. The resultant awkwardness of the second version moves one in favour of the first.[15]

B_4 was the belief that there was at least one possible agency "ww→" chain to F/E_1, the embarrassment of the government. The agency event, Ag/E_x, may still be either an event type or event token in the agent's beliefs. But at this early stage in the identification of the event which is to be the embarrassing of the government, it is unlikely that it will be identified beyond its type. In any case, a plurality of agency chains believed to culminate in F/E_1 would entail a plurality of particular events of agency to be considered.

I_2 is the result of the incorporation of the content of B_6 — the belief that the release of the secret papers is the best available means to the government's embarrassment — within the content of the intention I_1. The causal chains believed to be available may not come into play for the agent until he has faced planning considerations. That is, the process toward action may cut into the planning considerations[16] at I_1. Once there is intention, or some commitment to action, the question of which particular causal strand to employ may be deferred pending the larger and more complex question of how best to arrange the pathways to serve one's various commitments.

One particular pathway to F/E_1, and therefore a particular event of agency, will finally, however, have to be fixed by the agent in order for intentional action to occur. The event which is believed to have the properties which will initiate the 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_x, which is this movement of his finger, H/E_3, which causes the release of the secret papers, G/E_2, which causes the embarrassment of the government, F/E_1. And thus the final and efficacious intentional state, the one which will cause the event of agency which has been identified and evaluated as the event E_3 with the causal relations to F/E_1, has the following form:

I_3:

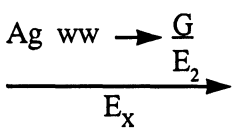


This is the form of a possible state of efficacious intention. More is said of it in the immediate sequel.

3. THE PRACTICAL SYLLOGISM: DESIRE

The nature of the practical syllogism with a desire premise is an essential stage in practical reason and deserves further attention.[17] The active and cognitive attitudinal states we attribute to agents are important in our appraisals of them and their actions and since there is some controversy over which states practical reasoning yields as conclusory, this topic is worth addressing.

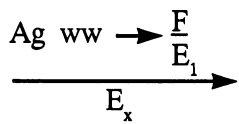
Suppose then that the agent desires the embarrassment of the government, F/E_1, and believes that the release of the secret papers, G/E_2, is the best means to that end. What practical or attitudinal conclusory state should a rational agent be in as a result of these two premise states? Does a rational agent find himself, for instance, in the resultant state of desiring to G/E_2? We must, as we know, not take this putative conclusory state to be the liking of G/E_2, i.e., $-G/E_2\rightarrow$; that would be the wrong object of practical reasoning. We must take it as the desire to G/E_2: the desire for the event of agency which would cause G/E_2:



This desire-to state is supposed by some to be a rational result of the above premises. To assess this claim about the process of practical reason we ought to recapitulate the nature of desire. We think of a desire state 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, something

the agent can himself cause, such as the movement of his arm. What the agent does desire when desiring to embarrass the government is the event, E_x, he represents as Ag with causal relations, (perhaps) through other events, to the event which is F, F/E_1. That, and not F/E_1 itself, is the object of his attitude. To "desire" F/E_1 directly and without causal relations to an event of his agency is irrational so long as causation is necessary for actualization.

Now this desire does not change just because the agent has the belief that G/E_2 is a necessary event or means between Ag/E_x and F/E_1. If we need a principle to inform us on the identity of desires it would be that the identity of the event of agency being represented as the event with causal relations to F/E_1 determines the identity of the desire. The identity of the desire then does not change as the agent continues to articulate the nature of those relations which he must believe Ag/E_x has in order to become the event he will cause. The causal attitude the agent has to the Ag/E_x he will choose is explained by the causal relations Ag/E_x is believed to have in relation to F/E_1. Recall that the causal attitude to F/E_1 was originally transferred to E_x, some event which would cause F/E_1, because of the agent's belief in the necessity of causation for F/E_1, given its non-actuality. The causal attitude was subsequently connected to an Ag/E_x when the agent believed there was an event of his agency possible with causal relations to F/E_1. So those causal relations are essential to the attitude of desire remaining tied to that event of agency which would cause F/E_1:



That shows us the line of causation, and therefore reasoning, from the motive of liking F/E_1 to desiring to F/E_1 and explains why Ag/E_x is now the event which has inherited the attitude of desire. It is the perceived causal relations of Ag/E_x, the event of agency, to F/E_1 which causes the representation of the event of agency to become the subject of practical reason. The relationships which exist in a desire between the representation of the goal, the representation of the event

of agency, and the force of agency provide the structure for a significant part of the rationality of agency. No rational inference, then, from a desire premise to a desire conclusion, no effect of such a state to another desire state, whatever the additional premises, will be allowed which is not a transference of the attitude to an object which is the event of agency believed to have the causal relations to F/E_1, the goal. Unless this representation of Ag/E_x as the event with causal relations to F/E_1 controls the representational content of desires-to and intentions and therefore ultimately what the agent will cause, practical reason and rational agency are not possible.

What, then, about the practical inference from the desire to F/E_1 and the belief that G/E_2 is the best means thereto, to the conclusion: the desire to G/E_2? We can see at once that this conclusion does not preserve the relationships to F/E_1. If my conclusory state is the desire to G/E_2:

$$\text{Ag ww} \xrightarrow[E_X]{} \frac{G}{E_2}$$

then that may or may not get me F/E_1 since those events which are G's may not all be sufficient for F/E_1 even though some are. It would be irrational to be causal with respect to that event, Ag/E_x, which would cause G/E_2 without the proviso that it be a G/E_2 which would cause F/E_1. We don't then rationally come to desire the means because we desire the end! We don't desire the means, G/E_2, at all as conclusion. We desire G/E_2 *as means*; but that is only to make reference in a new guise[18] to the identical causal relations which were referred to by the description of Ag/E_x as the event with causal relations to F/E_1. What followed from the above premises was wrongly thought to be a new conclusory state of desire with a new object. Actually the conclusory desire state just above only articulates further the causal properties which the event of agency must have in order to facilitate the identification of the event to be caused. We don't, as conclusion to the premises, desire new properties of the same event, Ag/E_x; rather, we desire the same properties, although newly described, of the same event. Only in that way is rationality

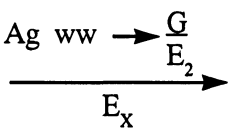
and therefore motivation preserved. And, *vice versa*. It is of course possible to independently desire the means not as means. It is also possible to come to desire the means not as means just as the result of their being in the premised states. But that would be a distraction and not an effect which conformed to rationality.

We must ensure that such distinctions are maintained if we are to have available in action discourse an accurate map of the causal psychology of agents.

4. THE PRACTICAL SYLLOGISM: INTENTION

If I am rational, what follows from my intention to F/E_1 and my belief that G/E_2 is the best means to F/E_1? Do I intend to G/E_2? If I intend to G/E_2, it follows on our view that I intentionally G/E_2 if I successfully act on this intention. So something of importance is at stake in how we describe the agent's states. Action discourse, which is directly instructed by our beliefs about the relevant states of agency, will, once a normative theory is put in place, have consequences for assessment and control of agents. So, do I, as conclusion to the above, intend to G/E_2? Do I intend the means?[19]

We know that an intention to G/E_2 is an intention to cause an event of agency with the properties to cause G/E_2.



We also know that this is not yet a complete representation of an intention: until positive evaluation of the event under this description has taken place, the agent is only in a state of desiring to G/E_2. The state of intending is, we recall, the state of desiring to do what is believed by the agent to be desirable. Is, then, the intention to G/E_2 a rational conclusion of our premises? It seems not since there is in the premises no element of evaluation of G/E_2. In order to rationally intend to F/E_1, it is necessary to have positively evaluated the event of agency, Ag/E_x, under the description "which would cause F/E_1", or to

value the agency event for its causal relations to the goal: it is to have positively evaluated that goal-functionality of the event of agency. Since evaluation is tied to intensional contexts, Ag/E_x is, in this case, valued essentially or non-substitutionally under that description, or for the having of those relations to F/E_1 which is what that description describes. The causal relations from Ag/E_x to F/E_1 which *PRAGMA* values may be equally identified or represented as the causal relations which G/E_2, the release of the secret papers, has if it has both effect relations from Ag/E_x and cause relations to F/E_1: if, that is, G/E_2 is means from Ag/E_x to F/E_1. The causal relations which Ag/E_x has to F/E_1 are, then, identical to those which G/E_2 has when G/E_2 is means for Ag/E_x to F/E_1 since G/E_2 is means for Ag/E_x to F/E_1 if and only if it is that effect of Ag/E_x which causes F/E_1. So the causal relations we intend and value are referable to as those which the means to F/E_1 from Ag/E_x has. The causal relations we intend are equally the causal relations of the release of the secret papers. And those relations are also identical to those which my embarrassing the government has. Since the embarrassment of the government is the desired effect along the same line of causal relations from Ag/E_x to F/E_1, its position on that line of relations 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, namely: an event as event of agency, Ag/E_x; an event as means, G/E_2; and an event as goal, F/E_1. The concepts of *action*, *means* and *goal* name these positions of *relata* in the line of causation.

Let us return to the syllogistic form in order to see this identity of reference at work there. The agent intends to embarrass the government. He believes that the release of the secret papers is means. With this 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, those to F/E_1. The means premise allows the creation of a co-referential description of that same set of causal relations. They are now describable as the causal relations

which the event of agency must have to the release of the secret papers since the release of the secret papers is an event also describable as the cause of the government's embarrassment. So the same span of causal relations is encompassed in the conclusion as was in the reference of the first premise. The practical reasoner is thus entitled to conclude that the intention to embarrass the government is the intention to release the secret papers, and, with the further requisite means belief, to further conclude that the intention to depress a certain computer key is the intention to move his finger so. The requisite means beliefs supply descriptions of events which are *relata* within an identical causal chain from Ag/E_x to F/E_1, from the movement of his finger to the embarrassment of the government, 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 beliefs 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 desired. These new representations which practical reason concatenates allow the agent to sufficiently identify the particular functionality of the event which his agency will cause.

The agent will of course be aware that the event of his agency will not only have the desired and, *pro tempore*, desirable causal relations to F/E_1 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 Ag/E_x to F/E_1, but also those which take us through an event which is a release of secret papers. So his event of agency has this property as well as its causal properties. Just as it was a requirement on a certain level of rationality that the agent evaluate what he desires, it is a 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 F/E_1. 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

identifies the event which he will cause through additional representations of that event.

Despite these evaluations prompted by the additional representations of the causal relations to F/E_1 desired and intended by the agent, he continues to desire and intend only those relations, i.e., only the goal-functionality of 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 doing of F/E_1 as having. But it is the doing of F/E_1 which has engaged his active attitude, not any of its other properties. 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.

4.1 Minimal Means

In *active* practical states of agency there exists actual and/or potential causation as well as a representation of what that state will cause: its intentional object. The represented or intended causation is of a pathway whose function is to link what the agency state will actually and directly cause with *PRAGMA*'s goal. What *PRAGMA* represents as means is just that functional (causal) pathway. Any points on that pathway by which he 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 functionality 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 reason is: (1) the event of agency — what *PRAGMA* will directly cause, (2) the particular functionality of what *PRAGMA* will directly cause in order to link him to, (3) his goal.

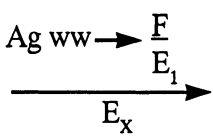
5. INTENDING

The nature of intending, since it is, we assume, a state in a process, requires for its explanation placement within that process. We look again then briefly at the process within which intending must fit as a prelude to a closer look at its nature.

The generic features of agency which precede and culminate in action, we have argued, consist of the following. There are two generic causal functions: one is the attitude of belief, the other the attitude of desire and intention. Together these functions take content to and from the world, and thus they are the functions which define our relations with the world. The belief function takes on content from the world via perception and gives off content to the active function which, when an efficacious intentional state, has the capacity to cause what it represents in the world. The function of these cognitive-*cum*-causal states is not explicable without the notion of representation. This, we claim, as the reader will by now be aware, is not a metaphysical plea; only a structural remark: the nature of the metaphysics surrounding and comprising the two functions of *cognitive efficacy* and their content is not urged. The content of the *active*, as opposed to the *epistemic*, states of agency has a structure which is to represent the event the state will cause in the world. Since the only events *PRAGMA* (directly) causes are bodily events of his, the function of the content of this state of his agency is to enable him to sufficiently represent the event of his body that would have the relations and properties he intends. This describes part of the process or logic which *PRAGMA* must be capable of if he is to be rational. One of the functions of belief is to provide sufficient identification of the object which will constitute his action. The other use of belief is to provide evaluation of the object of agency as its identification proceeds by means of concatenation of those properties and relations believed sufficient for the object to be the object of agency. So identification and evaluation constitute the rational aspect of rational agency.

We argued further that it was attractive to postulate a fundamental rational move which structured the content of an active state of agency 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) called '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 would allow identification of the cause of the motivating event. We held, then, that an unrationalized 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 causation and therefore no representation of an event which is linkable to agency. For rational systems, the belief in causation formulates the basic identification of the event which it is rational for agency to address, as *the event which would cause the goal*:



This is a state of desiring that F/E_1 come to be. The object of the attitude now has a rational basis in causation. A desire is a rational state just because it is in a form which is the result of a previous state of cognitive agency and a belief which has commenced the identification of the object of agency. The previous state, $-F/E_1\rightarrow$, was uninformed by belief and would result in behaviour which was non-rational. The desire that F/E_1 come to be, although perhaps insufficiently identifying its object, E_x, as the event which would cause F/E_1, is not non-rational but as yet perhaps not fully practical. It would result in teleological behaviour unlikely of success but not non-rational behaviour.

The process of rational agency described thus far has made use of the causal functions of belief and desire and representational content. But evaluative beliefs have not yet come into the process. We assume that the evaluation by the agent of any of his states with a causal function is rational. So evaluation in a rational system 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 one's causal capacity and those events to

which that capacity is limited: only events of one's body may stand in the effect relation to an agent's internal states of agency. A system which lacked beliefs about, and therefore representations of, the limitations on the causal relations to which its agency could stand would not be a system capable of (rational) action. Its behaviour might be like that of an infant where there existed no beliefs about which causal entry points into the world were available to it and which of those would lead to its motivating events. Where those beliefs do exist and an agent is convinced that an event he can directly cause, an "entry point", does exist which has the causal properties required to satisfy his motive, he is caused then to be in the state of desiring-to. Now he is in a practical state. He is into that part of the process which will identify his act. Heretofore his attitude had been first directed toward a motivating event F/E_1 *simpliciter*, and then toward whatever might cause the motivating event, *viz.*, E_x. Now, however, with the desire-to, his agency is directed at an event of his agency, Ag/E_x.

We accept that an agent intends only such objects. He cannot coherently intend the object of his desire-that F/E_1. Intentions have as their objects the same type of object as do desires-to, whereas to desire-that is to be in an attitude toward a "larger" object. All that separates a desire-to from an intention is the positive evaluation of the desire-to. Where the system 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 the system is in a state of intending to F/E_1. The first positive evaluation of an action has occurred and with it a significant increase in rationality has occurred. We shall urge that this state has all the properties of intending.

It has been claimed[20] that intending has the following properties. We describe each of them briefly in terms which our view of rational agency supplies in order to show that our conception of intending has all the properties arguably required of it.

5.1 Relations with Desire and Belief

For us, both desires and intentions are cognitively causal states whose

presence we are able to report upon. Both may have their content determined by beliefs and both have a structure to their content which "asks" for the identification of the event of agency which would cause the goal event and to which event of agency the efficacy of the state is directed. The similarity of the structures of desire and intention shows us that they are close parts of a common enterprise which is the rational process toward action. That the difference between them is constituted by the belief in the justifiability of the enterprise in which they are involved shows that their difference is one of added rationality.

5.2 Commitment

We claim that 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 rational process. Since the difference between a desire-to and an intention is the belief in the justifiability 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 positive evaluation, or it is just identifiable with that increased rationale for the doing. In either case our view 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 rational agency as 'intention' marks. It should be noted that scalar increase and decrease in belief in the justifiability of an act match the scalar increase and decrease in commitment to a doing of which an intention is capable.

5.3 Purity

This is the property which denies the possibility of the reduction of intentions to desire-belief complexes. On our view, again, an intention results from a desire to A and the belief in the justification of doing A. Intentions are results from such antecedents in rational practical systems and not identities of such antecedents. Moreover, on our view intentions would be states which one might find oneself "parachuted" into rather than caused in the usual cognitive way: they

might just occur in us as the result of non-cognitive rather than cognitive antecedents or as the result of some process other than that of practical rational agency. The possibility of such "parachuting in", or of some other process of formation of states of intention, emphasizes their purity since in neither case is an antecedent desire state even a necessary condition for intention. Actually, even in practical rationality, desires are not necessary antecedents to intention but rather only possible antecedents. The plasticity of rationality would allow evaluation of the causation of one's motivating event before desiring its causation and the subsequent direct move to an intention without an antecedent desire.

5.4 Relations with Planning and Other Intentions

We have argued that intentions occur at a crux in practical reasoning. At that crux, and certainly not sooner, planning becomes rational.[21] Since the first justification 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 and justifiability in relation to other intentions as well as to other planning considerations such as space and time might require.

5.5 Sufficiency for Action

For any of *PRAGMA*'s intentions the questions arise of when it becomes effective in the world beyond the mind and what properties such a state must have in order to be properly sufficient for action.[22] This is for us the same question as that of when the processes of identification and evaluation are sufficient for action. But if *PRAGMA* is rational and an agent there must come a point where, for at least some intention, he need not/will not ask of his beliefs how to implement that intention or for its further evaluation in the light of an added representation. To have an active attitude which is rationalized to a certain point *is* to be prompted to act if rational agency is possible. For an intention to be effective in the world it is not always necessary to call up a belief with respect to means. In order to release the secret papers he must, if he is a rational agent, finally just become capable of affecting the world. Where he is finally so capable is

precisely where the question of means or of further identification and evaluation does not enter. Thus if *PRAGMA* is 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 release the secret papers he believes, say, that he must move the papers in his hand toward the reporter, or move his fingers thus on the computer's keyboard. But in order to move his hand 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. 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,[23] 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.[24]

The above addresses the nature of the very idea of the sufficiency of an intention for rational action. The factual question of when an intention attains sufficiency is the question of when the practical rational process actually attains a state of sufficiency. 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 achieves a state of sufficiency may be determined by the occurrence of the belief that rationality, or identification and evaluation, has been satisfied. And *when* that belief occurs will vary from case to case and agent to agent. When that belief *should* occur, when the agent should act, and when not, are further questions which require for their answers a theory of rationality the purpose of which is to provide the conditions of justification for a belief that an intention is rational. 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.

Desires-to and intentions, as well as all the other active attitudes, share the property of causal functionality with respect to the body.[25] The various active attitudes, such as proto-desires, wishes, hopes,

desires-that and -to, and intentions, are distinguishable by their content and by their functional position in the process. A desire to *A* together with the belief in the desirability of *A*-ing causes, in a rational system, the state of agency called 'intending to *A*'. Intending differs from previous other agency states in that it is the first state in which the normative condition on acting has been addressed. We may regard all the practical reasoning in rational systems as the satisfaction or dissatisfaction of a set of *ceteris paribus* conditions which are deemed "rational" and have been described earlier. Practical reasoning, which consists of the identification and evaluation of the act intended, functions in practical rational systems as the set of *ceteris paribus* conditions on the sufficiency for action, or the efficacy of the intentional states. Thus, in practical rational systems intentional states would be causal except for their control by reason. As the process of identification and positive evaluation unfolds, cognitive content is added to the active states, marking the increasing rationality, which is to say 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.[26] Until practical reason is satisfied what exists is only causal sufficiency *ceteris paribus*.

This suggests that the attitude itself, this power of agency, its causal functionality, is present from the beginning of each complete segment of rational agency, i.e., a segment which is explanatory of action. The remainder of the process is a series of identifications and evaluations — additions of information — which constitute rationality and operate as a series of *ceteris paribus* conditions upon the efficacy of the active component of agency. When these *ceteris paribus* conditions are removed, which is to say that the state of efficacy has sufficient informational content, causal sufficiency exists in a rational system.

It is possible, however, for the active states of agency to become effective at any stage of practical reason without the benefit of full

rationality. That such impulsivity 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, capable of transgression against rationality. Breaches of rationality by impulsivity are commonly called 'compulsions'. And compulsions sit on a continuum which has depression at its opposite pole with *akrasia* in between. Whereas compulsion is the paradigm case of the failure of practical rationality 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.

A word on *akrasia* is apposite here. *Akrasia*, the converse of recklessness, is the failure of an intention to cause its intentional object despite the presence therein of sufficient informational content for practical rational efficacy. 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 for sufficiency of action. It attacks any such offered conditions with the argument that their satisfaction is compatible with insufficiency for action. Defenders of a sufficiency thesis, thus attacked, counter with the charge that to deny action under their conditions is incoherent. But if action is a viable concept then there is entitlement to the assumption that there are *bona fide* sufficiency conditions for rational action. We are entitled to believe that some rational states must be sufficient for some actions. We can, in these states, separate their causal sufficiency or their active causal function 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 rationality which consists of the identification of the event to be caused and its evaluation is the set of *ceteris paribus* conditions which such active attitudes must satisfy before their rational release into activity. Thus when states sufficient, *ceteris*

paribus, for action have accreted all the content which practical rationality adds to them as they pass through the process of identification and evaluation, they achieve at some point sufficiency *simpliciter* for action. If we accept this picture of sufficiency's presence at the beginning of practical reason, and of practical reason being the control of rationality conditions over sufficiency, then we have the following options of how to see the problem of *akrasia*. Where the agent reports (1) that he intends to *A*, and (2) that he has removed all its *ceteris paribus* conditions — has satisfied practical reason — yet has not so acted, then either:

(a) Report (2) was false and he has not satisfied practical reason; or (2) was true and report (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 held to be in a state of rational 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 sufficiency of his active state itself, or to its causal "override" by some other active state, of which the agent may be unaware.

5.6 Relations with Intentional Action and Responsibility

Intending, and particularly the intention which has satisfied all the *ceteris paribus* conditions of a rational system and is thereby sufficient *simpliciter*, stands in the causal relation to an event. So-called *future-directed intentions*[27] stand in the causal relation to events *ceteris paribus*, or subject to the further conditions of practical reason. But, we claim, that an agent's intentional state has caused an external event is only a necessary condition for the agent to be relatable to the object called an action. An agent is relatable to an action through their explanatory relationship. An intentional action is 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. That object makes essential reference to its *explanandum* and so fits it for its role 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 realization of the concept of agency requires, as we may now expect, the working presence of a state which is both cognitive and efficacious, one which has cognitive efficacy, at that very point where agency and its object, the action, may be said to enter into their causal relation. That presence is recognized by most writers and has borne the names of "present-directed intention" (Bratman), "immediate intention" (Brand), "intention in action" (Searle), and "all-out judgement" (Davidson), and it is that state whose presence both marks agency or the causal sufficiency of persons and explains the emergence of those objects called actions which we attribute to agency. 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" which has not only the aspect of cognitive content but that of the causal sufficiency of certain cognitive states present antecedently and throughout the process of practical reasoning. Views which see causation emerging suddenly and not until the point of 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 process of practical 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, unless one holds the unlikely emergent view of their relationship, the 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 that 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 sufficiency.

Thus, while it is true that any theory of action must account for why causation occurs at the point of action and not earlier or later in a rational system, 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 efficacy emerge. A superior theoretical option, we claim, is that cognitive efficacy or sufficiency 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 the gates in the process so that cognitive efficacy already present therein, and reportable upon, may emerge, or not, depending upon its rationality.

While beliefs, as they function in practical reason, 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 themselves? This is really a question of whether we need two separate causal functions for action theory or only one.[28] We openly assume that two distinct functions must be made out however apparently unitarily they are packaged. But the point to hold on to in this matter is that it is the belief function with its injection or withholding of information which

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

This is what is meant, then, by 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 is otiose.

Another feature of smoothness is the concatenation which occurs in the cognitive content of *PRAGMA*'s desire or efficacy. The agent, in our example, begins with the cognitive content of his proto-desire for the embarrassment of the government. As the desire, or, as we may alternatively speak, as the state which is causally sufficient subject to practical reason, proceeds through the syntax of practical reason, it gathers content. When a means to the goal of embarrassing the government is fixed, the intention to embarrass the government becomes, in our example, the intention to embarrass the government by means of the release of the secret papers. And when that intention is finally efficacious it has become the intention to embarrass the government by means of the release of the secret papers ... by means of moving one's arm thus. When the intention is efficacious, when it is sufficient for the initial event in the contemplated chain of events, when *PRAGMA* is *en acte*, the intention has a concatenated content provided by practical reason's injections of information about the believed best causal route to the goal. When *PRAGMA* is in this state he will be in a state of sufficiency for action, if he is rational. Agents sometimes come to this state non-ideally, or not having followed the syntax of the practical rational process. Sufficiency for action will be present but the intentional state may not be justifiable. It is worth repeating that in all of this, a theory of what is rational is needed. Such a theory has some stable generic points as the roles of

identification and evaluation in rationality exhibit. But just where these operations are justifiably believed satisfied, and rationality therefore satisfied, and sufficiency justified is not a settled matter. Wherever the theory of rationality lights on these matters, however, rational sufficiency is producible. Reason will obey.

There is one further aspect of the smoothness hypothesis worth remarking upon. It may be that there is really only one active attitude throughout the practical rational process. That attitude would be a causal state, or, put computationally, a state which has a function of addition — a function of moving to the next state provided by the syntax of the process. Thus, desiring and intending would be generically the same in their function: move to the next state in the process. Their difference as attitude 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.

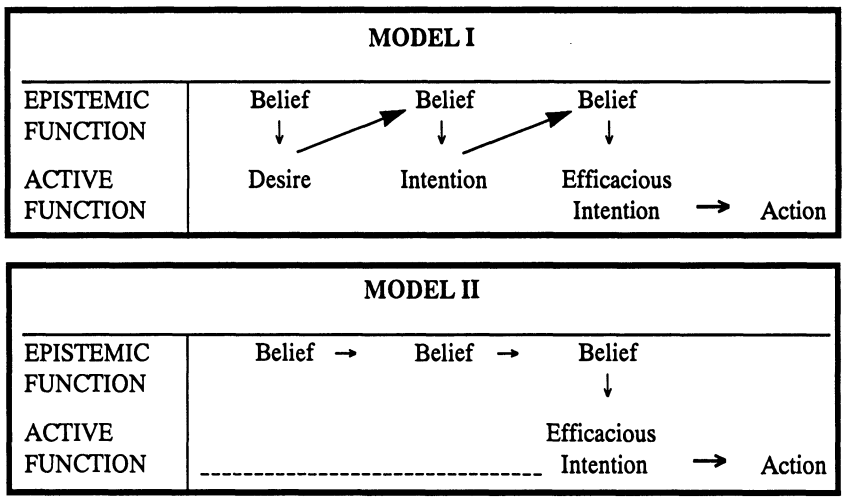
7. MODEL II

So far our theory of the process antecedent to rational action finds states such as desires and intentions — those states with the *active* causal function — included along with those states with the *epistemic* causal function as stages in that process. The process is taken to consist of practical syllogisms which move matters along from conclusory desire and intention states through planning to action. The import of this model is that the agent is to be found in such conclusory active states at points throughout the process, where these active states have thus picked up the content developed to that point in the epistemic function of the process. The motto of Model I on this aspect could be: "Co-functionality throughout the process until efficacy." An alternative, Model II, would envisage the two functions remaining separated until a final active or efficacious intention is created through their coalescence. Its motto would be: "Only the belief function until co-function is rational." All operations on content, all practical results of rationality, would, on this second

model, be restricted to the belief level until the efficacious intention is created through the passing of that concatenated information to the active causal function which would have remained "content-less" to this point. The beliefs which would be sufficient for the creation of an efficacious intention would include the following: that the process of adequately identifying the event, Ag/E_x, which is the event with the properties of being the first cause in a chain of events which would include the goal, is complete; that this event is the best event to cause, all things considered; and that it is an event caused by agency. When such an "all-out judgement," to use Davidson's phrase,[29] is the case for a rational agent, then the information which identifies and evaluates the event which is the subject of that judgement is passed to the active causal function. A theory of rationality would have to address when the identificatory specifications and their evaluations in the concatenated content of the agent's beliefs are sufficient to cause co-functionality or efficacy in a rational system.

Models I and II, and there are surely others, differ only in the nature of the stages of their processes. The reasoning, the structure, and the concatenation of content are the same. Even their implications for autonomy are somewhat the same, although Model I lends itself, as Model II does not, to a limitation upon autonomy to be described shortly. Model II, we claim, liberates us in the views it allows of the relations between the rationality and causal sufficiency of agents. Once the liberal nature of their relation is seen in Model II it may be transferred to Model I. Their stages are different in that Model II insulates the two causal functions until their co-functionality at the point penultimate to action, while Model I has co-functionality occurring throughout. Models I and II may be represented as follows:

FIGURE 3. MODELS I AND II



There is a problem with Model II but it appears to have a solution. There are, however, at least two advantages to Model II. The first is its simpler engineering and the second is what it suggests about autonomy.

The problem for Model II is that we apparently report on desires and intentions throughout the practical mental process which precedes action. These reports, apparently about active states with content derived in the process, deny the principle of insulation which distinguishes Model II. The phenomenology of desire and intention, one might say, is on the side of Model I. The question is, however, whether such reports actually are about the existence of stages of co-function. Proponents of Model II would have to argue that to report a desire or intention is to report on a belief which would, *ceteris paribus*, achieve the co-functionality of a desire or intention. In Model II, desire and intention reports report upon certain points in the process of intention formation which, if the remainder of the process yields the required content, will produce co-functionality and action. Model II views desires as beliefs[30] with, given the process they inhabit, a propensity toward action through co-function. And since that propensity seems to be the essential point about desires, Model II

may not be required to deny mid-process reports on desires and intentions. But they would not be reports on occurrent co-functionality. They would be reports on occurrent epistemic functionality and therefore dispositional co-functionality at the point of action.

There appear, on the other hand, to be two clear advantages to Model II. To insulate the active attitude, or function, from content until that content is fully rationalized is a simpler engineering of a system fully equipped for rational agency. The point of the separation of the evaluation and identification functions, which are belief functions, from the active function is that the desirable goal of having only *fully* rationalized actions is more readily achievable. In Model I, where co-functionality occurs throughout the process antecedent to action, beliefs themselves could operate as yes/no gates to those co-functional states which are active. The question would be whether that epistemic function alone is adequate to the task of controlling the active function of desires and intentions. Model I could, alternatively, have a structure for the handling of desires and intentions to which beliefs merely control the entrances. There may then be only the advantage of simplicity for Model II over Model I. While Model II needs an engineering or logic for beliefs which will constitute rationality and a specification of the nature of the content a belief must have in order to cause the fusion of the two functions which precedes rational action, Model I needs that engineering as well as one which will control the progress of active desires through the process. There will have to be a logic of desires in addition to that of belief and a specification of the state at which the coalescence of the two functions is sufficient for action.

Apart from simplicity, the main advantage of Model II may be in what it suggests for autonomy through the isolation of the two causal functions of rational agency until their co-functionality in a state of causal sufficiency for action.

8. AUTONOMY

Our Model I of practical reasoning is compatible with a well-known view on the nature of autonomy. According to that view, an agent's

potential for autonomy is limited by the fact that practical reasoning which yields an active attitude with action potential as conclusion, requires in its premises an active attitude from which the conclusion is derived. Thus, desires and intentions are not themselves the product of belief states, nor of any other rational or cognitive state, alone. The content given with the antecedent or premise active attitudes marks the limit, on this view, of the exercise of rationality over action. Belief may have a large role in the determination of the content of an efficacious intention, but there must always be such a premise state with content given if an active attitude is to emerge from a rational practical system: no act will be explicable without some active attitude with its content presupposed functioning at the commencement of the process to be explained. Our Model I has an active attitude, the proto-desire $-F/E_1\rightarrow$, as the initiating state of the rational processes of agency,[31] and a further rationalized active attitude as conclusion and premise state of every subsequent rational process.

It is open to such a view of practical reasoning and autonomy to hold the content of any given active attitude subject to the agent's beliefs about what is desirable — beliefs, that is, about what the agent's desire content should be. But with the apparent explicability requirement of Model I, another desire must be invoked to function as premise to the conclusion which amends any given desire. Thus, no matter how morally or rationally sophisticated that desire may be which is called for by the explicability requirement for action in Model I, it will be itself inevitably unopen to amendment through beliefs about desires alone. Suppose that Model I does not covertly and inconsistently assume that the content of desire is amendable by beliefs alone. Then any amendment of desire which is possible is entirely out of the hands of reason alone, dependent, as it is, upon the sheer (physical) occurrence and givenness of the desire required for the amendment.[32]

How far back we are able to take the process of the justification of active attitudinal states, or how far back we can drive the premise states before we reach a desire state for which we cannot find a rationalizing attitudinal antecedent, draws the limit of autonomy on

this view. We might find that the limit occurred with that state where we found ourselves with the desire to be rational: we cannot produce an antecedent desire state which together with a belief would produce the desire to be rational. We simply find ourselves with the desire to be rational; it is a part of our nature to be so motivated and its presence constitutes the premise of our rationality. There would be, on the view being described, no antecedent causal cognitivity to the desire to be rational. Our autonomy would stop at this point since, whatever may be meant by that notion, its governance does not exceed the bounds of the cognitive, and the essence of the view we are examining is that some active cognitive state with some content, however general, however desirable, however justificatory or capable of supporting subsequent desires or intentions, must be of necessity beyond the reach of belief and therefore beyond autonomy, or practical reason cannot be explained. Since the exercise of autonomy is identical with the exercise of practical reason and practical reason presupposes intentional premises, so, therefore, must autonomy presuppose and be limited by at least one such premise.

Our view makes it possible to take autonomy a significant step further. We hold that practical reason, while it must assume an antecedent active causal function such as that contained in an intention or desire in order to explain the occurrence of a rational conclusory active attitudinal state, need not assume any content to that active causal function. There is, we claim, no incoherence to the "un-contented" state of $\rightarrow$. Without such a state of functionality one admittedly cannot produce syllogisms which, on the basis of beliefs alone, conclude with such states. But such premised states need not have content which is inextricably given with them. Their content may come entirely from belief. This would mean that all content was determinable by belief or non-heteronomous reasoning rather than some content inevitably heteronomously supplied and, therefore, systematically beyond reason. Practical reason on our view assumes the possibility of active attitudinal states nude of content, $\rightarrow$.[33] Reason would, through belief formation processes, determine the content of such states. The content might be to be rational, and a desire or intention to be rational would result which could now

function as a premise in a practical syllogism in the ordinary way. What we have just described is the exercise of autonomy in an area which the standard view[34] claims is unavailable to rational agents.

The increase in autonomy is significant in that on our view all content of the active attitudes, including the most gnomic, is rationalizable whereas it was not on the standard view. The unrationalized content systematically present in the standard view, though perhaps small in amount, is reciprocally large in its justificatory import and allows therefore our highest values in these matters, rationality for example, to be beyond choice and reason and present only as a structural given. Moreover, it is, on the standard view, not clear that there exists content which is so gnomically jusificatory that practical reason would need only one such heteronomous state. It needs at least one heteronomous state unless explanation of the rationality of all our choices runs back to a single desire premise.

Of course, autonomy itself, even on our view, is a captive of beliefs. But belief formation, or the "contentifying" of the belief function, is the essence of rationality and to identify autonomy with that, as Kant did, is surely to identify it with the most self-determining aspect of our nature. If autonomy is a value it is in connection with the virtues of belief formation or reason that this value must be explained. Our argument, then, is really a defence of Kant in that it denies that practical reason and therefore autonomy is necessarily an exercise in at least some heteronomy.

4

THE SEMANTICS OF ACTION

1. ACTION: THE EXPLANATORY OBJECT

We hold that there is no difference, *qua* act, between an intentional act and an act *simpliciter*.[1] The relationship which exists between acts *simpliciter* and unintentional or intentional acts is as follows. 'Unintentional' means that the standard case of an act has been generically departed from, that the object being described, although having some of the features of a standard action, nevertheless lacks some of its necessary ingredients. It marks a standard deviation from the product deemed the standard of agency. To mistakenly, and therefore unintentionally, call on a person you falsely believe to be another is for an event of your agency to be caused through an intention in whose formation a false practical belief functioned. This object although caused by your agency is not the object you intended: it is not the event you caused *as you represented it*; it is *anisomorphic* with your intention. On the other hand, unqualified agency occurs when one's efficacious intention is informed by true beliefs, and therefore when the intention finds its intentional isomorph effected. Thus the form of deviation from the standard called 'mistake' falls under the main generic term for action deviations: 'unintentionality'. The place for the claim of *intentionality* is in reaction to one of unintentionality, and, of course, such a denial of unintentionality carries with it the implication of the return to unqualified or standard agency. Where we have such terms as 'unintentional' which

purportedly mark the absence of a feature necessary (and therefore important) to our central and standard cases, we also have terms which rebut such purport. Terms of rebuttal, such as 'intentional', consequently add no additional features to the standard case, have no additional semantics of their own, but have their total function absorbed in merely setting us back to that standard through their rebuttal of the presupposed assertion or suggestion that the standard has been essentially deviated from. It is, as we have argued,[2] a mistake in strategy, then, to look for something called 'intentional' as over and above what is already in action *simpliciter*, just as it was an error to look for something additional called 'reality' in our objects, or something additional called 'freedom'[3] in our standard actions.[4] In order to discover what intentionality is we must identify the essential features of action, those elements of which the ascription of unintentionality denies the presence, and not look for intentionality in anything beyond their presence.

To explain the concept of action, however, we need to address not just the intentionality of the agent but the functionality behind the concept of action as well. Action, we have argued, is *doubly* intentional. First, an action is an object caused by a certain state of cognitive efficacy of an agent and whose properties are those under which the agent cognized the intended object. And second, this object is held to this structure because the concept of *action* has a practicality which impacts its semantics. A semantic interpretation of the question 'What did you do?', or 'What was your action?', which ignored the function of the concept[5] would, then, be misconceived.

Your act *simpliciter* is that object you, as its agent, truly and justifiably intended and caused, as you intended it. We thus ascribe an action unqualifiedly to an agent when and only when the case has the following features:

(1) The agent caused an event
(2) with a certain intention, where
(3) the representations intended were true, and
(4) the agent was justified in the beliefs which determined those representations.

The absence of each of the above features marks a well-known deviation from the case of a true action ascription:

(1) Where causation is absent there is no action ascribable; and *alibi* is.

(2) Where there was causation but the causing event was not an intention, again there is no action; rather, there are merely physical or non-practical events causally related.

(3) Where there was intentional causation but with false information about at least one intended property of the event of agency, then *mistake* is ascribed. When mistake is ascribed a mis-intention is implicated.

(4) Where there was intentional causation but the event of agency lacks at least one intended property due to an unforeseen contributory field event, then *accident* is ascribed; and *vice versa*.

Mistake and accident comprise the *un*intentional. *Inadvertence* is *extra-intentional*:

(5) Where there was intentional causation and the event of agency has the properties intended of it then any additional properties which were not intended together with those events which were side- or end-effects of the intentional comprise the inadvertent or extra-intentional.

There remain cases in which an active attitude, sometimes an intention, was causal but was not formed according to certain normative standards: that is, cases in which casual states were not formed *rationally* or *freely* or were formed *negligently*.

(6) Where there was cognitive causation but no positive evaluation then an active attitude was causal due either to (a) *compulsion*: an overpowerful active attitude, or (b) *recklessness*: weakness of the information function. The form of irrationality called 'self-deception' occurs when an intention is formed by a process which includes a false evaluative belief which was caused by an overpowerful desire state rather than by other beliefs relevant to evaluation.

(7) Where an agent acted but the goal-functional component of his intention, namely, his means and goal, was determined either by another agent or by events in the world, then *coercion* or *necessity* are respectively ascribed but not *unintentionality*. An (standard) action would be ascribed in these cases but there would be a *marker* ('coerced' or 'necessary') to indicate that the circumstances of intention formation had been irregular, were *non-autonomous*, although the intention was successful, as opposed to non-standard cases such as *mistake* and *accident* in which the intention was not.

Just as there are the action adjusters, 'mistakenly', 'accidentally' and 'inadvertently', there are intention- and belief-formation adjusters which mark deviations from the functional norms which govern these formations: 'compulsively', 'recklessly', 'impulsively', 'self-deceptively', "coercedly", and 'necessitously'.

(8) Where all the required functions for intentional causation are in place; where, that is, goal-functionality has been sufficiently identified and positively evaluated so that an intention has been formed and no belief of the agent stands in the way of its efficacy, yet causation does not occur, then *akrasia* is the case. *Akrasia*, then, has a structure inverse to compulsion in certain respects. In compulsion, the active function is of such strength[6] that the information function is to a large degree irrelevant to the causal sufficiency of the active function. In *akrasia*, the active function of the intention is of such weakness, for whatever non-informational reason, that despite rational sufficiency, the intention fails to be causal.

(9) Where there was true representation of the event of agency but the beliefs which determine the content of the efficacious intention were not justified or were believed to be unjustified by the agent who employed them (i.e., they were doubtful beliefs), we have cases of *serendipity* or of *trying*, but not, again, the full normativity of agency.[7]

If causation holds as a real relation between events,[8] then, if agency is causal, causation must relate individual events in the matter

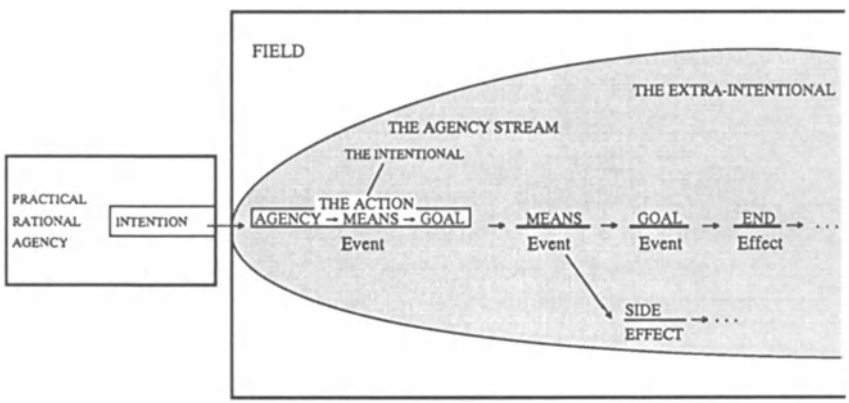
of agency. So, our relationship to our actions, if realistically causal, has a state of our agency as one term and an event which is its effect as the other. The ultimate state of one's agency, where it is rational, is an efficacious intention developed through the rational process we described in chapter 3. The effect of agency is an event of agency, perhaps a bodily movement or change, but it could be any event so caused. Given the purpose of the concept of action, however, the action will be the intended properties of the event of agency. An action *simpliciter* is that kind of entity whose properties are defined by, and limited to, those which correspond to the properties intended of it. This definition and limitation of the object are in place in order to serve the informational-*cum*-explanatory function of action. Such practical objects are as capable of appearing in extensional contexts as is any other object of reference.

The double intentionality of action concepts entails, first, the intentionality of the agency whose efficacy is the condition which separates acts from other events. Secondly, there is the intention or practicality which makes sense of the use of action concepts and their larger family. Davidson's drive for an extensional treatment of action allows him to deny an essential role to the explanatory properties of action. What the naval officer who mistakenly sank the Bismark *caused* is, we know, extensionally conceivable apart from his intention. But we cannot separate the idea of *action* from its practical *raison d'etre*. The purpose of action ascription is to give information about the intentional state which was causal in the world by means of a reference to its worldly isomorphic effect. The service of that goal requires that the intended properties of what was caused be an essential part of what is referred to with action ascription. This brings in as essential the intention which Davidson found inessential to what an action is but without which its fundamental practical role cannot be addressed. The move which brings the agent's intention in through its worldly isomorph is also, as we have seen, a move against those who hold that *intention*, as it modifies mental events, is a different concept from *intention* as it modifies events in the world.[9]

2. INTENTIONAL ACTION

Recall that the logical space of action, which contains the act-objects themselves as well as all the other items relevant to action ascription, consists, first, of a chain of events called 'The Agency Stream' which begins with an event of agency and includes all events causally downstream from that event. Recall further, that depending on the nature of the case, the agency stream will be comprised of parts which are intentional, unintentional, and extra-intentional. The relation a segment of the stream bears to the intention which caused it determines the category to which it belongs. If it was intended, then it will be a member of the intentional part of the agency stream; if it was the result of unintentionality, then it will be a member of that part of the agency stream; and if it was not part of the intentional but "flows off" the intentional, it will be part of the extra-intentional agency stream. Besides agency stream events there are field events which either converge causally with a member of the agency stream or stand in no causal relation to any such event. Each event in these logical spaces will also be either foreseen or unforeseen. The intentional properties of events are, *ex hypothesi*, foreseen (but not vice versa). *Foreseeness* or *unforeseeness* is essential information about the nature of the agency involved, and an explanatory or normative commentary on the agent would be incomplete without it.

FIGURE 1. THE LOGICAL SPACE OF INTENTIONALITY



We begin our account of the semantics of action with the construction of a case of action *simpliciter*.

An action is an object which necessarily makes reference to the nature and certain of the content of the antecedent process of agency. This is so, we claim, since the language of action ascribes the sort of object whose properties constitute an explanation of why an event was caused by an intention. An object which is adequate to explain why an event of agency with certain properties occurred will consist of those properties which were represented in the intention which caused it. Such a diagnostic object is only possible because of the relation of representation which holds between an intention and what it causes.

Suppose the agent "intentionally calls upon his neighbour," as it is over-described in the literature.[10] This is to ascribe to him that object which is comprised of the event he caused, as he truly[11] and justifiably[12] intended it and not as it is otherwise truly representable; not, that is, in terms of an anisomorph of the intention, which would render the object undiagnostic. His intention will consist, in this case of full agency, of a motivation to call upon his neighbour and an evaluation and true identification of the causal relations to the event which would satisfy his motivation. That is, his intention will consist in a certain goal-functionality.

If it is true of the agent that he unqualifiedly called on his neighbour then the act or object ascribable to him, being a worldly isomorph of his intention, has all and only the causal relations to his goal which were intended. The information contained in his efficacious intention is derived from the *concatenants* of the process of practical reasoning or rational intention formation. That he so acted presupposes the following:

(1) he was motivated;[13]
(2) his motivation was for the calling on his neighbour;
(3) he evaluated that motive positively;
(4) he identified an event of his agency which would have causal relations to his goal event;
(5) he evaluated those relations positively;

(6) his evaluations and identifications were informed by true, justified beliefs;
(7) a state of intention with the concatenated content from the above resulted and was effective.[14]

If the above process occurs, then the way the world unfolded from the agent's efficacious intention to the goal which he had therein represented was the way he intended it to unfold. Therefore, that one worldly event in this unfolding which was directly under his control — the body-event of his moving thus and so — was the first event in the agency stream which includes the causal chain of events from his body to the event of his neighbour being called upon by him. The intended causal properties of this event of his agency, in virtue of which it has causal relations to his goal, comprise his action, an object constructed to explain the agent's exercise of his agency, or for obtaining information concerning the use of that power.

An action ascription refers, then, to the goal-functionality of the agent's event of agency. We represent that essential practical structure as follows:

$$\frac{\text{AG ww} \rightarrow \dfrac{F}{E_1}}{E_x}$$

Further descriptions of an action are possible and could include references to the event of agency through the use of descriptions of its means events: 'He took the right fork in the path'. There will be, in the process, positive evaluations of these means structures, as represented, which are conclusions of means-end reasoning concatenated into the intention from premises which are part of the process of intention formation. This fuller information includes the premises behind the identification of the agent's event of agency as the one with causal relations to his goal and as the one which he should therefore cause, as well as the premises behind the positive evaluations made of the causal properties of the event of agency being identified.

3. UNINTENTIONALITY

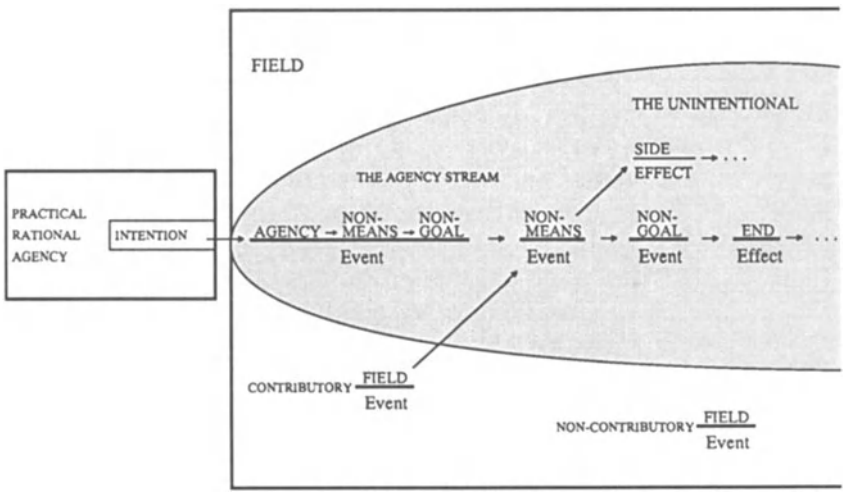
The purpose of action attribution, we have argued, is to explain the actual properties and relations of an event of agency by means of the intentional structure of the agency-state which caused it. Thus terms such as 'means' and 'goal' describe causal relations of an event of agency in terms of structural features of the intention which was causal. When there is intentionality, and therefore action *simpliciter*, what gets explained by a true action ascription is the event of agency with all and only those properties which were intended of it. What does the explaining is the mental isomorph of this intended effect.

However, as we are all too aware, there are cases of agency which go awry, where isomorphicity between the intention and the properties and relations of its proximate effect does not hold. This failure of matching marks the *unintentional*. Mistakes and accidents, which comprise the category of unintentional explanatory objects, come into play when the informational content of an intention is only partially explanatory of the event of agency it causes. What is required for explanation of such an event of agency is an unintentional act-object which will consist of that content of the intention, if any, which has corresponding properties to its effect together with those actual properties of the event of agency which explain why it failed to have all the properties intended of it. Thus, for unintentionality, the explanation of why the event of agency had the properties and relations it did is in terms of the intention which was causal plus certain actual properties of the event of agency which were not intended.

Suppose that while intending to visit your neighbour you mistakenly take the path to the beach in the false belief that your neighbour lives that way. If so, then the intention which was causal represented *certain* properties of the event of agency but *falsely* represented certain essential others, thus *tainting* the entire agency stream with unintentionality. Where the falsity of a means belief so affects its intention that what it causes is not what was intended, the intention alone cannot explain what it causes. Beyond the point of divergence between what was intended and what was caused the

structure of the world must be used to explain why the intention had the effect it did. Thus, in order to be the explanatory unintentional object it is, a mistake must consist of represented and *mis*represented, and therefore *un*intentional, properties and relations of the event of agency.

FIGURE 2. THE LOGICAL SPACE OF UNINTENTIONALITY



To say, then, that you mistakenly took the path to the beach is to say the following:

(1) Your intention was (proximately) causal of the corresponding event of agency.

(2) The practical content of your intention — the representation of the intended causal route to the goal — was only partially isomorphic with, and therefore false of, the actual causal properties of the event of agency.

(3) The event which gives its name to the mistake (*taking that path*) was either misrepresented in your intention as to its relations — it was not a means — or it is an event which occurs as the result of

this misrepresentation and is therefore not the goal (*arriving at the beach*). In mistake, every event in the agency stream is unintentional.

The ascription of unintentional acts, such as mistake and accident, proceed through the denial of the syntax of intentionality. It is to deny, for example, that the object imputed is isomorphic with the informational content of the state which was its cause.

3.1 Action vs. *Mistake*

Mistake is one of the categorizable cases of deviation from intentional action. The object which will be ascribable to the agent with the claim that he mistakenly went to the beach will be derivable from adjustments to the object which is truly ascribed to him with the claim that he unqualifiedly went to the beach. We proceed, therefore, to first construct the object the agent's *going to the beach* within the *Square of Intentionality* and will then construct his *mistakenly going to the beach* within the *Square of Unintentionality* by means of adjustments to the first.[15] These squares contain quadrants of information essential to the explanation and appraisal of agency.

Since a mistake is a standard deviation from an action, the nature of the mistaken act-object will be revealed by noting which of the properties of the object ascribed to an agent with a claim of action are being denied with the claim of mistake, with what they are replaced, and what remains in place from action. The conditions which determine the nature of the act-object *going to the beach* are:

(1) the agent was motivated,
(2) the motivation was for going to the beach,
(3) he positively evaluated this motive or state,
(4) he identified an event of his agency which would have the causal relations to bring about his goal event,
(5) he positively evaluated those causal relations,[16]
(6) his identifications were informed by true, justified[17] beliefs, and
(7) a resultant state — the intention — with the concatenated content from the above occurred and was effective.

Where the above conditions hold, the object or act ascribable to the agent has all and only the practical properties intended: the act is the

object isomorphic with the practical representational content of the state which was its cause; namely, the state described in (7).

The intention to go to the beach, which would be in square (1) of a Square of Intentionality, would have the following structure:

> of his agency, Ag, which is his moving thus and so, H, which would cause the event, E_2, of his moving down that path, G, which he believes would cause the event, E_1, of his being at the beach, F.
>
> ——————————→
>
> the event(s) E_3

Further symbolized, this intention to go to the beach is as follows:[18]

$$\text{Ag+H ww} \rightarrow \underset{E_2}{\underline{G}} \text{ ww} \rightarrow \underset{E_1}{\underline{F}}$$

$$\xrightarrow[E_3]{\qquad\qquad}$$

The principle we see at work in the ascription of an act is that the act-object ascribed to the agent is what was caused by the agent only as he truly and justifiably intended it. In terms of the Square of Intentionality, this means that the intentional state of the agent, represented in square (1), must, if it is to be a true representation, accord with what happens in the world, which is represented in square (2).

(1) INTENTION

$$\text{Ag+H ww} \rightarrow \underset{E_2}{\underline{G}} \text{ ww} \rightarrow \underset{E_1}{\underline{F}}$$

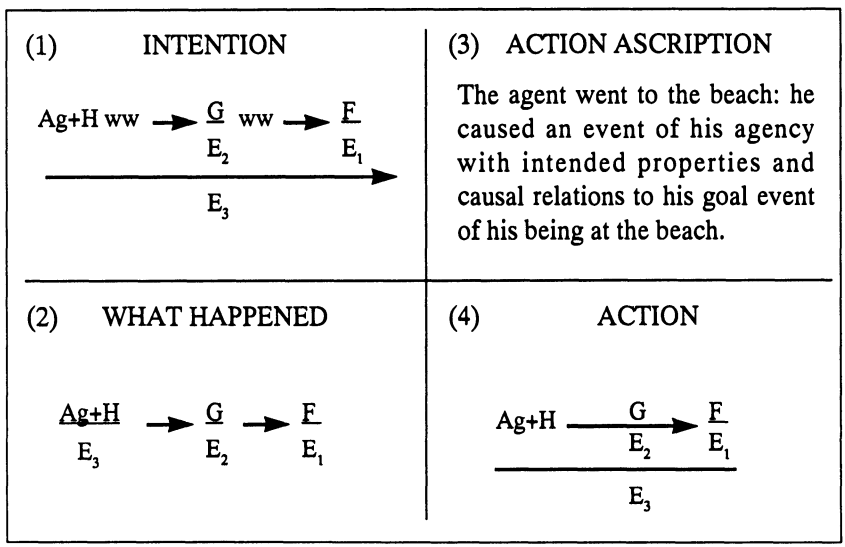
$$\xrightarrow[E_3]{\qquad\qquad}$$

(2) WHAT HAPPENED

$$\underset{E_3}{\underline{\text{Ag+H ww}}} \rightarrow \underset{E_2}{\underline{G}} \rightarrow \underset{E_1}{\underline{F}} \ \ldots$$

When this is the case, then, according to the principle of act-ascription, we may create an object in square (4) which is what was caused by the agent as and only as he intended it and which is attributable to the agent.

FIGURE 3. THE SQUARE OF INTENTIONALITY FOR GOING TO THE BEACH



F/E_1 is the event of the agent being at the beach which is his goal event and which he believes will be caused by G/E_2, his moving down that path which he believes will be caused by H/E_3, his moving thus and so.

Against this background, the amendments asserted and implied by the ascription of *mistake* can be made. Those amendments to the standard cases of action leave us with a picture of the agent's intentional state and its semantic counterpart as he was causal: the explanatory object referred to in mistake attribution. The first

amendment is that being at the beach was not his motive or desire. The event he was endeavouring to identify in his practical rational exercise was not the event of his agency which would lead him there. So we deny the ascription to him of that act-object which has causal relations to his being at the beach. Nevertheless, since with the ascription of mistake *agency* is not being denied but asserted, some other not yet specified motive was at work, some other intentional state is ascribable to the agent. Let us assume that he actually wanted to see his neighbour. The second amendment then is that the agent held the false means belief that the path to the beach leads to his neighbour. So he misidentified the event he wanted to cause, which was the event which would have causal relations to his being at his neighbour's, as the event which also had causal relations to his moving down this path. The agent then harboured the following structure as he acted:

> of his agency, Ag, which is his moving thus and so, H, which would cause the event, E_2, of his moving down that path, G, which he believes would cause the event, E_1, of his being at his neighbour's, N.
>
> ——————————→
>
> the event(s) E_3

Further symbolized, this becomes:

$$\text{Ag+H ww} \rightarrow \underset{E_2}{\text{G}} \text{ ww} \rightarrow \underset{E_1}{\text{N}}$$

$$\xrightarrow{\qquad\qquad} E_3$$

The representational structure contains a false means-belief (the belief that G/E_2 would cause N/E_1) and is thus a misrepresentation of the causal properties of the event of agency which the agent was motivated to cause. The representation which would otherwise have been true of the event he causes is truncated in the example at the causal connection between the movements down the path and the goal: that goal is not truly representable as causable by those movements. The object which is ascribable to the agent will, then, be his event of agency with causal relations to an unintended event

which was caused by his mistaken means event, an event which is therefore *not* means just because its effect was not his goal. So the syntax which binds intentions and actions is not applicable in mistake: the event of agency had relations to events which are neither means events nor goal events. The event of agency is thus not representable in the practical terms which it would need to share with its intentional cause in order to be an action. Nevertheless, these denials of action syntax to mistake constitute an explanation of why the intention failed.

FIGURE 4. THE SYNTAX OF ACTION AND MISTAKE

ACTION:

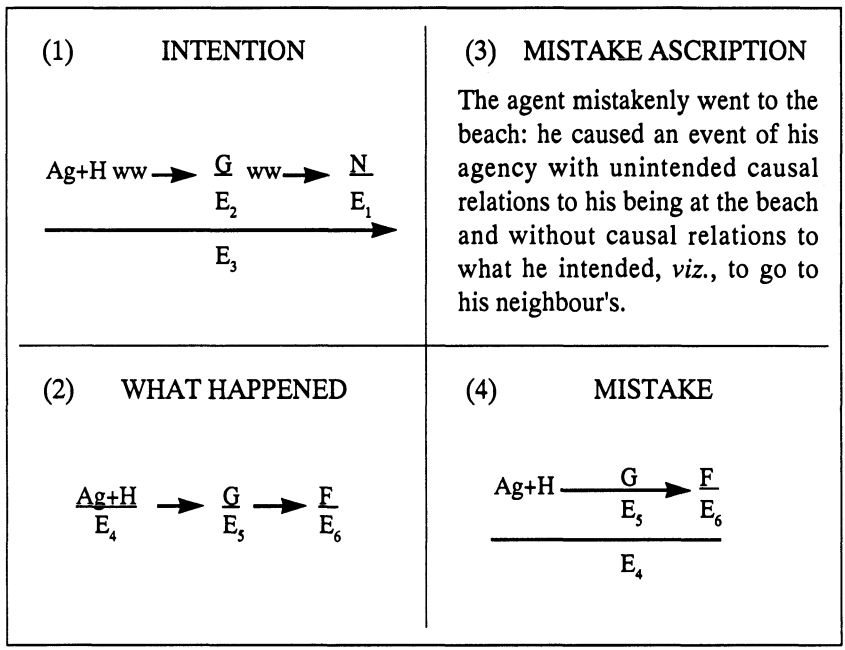
$$\frac{\text{[AGENCY] [MEANS] [GOAL]}}{E_x}$$

MISTAKE:

$$\frac{\text{[AGENCY] [NON-MEANS CAUSAL RELATION] [NON-GOAL CONSEQUENCE]}}{E_x}$$

In summary, if an object is to be the object intended it must have the intended causal properties for the goal. The object which has causal relations to a "false means", and therefore not to a goal, is not the object intended. So what was intended did not occur and what occurred was not intended. To explain what did occur beyond the false means we need the structure of the world itself, its non-agency causes. A mistake, being an explanatory object where intention was causal but anisomorphic, must then use both the intention and unintentional to be explanatory. That is what a mistake is: a case of incomplete intentional explanation together with an explanation of its incompleteness and an implication of what it would take for completeness: an object which explains itself partly by its causal relation to the practical intentional state which caused it and partly by its causal relations to events in the world. A mistake is an object which explains its own unintentionality.

FIGURE 5. THE SQUARE OF UNINTENTIONALITY FOR MISTAKENLY GOING TO THE BEACH



N/E_1 is the event of the agent being at his neighbour's which is his goal event and which he believes will be caused by G/E_2, his moving down that path which he believes will be caused by H/E_3, his moving thus and so.

F/E_6 is the event of the agent being at the beach which was caused by G/E_5, his moving down that path, which was caused by H/E_4, his moving thus and so.

3.2 The Sinking of the Davidson

For comparison, consider Davidson's treatment of *mistake*. In his well-known example, a British naval officer mistakenly sinks the Bismark in the false belief that it was the Tirpitz. On Davidson's analysis of the nature of the mistake, we ascribe to the officer the event of his agency: that event of his finger moving against the firing button which caused the torpedo to be launched so that the Bismark,

not the Tirpitz, sank; that is, his bodily event with those relations. Since this event of the officer's agency had those causal relations which unfolded in the Bismark's sinking beneath the waves, that event is his action. Davidson believes that the officer intentionally sank that ship he sighted on the horizon since he intentionally moved his finger ... so that the ship, although actually not the Tirpitz, would sink. If he had those intentions, then what we ascribe to the officer as his act, differs, on a Davidsonian account, from what our theory finds ascribable to him. For Davidson, since the officer intended and caused that movement of his finger, he intentionally so moved his finger. And since that moving of his finger was also, by virtue of its causal relations, the event which is referable to as the sinking of that ship, then since that ship was the Bismark, he intended not only to sink that ship, but, by substitution, to sink the Bismark.

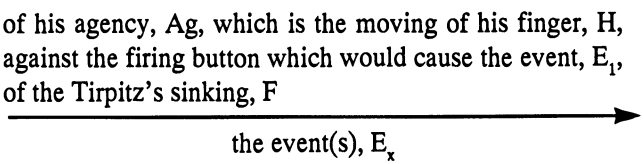
> Suppose an officer aims a torpedo at a ship he thinks is the *Tirpitz* and actually sinks the *Bismark*. Then sinking the *Bismark* is his action, for that action is identical with his attempt to sink the ship he took to be the *Tirpitz*, which is intentional.[19]

According to Davidson, so long as the event of his moving his finger was caused by the officer's relevant intentional state, which would be that state which contained some true representation of its effect, that event was his action whatever the other subsequent properties of that event.

On our view, however, the officer's act is the object which mirrors in its structure the intentional state which was its cause. In the fully intentional case, the act-object ascribed is an artefactual object, one extracted for a purpose, whose properties are defined by the nature of the effective intention in order to address that intention through the specific object it represented. For Davidson, the act-object has both intentional and unintentional properties and relations all of which may be used to refer to the action and therefore none of which is essential: it is the event caused whatever its properties, so long as it was caused by an effective intention with at least one true representation of it. If we know that the naval officer sank the Bismark, the object so identified, on the Davidsonian view of action,

gives us no information about the nature of its cause. Action, on this analysis, is an uninformative concept except for its import of some unspecified intentional causation. If so, we would do better to throw out action discourse altogether and work with events alone. But then we should be deprived of the service of the powerful, self-explanatory and diagnostic object which an action, we argue, is.

In contrasting our view of the semantics of *mistake* with Davidson's, we shall see why unintentional objects, although partly isomorphic with their cause and thus not the objects intended, are nevertheless self-explanatory. The officer, it would be agreed, did cause his finger to move. And the movement of his finger did have those causal relations to the world so that the Bismark sank. The movement of the naval officer's finger is an event which has the causal relations required for that sinking. But it is not yet the intentional sinking of the Bismark, not that act-object, until we determine how the event of the officer's finger movement was represented by him since his act will be that object with and with only those causal properties intended in his effective intention. Since it was a mistake, we know that the officer did not, in the formation of his effective intention, represent the event he intended to cause — his finger movement — as the event which would have causal relations to the Bismark's going under. What he intended to cause was the event which would initiate the Tirpitz' going under. His intentional state was:



We must not forget that this state, being an intention, would have passed an evaluation. We may then add that the event of his agency with the relations and property so far identified was evaluated as desirable,[20] as worth doing.

Next, the agent would have endeavored to identify further the event he will cause by means of alternative descriptions of the causal

relations he believes it will need in order for it to be the event of his sinking of the Tirpitz with, as we know, an evaluation of that doing at each stage in the development of its identification. The officer now comes, according to the example, to the false belief that the ship he descries on the horizon is the Tirpitz. The new description of his intended act as the sinking of that ship, is evaluated. However, at this point, and subsequently, the officer is identifying and therefore evaluating the *wrong* object. The event he will cause with the causal relations he has identified, will not be the event he identified and intended in his first intentional state described above. That would have been the event of his agency which would have had the causal relations to the sinking of the Tirpitz. Now, however, the event of his agency which he seeks to identify is the one which will be, unbeknownst to him, the sinking of the Bismark.

We should now be aware that wherever a false belief relevant to the identification of the event of agency enters that process, the event the agent identifies and evaluates is not the event he intends.[21] It is not the event which fits into the *original frame of the identifying description*: 'the event which will be the initiating cause in a chain to the sinking of the Tirpitz'. For a false belief about means to enter this frame in which attitude and content are fused together is for the event of agency being described and identified to have become the wrong event, not the intended one. It cannot have the relations required by the framing description where false beliefs have determined those relations. But the object both desired and intended is the event with the causal relations to the sinking of the Tirpitz. The active attitude of intending is fused to that structure, not another, even when that other is (mistakenly) causal.

So the officer has identified and evaluated and caused an object which is not the object he intended according to the framing description in his first intentional state. Relative to that motivating state of his agency, relative to the object which he intended and was therefore set to identify and evaluate, he has misidentified and misevaluated. Because of the false belief that the event of his agency is the event which will cause the sinking of that ship on the horizon, actually the Bismark, the proper causal relationship of the event of his

agency to his goal is lost. The event he causes is not the event he intends or intended. He is now *misintending*: he has a false intention. But Davidson holds that the officer intended to sink that ship which he took to be the Tirpitz.[22] That he had this intention, however, is false. If he intended to sink the Tirpitz and he intended to sink that ship which he took to be, but was not, the Tirpitz, then within the formation of a single intention he both intended to sink the Tirpitz and intended to sink that ship which is not the Tirpitz. He therefore intended an action which was both the sinking of the Tirpitz and the sinking of that ship which was not the Tirpitz. The account of intentional states which Davidson requires in order to explain the Bismark's sinking as the naval officer's action ignores identity considerations for intentions and therefore violates the consistency requirement on intention formation in rational agents.[23]

The source of this trouble for Davidson's account is that it has an improper conception of the nature of what is intentional. The proper conception takes the intentional to be the object caused with all and only the intended goal-functionality. In untying the intended object from its goal the above inconsistencies are made possible. The identities of intentions require that the tie between the object of the intention, which is an event of agency, and the goal, which defines the causal relations of the intentional object, be maintained.

If the naval officer misintended to sink the Bismark, he therefore sunk the Bismark by mistake. What object is his mistake?

The mistaken object is the event of the naval officer's agency, his moving his finger so, with causal relations to the sinking of that ship on the horizon, actually the Bismark. The object has that structure. It is furthermore an object which occurred due to an intentional state but which was not intentional. There is not in the naval officer's intentional state *one* true representation of his event of agency.[24] The structure of his intentional state is anisomorphic with the event of agency since he mis-identifies the event he caused as the event he intended.

With the attribution of the unintentional object mistake, we have, in terms of the above example, the following:

(1) The officer moved his finger in such a way that it had causal relations to the depression of the firing-button, to the firing of the torpedo, and the sinking of that ship, which was the Bismark. He caused an event of agency which had causal relations to at least one event he did not intend his event of agency to have, and
(2) he intended at least one causal relation which it did not have. In this case, it lacked causal relations to the sinking of the Tirpitz, his goal, and
(3) the causal relation intended — the sinking of the Tirpitz — is falsified by the actual causal relations of the officer's event of agency.
(4) The actual event of his agency is then not the event he intended. He intended the event with causal relations to the sinking of the Tirpitz. He intended that singular event: the event with that goal-functionality. The singular event which he caused was some other event entirely.
(5) But the event of the Bismark's sinking stands in an effect relation to the event of his agency, his finger moving, precisely because the officer believed that ship on the horizon to be the Tirpitz, not the Bismark. That belief caused the officer to misrepresent the event of agency he had been intending. The resultant misintention is what explains the mis-sinking of the Bismark.

In this way, the ascription to the officer of having mistakenly sunk the Bismark explains the event of his agency having the relations it had by means of a reference to the intentional state which was the cause of that event and through certain assertions and denials of its content. We are given, with mistake, an object with the properties above described, and through which, the properties of the intentional state which was its cause are sufficiently described to constitute an explanation of the mistake.

3.3 Accident

Suppose the driver accidentally struck the pedestrian. Then it is false that there was an event of agency which the driver represented as the event which would initiate a causal chain to his goal of the pedestrian's being struck. The denial of this ascription is based upon the fact that the pedestrian's being struck was not represented as goal by the agent since that representation was, as accident entails, not among the agent's causally relevant cognitive or attitudinal states. What is ascribed with accident is, as with mistake, an act-object which is at most partially isomorphic with its intentional cause but not, as it was with mistake, because the object lacks causal connection to the agent's goal event due to a false belief which caused him to misrepresent the event he caused as being the event with the causal relations he wanted, but rather, in accident, because the event he caused has had cut, by unforeseen other events not caused by the event of agency, the causal relations it would otherwise have had to the agent's goal. The syntactical structure of accident will contain the same denials of the syntax of action as did mistake but will include a different syntactical component as reason for the denials.

FIGURE 6. THE SYNTAX OF ACCIDENT

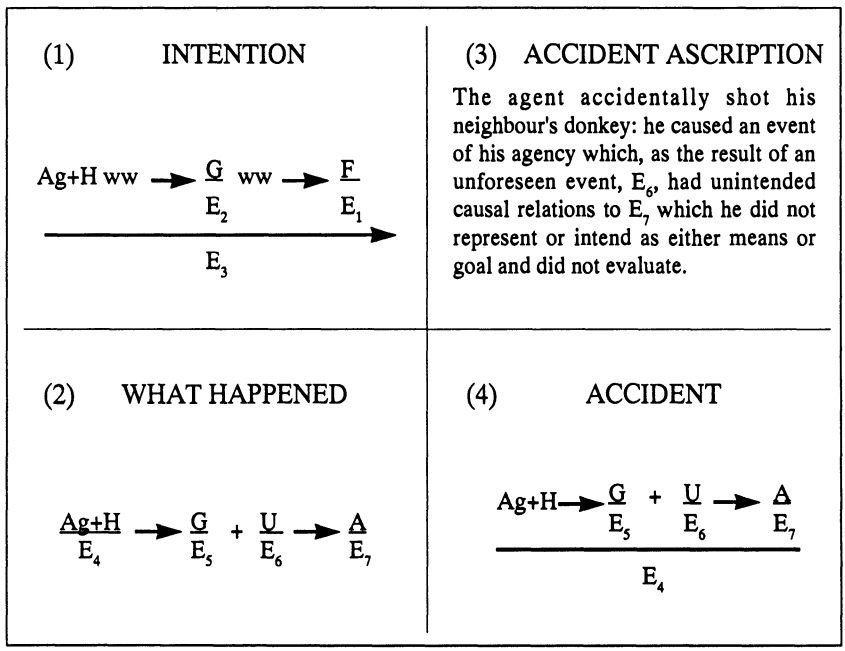
$$\frac{[\text{AGENCY}]\ \begin{bmatrix}\text{NON-MEANS}\\ \text{CAUSAL}\\ \text{RELATION}\end{bmatrix} + \begin{bmatrix}\text{UNFORESEEN}\\ \text{INTERVENING}\\ \text{CAUSE}\end{bmatrix}\ \begin{bmatrix}\text{NON-GOAL}\\ \text{CONSEQUENCE}\end{bmatrix}}{E_x}$$

Accident, like mistake, is an object which explains the unintentionality of the event of agency; explains why, that is, the event of agency does not have the properties and relations intended of it although it is an event of agency.

Consider Austin's famous case:

You have a donkey, so have I, and they graze in the same field. The day comes when I conceive a dislike for mine. I go to shoot it, draw a bead on it, fire ... but as I do so, the beasts move, and to my horror yours falls.[25]

FIGURE 7. THE SQUARE OF UNINTENTIONALITY FOR ACCIDENTALLY SHOOTING THE NEIGHBOUR'S DONKEY



F/E_1 is the death of my donkey which is my goal event and which I believe and intend will be caused by G/E_2, the firing of my carefully aimed gun, which I believe and intend will be caused by H/E_3, the moving of my finger against the gun's trigger.

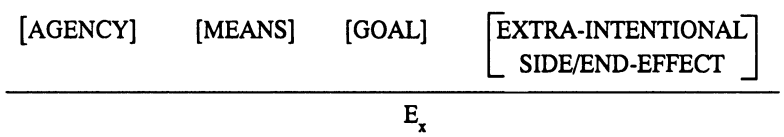
A/E_7 is the death of my neighbour's donkey which was caused by U/E_6, his unforeseen stepping into the line of fire of the bullet which was discharged by G/E_5, the event of my gun being fired, which was caused by H/E_4, the movement of my finger against the gun's trigger.

4. THE EXTRA-INTENTIONAL

Inadvertence, the paradigm of the extra-intentional, does not, as do mistake and accident, deny the applicability of the syntax of action to the event of agency. The inadvertent object is one with:

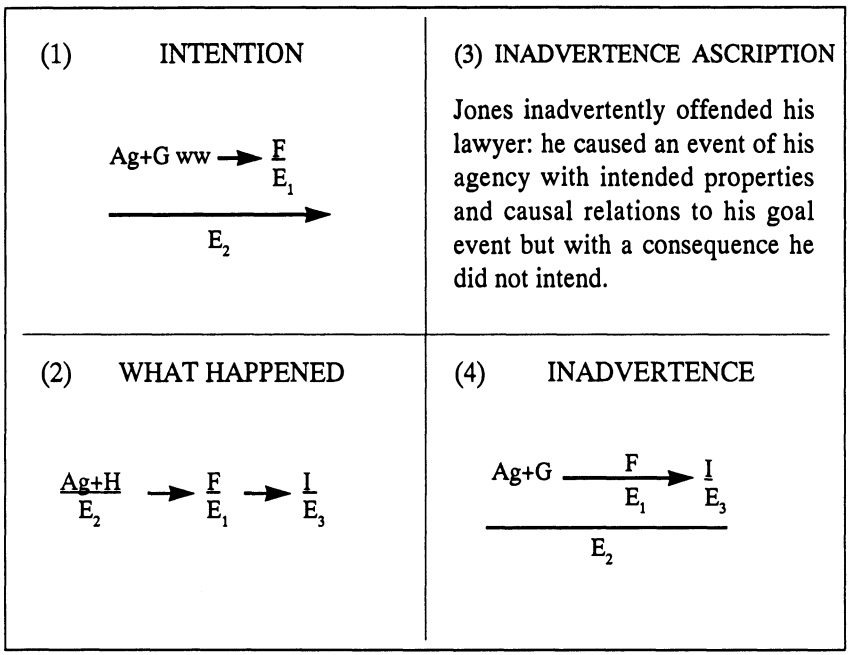
(1) All the properties of the event of agency which were represented and intended as either means or goal, but with
(2) at least one relation or property which was not represented as either means or goal.

FIGURE 8. THE SYNTAX OF INADVERTENCE



An inadvertence, then, contains, as a proper part, a fully intentional object, an action — the event of agency as the agent intended it — plus at least one further causal relation or property not intended. (The extra-intentional aspect cannot be due to either accident or mistake due to the first condition above.) If Jones moves in order to speak to his dinner companion, an intentional action, let us say, and so turns his back with offending consequences to his lawyer, then whether or not he realized the offence would occur, it was inadvertent: Jones inadvertently offended his lawyer. The offence was extra-intentional precisely because this consequence falls outside the structure of his intention. Recall that this is in itself not a normative judgement about Jones' carelessness. However, side-effects and end-effects of one's actions are precisely the point at which normativity may enter.

FIGURE 9. THE SQUARE OF EXTRA-INTENTIONALITY FOR JONES' INADVERTENTLY OFFENDING HIS LAWYER



F/E_1 is Jones' facing his dinner companion which is his goal event and which he believes will be caused by G/E_2, his moving thus and so.

I/E_3 is Jones' lawyer taking offence which was caused by F/E_1, Jones' facing his dinner companion which was caused by G/E_2, Jones' moving thus and so.

When we truly attribute an intentional object — an action — to an agent we impose a structural template upon the event he caused. This structure is the one shared with and derived from the intentional state which was, *ex hypothesi*, causal. That state, being itself a stage in the process of practical rationality, has, we know, the following generic syntax:

$$\frac{[\text{AGENCY}]\ [\text{MEANS}]\ [\text{GOAL}]}{E_x}$$

When an act is truly attributed, then the object ascribed and the state which was causal, are, since they are held in the truth relation, both bound by this syntax.

Within that structure we know *means* and *goal* are interdependent pieces of syntax. An event is a means *iff* it has the intended causal relations between the event of agency and the goal; and an event is goal *iff* it has the intended causal relations with means events and/or an event of agency. The event of agency, however, is not engaged in this interdependence between means and goal. Since it is the event which is the effect of a state of intending, it is assumed by the others, but need not, as does the attribution of action, assume them.

When an intentional action is truly ascribed then the above syntax is in place and events, as intended, are in these relations. When, however, unintentionality or extra-intentionality are truly ascribed, some deviation from this syntax has occurred. With both, an event of agency will have occurred: this is the very basis of the connection between intentionality and its deviations — unintentionality, and extra-intentionality — the very reason that the latter two are qualifications upon action. With unintentionality, either means or goal is denied and where either is, both are, as we have seen. With extra-intentionality, rather than being denied, means and goal are asserted, but with that, so is consequence. The three types of action deviations just are the objects which explain why certain properties and relations of the event of agency are not representable in the syntax of intentional action. That is because the property or relation is present due either to false means-belief, the unforeseen intervention of a non-goal-causing event, or is an unintended, although perhaps foreseen consequence.

5. RESPONSIBILITY

On our theory of action, the informational-*cum*-explanatory function of action ascription is *the* determinant of the nature of the act-objects. With *action* ascription, what is ascribed, and what therefore explains, is the intended goal-functionality of the event of agency. With *mistake* or *accident* ascription, what is ascribed are those unintentional properties and relations of the event of agency which explain why it

was not goal-functional. An *inadvertence* is an object which explains why an event of agency had properties and/or relations in addition to those intended.

The function of a *responsibility* ascription, however, is to provide all the information relevant to appraisals of an instance of agency.[26] Responsibility is, as it were, the *accounting* relation, or the *writing to one's account* relation, and normative judgement or valuation of what is written to the agent's account determines his worth, as it were. Thus, although responsibility is not itself a normative notion, it is an essential premise in the normative judgement of agency.

When an agent has been causal, the non-normative information relevant to a normative judgement of him will come from the agency stream. The properties and relations which occur in the agency stream are what the normatives judge; but through that judgement, the agent of the change is judged. *How much* of the agency stream is considered relevant to appraisals of agents, and therefore how much of the stream is ascribable to the agent as his responsibility, will be a function of how much is relevant to normative theories. With the present state of normative theory, an agent will be responsible for at least that part of his agency stream which is *intentional.* He may be responsible for its *unintentional* and *extra-intentional* properties as well, but since *unforeseeable* properties and events are irrelevant to such an appraisal, the area of responsibility will be limited to that part of the agency stream which is *foreseeable.*[27] Sub-areas within the area of responsibility will include the *foreseen* and the *unforeseen* since these properties of members of the stream will also be relevant to normative appraisals of the agency involved.

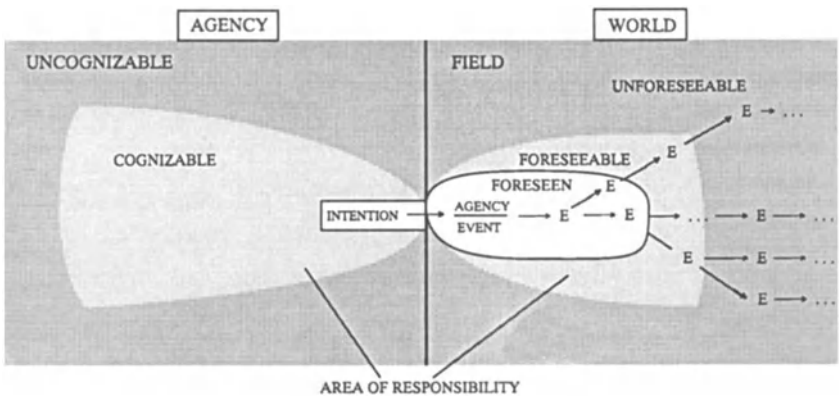
These demarcations define the area of responsibility whenever agency has occurred, whether what was intended was realized or not. It should be emphasized that the properties upon which the above demarcations are based comprise all the non-normative information about properties and events resulting from a case of agency which is necessary for the normative evaluation of that agency.

Thus our view of the functions of action ascription and responsibility ascription differs sharply from those theorists, such as Michael Bratman and like-minded legal theorists, who would allow

"our concerns with the ascription of *responsibility* to shape our classifications of actions as *intentional.*"[28] In the interest of holding agents responsible Bratman would have us distinguish the concept of *intention* which modifies mental events from the concept of *intention* which modifies physical events so that one may do intentionally what one did not intend to do.[29] For us, not only is the business of action ascription distinct from the business of responsibility ascription, but we are able to ascribe actions and responsibilities while preserving a *univocal* conception of *intention.*

The extent of the responsibility object is, as we have seen, determined by the range of certain properties of agency. But responsibility also extends over the processes within agency itself. There also, the extent of responsibility is determined by those properties which are relevant to the appraisal of the agent. The limit of responsibility in the area of intention formation is where the cognizable ceases and the uncognizable commences. This boundary resonates with, but is not identical to, the unforeseeable and the unforeseen.

FIGURE 10. THE AREA OF RESPONSIBILITY



5

A DESIGN DOCUMENT FOR *PRAGMA*

1. INTRODUCTION

In this chapter we adumbrate the process of case analysis which a computer program such as *PRAGMA* might use to analyze cases of agency according to our theory. While the steps of the analyses which follow will almost certainly lack fidelity with *PRAGMA* itself, even such a superficial account will help the interested reader to further understand the theory and the program.

CASE LIBRARY[1]

> You have a donkey, so have I, and they graze in the same field. The day comes when I conceive a dislike for mine. I go to shoot it, draw a bead on it, fire: the brute falls in its tracks. I inspect the victim, and find to my horror that it is *your* donkey. I appear on your doorstep with the remains and say — what? 'I say, old sport, I'm awfully sorry, &c., I've shot your donkey by *accident*'? Or 'by *mistake*'? Then again, I go to shoot my donkey as before, draw a bead on it, fire — but as I do so, the beasts move, and to my horror yours falls. Again the scene on the doorstep — what do I say? 'By mistake'? Or 'by accident'? (J.L. Austin, 'A Plea for Excuses')

> Suppose that I am playing a pair of identical video games in which I am to guide a missile into a certain target. The games are known to me to be so linked that it is impossible to hit both targets. If both targets are about to be hit simultaneously the machines just shut down. Both targets remain visible to me, so I can see which target I

hit if I hit either target. I know that while I can hit each target, I cannot hit both targets. I am quite skilled at such things, and ambidextrous, but these are difficult games and I am doubtful of success. Since I know it is difficult to hit either target, I decide to play both games simultaneously. I proceed to try to hit target 1 and also to try to hit target 2. Suppose I do hit target 1 in just the way I was trying to hit it, and in a way which depends heavily on my considerable skills at such games. Do I hit target 1 intentionally? (Adapted from Michael Bratman, 'Two Faces of Intention')

In writing heavily on this page I may be intending to produce ten legible carbon copies. I do not know, or believe with any confidence, that I am succeeding. But if I am producing ten legible carbon copies, am I doing so intentionally? (D. Davidson, 'Intending')

Suppose a naval officer aims and fires a torpedo at a ship he knows to be the Tirpitz and sinks it. Does he intentionally sink that ship? Suppose he aims and fires a torpedo at a ship he thinks is the Tirpitz and sinks the Bismark. Does he intentionally sink that ship? (Adapted from D. Davidson, 'Agency')

A house is on fire. A father is trapped in the attic with his two little girls. He comes to the conclusion that unless they jump they will all be burned alive. But he also realizes that if they jump they are all likely to suffer serious personal harm. The children are too frightened to jump and so in an attempt to save their lives he throws one out the window to the crowd waiting below and jumps with the other one in his arms. All are seriously injured, and the little girl he threw out the window dies of her injuries. On the Select Committee's proposal that unfortunate man would be a murderer. That is because he was aware that "in the ordinary course of events" the little girl would suffer serious injury and therefore by the proposed definition he is taken to have intended to cause serious personal harm to her. He was also of course aware that his act might cause her death. He is therefore guilty of the murder of his own girl when he was actually trying to save her life. (J.C. Smith, 'A Note on 'Intention'')

> Welcome to *PRAGMA*, the PRactical RAtional AGency MAchine. This is a user-interactive system for analyzing cases of agency according to the theory developed above. That theory sees actions as objects which are explanatory of their intentional cause, and action ascription, therefore, as an enterprise informative of the practical rational state which was causal. The system will, with help from the user, answer such questions as:

(1) What explanatory object is ascribable to the agent in a case of rational agency?
(2) Is this a case of (intentional) action, mistake, accident, or inadvertence?
(3) To which intentional category from the *logical space of action* do the objects and properties in his *agency stream*, and any other relevant objects and properties, belong?

The case-analysis is divided into sections in which the user is led through the construction, based on information contained in the case-text (and on allowable inferences and assumptions, if necessary) of: (1) the agent's intention; (2) what happened as effect of the agent's efficacious intention; (3) the act-object ascription or other commentary which is appropriate based on the relation between the intention, (1), and its effect, (2); and (4) the practical explanatory act-object ascribed and/or the intentional category from the logical space of action to which each item in (2) above belongs.

The system employs, as aids to the analysis, several diagnostic tools. There will be Side-text, Help Screens, a Glossary and examples available to help the user. In addition, there are data-structures, the *Squares of Intentionality and Unintentionality* and the *Object Frames*, which represent and organize the information contained in and derived from the language used to describe the case.

2. ACTION

A naval officer aims and fires a torpedo at a ship he knows to be the Tirpitz and sinks it. Does he intentionally sink that ship?

2.1 Introduction

(1) Is this a case of (rational) agency?

(a) YES (b) NO

COMMENTARY: If the answer is 'NO', then, since *PRAGMA 1* is a program for *action* ascription, and 'NO' entails that there was no event of agency and therefore neither action, mistake, accident nor inadvertence, the case is of no interest to *PRAGMA 1*. However, the concepts of *negligence*, (act of) *omission*, and *alibi* are easily addable and are contemplated for *PRAGMA 2*.

As an aid to answering questions, the user may consult HELP and a GLOSSARY of terms. For example:

HELP

A case of rational agency is one in which a practical rational state of an agent is effective of an event of agency: that is, where an intention of an agent is the proximate cause of an event of his agency.

GLOSSARY

ACTION: a practical explanatory object which is the intended GOAL-FUNCTIONALITY of an EVENT OF AGENCY; an action is all and only the intended properties of an EVENT OF AGENCY which enable it to be the first event in the intended causal pathway to the agent's GOAL.

AGENCY PROPERTY: the property of an event in virtue of which it is an EVENT OF AGENCY; namely, its being the proximate effect of an agent's EFFICACIOUS INTENTION.

AGENCY STREAM: the chain of events which is causally downstream of an EFFICACIOUS INTENTION. Events not in the AGENCY STREAM are FIELD EVENTS.

PRACTICAL RATIONAL AGENCY: the causing of an event of agency by a state of an epistemological, teleological and effective system which has been rationalized by the process of practical reasoning.

(2) ANSWER:

YES there was (rational) agency.

COMMENTARY: The analysis of this case will proceed by filling in the required content of The Square of (Un)Intentionality. A description of that data-structure follows:

THE SQUARE OF (UN)INTENTIONALITY

(1) INTENTION A representation of the intention of the agent with its content which consists of the intended properties and relations of the event of agency.	(3) COMMENTARY Language which describes an object from square (2) in terms of the logical space of action: i.e., in terms which are relevant to explanation and appraisal of agency.
(2) WHAT HAPPENED A representation of the agency stream — the event of agency and all relevant downstream events — together with all relevant field events.	(4) LOGICAL SPACE OF ACTION A representation of the object or property described in square (3).

2.2 Intention (Square 1)

(3) What was the agent's efficacious intention?

COMMENTARY: If this is a case of rational agency, then there was an efficacious intention which was effective in the world. An intention has the following structure:

$$[Ag]\ [MEANS]\ [GOAL]\ [Ev^{+}] \longrightarrow [E_x]$$

That is, a rational and causal state, →, of an agent which represents the event, E_x, of agency, Ag, with the causal relations, MEANS, to the agent's GOAL, which the agent has judged to be desirable. To identify the agent's intention the user must identify the intended causal relations of the agent's event of agency. This will require answers to the following questions:

(a) What was the agent's goal event?
(b) What was the agent's means event(s)?
(c) What was the agent's event of agency?

Answers to (a) and (b) provide the goal-functionality of (c), the agent's event of agency, and (b) and (c) pick out the causal pathway from the agent to his goal event.

GLOSSARY

GOAL EVENT: an event with a GOAL PROPERTY whose causing is the object of a DESIRE-TO or INTENTION.

GOAL-FUNCTIONALITY: the intended causal properties of an EVENT OF AGENCY in virtue of which it is the first event in a chain of events to the GOAL.

GOAL PROPERTY: a relational or non-relational property of an event the causing of which is the object of a DESIRE-TO or INTENTION, and which was inherited by the DESIRE-TO or INTENTION from its PROTO-ATTITUDE ancestor.

GOAL: a motivating event or property the causing of which is the object of a DESIRE-TO or INTENTION.

INTENDED PROPERTY: a GOAL-FUNCTIONAL property of an EVENT OF AGENCY.

INTENTION: a causal and informational state of the type ACTIVE ATTITUDE whose function is causal (potentially or actually) with respect to the object it represents in its content, that object having been judged to be, on the whole, desirable.

INTENTIONAL EVENT OF AGENCY: an EVENT OF AGENCY which has all the properties and relations it was intended to have, i.e., those represented in the content of its EFFICACIOUS INTENTION.

MEANS: the intended GOAL-FUNCTIONALITY of an event; the intended GOAL-causing property of an event.

MEANS EVENT: an event with intended causal relations from the agent's EFFICACIOUS INTENTION and to the agent's GOAL EVENT. An event with a GOAL-FUNCTIONAL causal relation.

(4) What was the agent's goal event?

(5) ANSWER: The agent's goal event was the event:
the sinking of the Tirpitz.

COMMENTARY: The system assigns an event-name to the goal event and a property-name to its goal property:

Assign 'E_1' to the event.
Assign 'F' to the property: Tirpitz sinking.
Assign 'F/E_1' to the event, E_1, of the Tirpitz sinking, F.

The system places these values in the appropriate place in the structure of the agent's intention which will then appear in square 1 of the Square of (Un)Intentionality. That is, it places 'F/E_1' in goal position of

$$\underline{[Ag]\ [MEANS]\ [GOAL]\ [Ev^+]} \longrightarrow$$
$$[E_x]$$

The agent's intention:

(1) INTENTION

$$[Ag]\ [MEANS]\ \frac{F}{E_1}\ \ Ev^+ \longrightarrow$$
$$[E_x]$$

The system places this information into the Object Frame for this case:

PROPERTIES	EVENTS					
	E_1	E	E	E	E	E
INTENDED PRACTICAL PROPERTIES						
[AGENCY] [MEANS] [GOAL]						
AGENCY PROPERTY _____	___	___	___	___	___	___
MEANS PROPERTY _____	___	___	___	___	___	___
GOAL PROPERTY _____	F	___	___	___	___	___
ACTUAL PRACTICAL AND NON-PRACTICAL PROPERTIES						
EVENT OF AGENCY						
Causes _____	___	___	___	___	___	___
Is caused by _____	___	___	___	___	___	___
MEANS EVENT _____	___	___	___	___	___	___
Causes _____	___	___	___	___	___	___
Is caused by _____	___	___	___	___	___	___
GOAL EVENT _____	___	___	___	___	___	___
Causes _____	___	___	___	___	___	___
Is caused by _____	___	___	___	___	___	___
CONSEQUENCE _____	___	___	___	___	___	___
Causes _____	___	___	___	___	___	___
Is caused by _____	___	___	___	___	___	___
CONJOINING CAUSE _____	___	___	___	___	___	___
Causes _____	___	___	___	___	___	___
Is caused by _____	___	___	___	___	___	___
NON-CONJOINING _____	___	___	___	___	___	___
Causes _____	___	___	___	___	___	___
Is caused by _____	___	___	___	___	___	___
NON-ACTUAL EVENT _____	___	___	___	___	___	___

(6) What was the agent's means event(s)?

COMMENTARY: What event did the agent believe and intend would cause the goal event?

GLOSSARY

MEANS: the intended GOAL-FUNCTIONALITY of an event; the intended GOAL-causing property of an event.

MEANS EVENT: an event with intended causal relations from the agent's EFFICACIOUS INTENTION and to the agent's GOAL EVENT. An event with a GOAL-FUNCTIONAL causal relation.

(7) ANSWER:

The agent believed and intended that the event:

the torpedo firing

would cause the goal event of the Tirpitz sinking.

COMMENTARY: The system assigns an event-name to the means event and a property-name to its means property:

Assign 'E_2' to the event.
Assign 'G' to the property: torpedo firing.
Assign 'G/E_2' to the event, E_2, of the torpedo firing, G.

The system places these values and the symbol for *which would cause*, 'ww→', in the appropriate places in the structure of the agent's intention.

$$[Ag]\ [MEANS]\ \frac{F}{E_1}\ \ Ev^{+}$$

$$\longrightarrow$$

$$[E_x]$$

The agent's intention:

(1) INTENTION

$$[Ag]\ \frac{G}{E_2}\ ww \longrightarrow \frac{F}{E_1}\ \ Ev^{+}$$

$$\longrightarrow$$

$$E_x$$

The system adds the relevant information to the Object Frame.[2]

(8) Were there any other means events intended?

COMMENTARY: What event did the agent believe and intend would cause the means event? NOTE: This question will recur until an event of agency is identified. At that point, the intended causal chain from the agent to the goal event will have been identified.

(9) ANSWER: The agent believed and intended that the event:

his finger moving against the firing button

would cause the means event of the torpedo firing which would cause the goal event of the Tirpitz sinking.

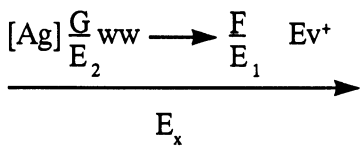
COMMENTARY: The system assigns an event-name to this event and a property-name to its property:

Assign 'E_3' to the event.
Assign 'H' to the property of the officer's hand moving just so.
Assign 'H/E_3' to the event, E_3, of the officer's hand moving just so, H.

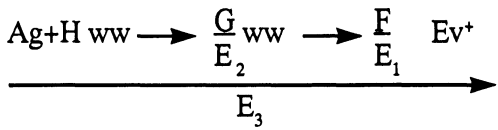
COMMENTARY: Since this event is an event of the agent's body which agent's may cause directly, i.e. without some intervening means event, the event of agency has been identified. The system, therefore, assigns the property 'Ag' to this event.

Assign 'Ag' to the event, E_3, of the officer's hand moving just so, H: Ag+H/E_3.

The system places these values in the appropriate places in the structure of the agent's intention and in the Object Frame.

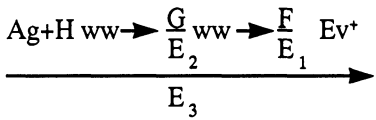


The agent's intention was:



This is a full representation of the agent's efficacious intention. The system places it in square (1) of the Square of (Un)Intentionality for this case:

SQUARE OF (UN)INTENTIONALITY FOR
THE OFFICER'S SINKING THE TIRPITZ

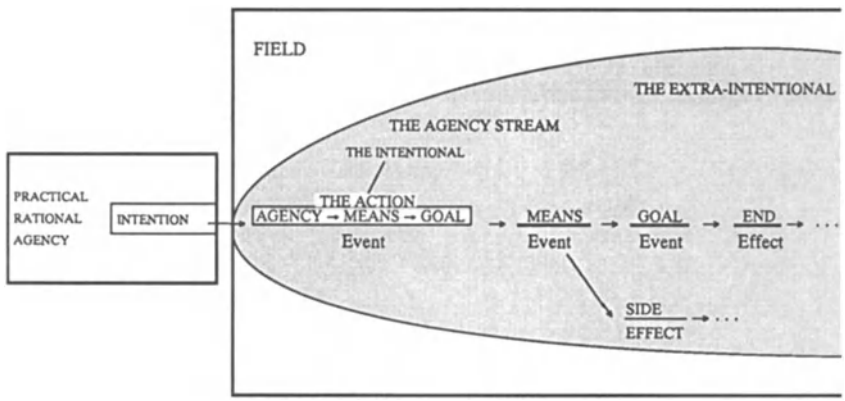
(1) INTENTION  	(3) COMMENTARY Language which describes an object from square (2) in terms of the logical space of action: i.e., in terms which are relevant to explanation and appraisal of agency.
(2) WHAT HAPPENED A representation of the agency stream — the event of agency and all relevant downstream events — together with all relevant field events.	(4) LOGICAL SPACE OF ACTION A representation of the object or property described in square (3).

2.3 What Happened (Square 2)

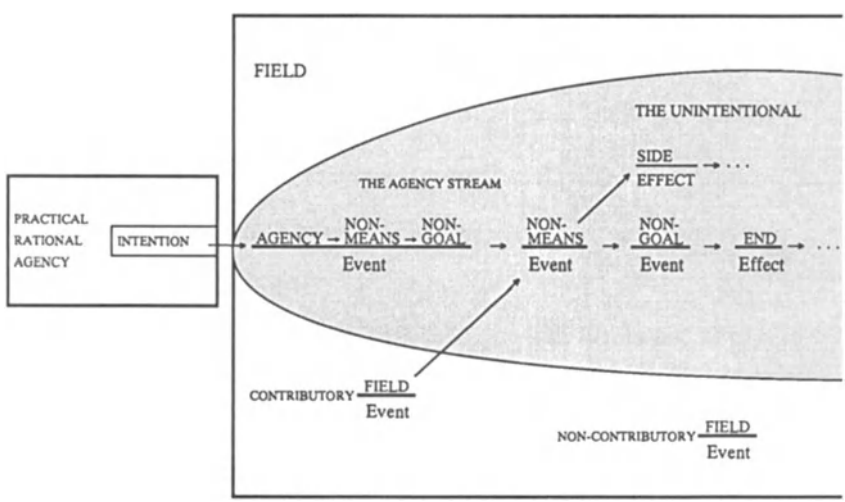
COMMENTARY: The system prompts the user for what actually happened in the world as a result of the agent's efficacious intention.

THE LOGICAL SPACES OF ACTION

THE LOGICAL SPACE OF INTENTIONALITY



THE LOGICAL SPACE OF UNINTENTIONALITY



Thus the user is required to identify the actual event of agency caused. Depending on the nature of the case, the agency stream may contain the *intentional*, which is the intended goal-functionality of the event of agency, and the *extra-intentional* which consists of everything outside or causally downstream of the intentional which was not intended. Or, the agency stream may be *unintentional*, namely, when the intention fails to cause the intentional. In either case, the agency stream will include an area called 'The Domain of Negligence' whose members will be either extra-intentional or unintentional and for which the agent may be liable. The user will also identify any relevant field events, such as a causally contributory field event which enters the agency stream and frustrates the causing of the goal in cases of accident.

GLOSSARY

AGENCY STREAM: the chain of events which is causally downstream of an EFFICACIOUS INTENTION. Events not in the AGENCY STREAM are FIELD EVENTS.

END-EFFECT: an event which is CONSEQUENCE of a GOAL EVENT and is therefore EXTRA-INTENTIONAL. See also SIDE-EFFECT.

FIELD EVENT: a event which is not part of an AGENCY STREAM. Field events are of two sorts: those which contribute causally to the agency stream and those which do not. Of those field events which do contribute causally to the agency stream, some will be goal-frustrating, some will be goal-contributing, and others will be ineffectual in this respect.

SIDE-EFFECT: an event which is an unintended collateral consequence of an EVENT OF AGENCY, a MEANS EVENT, or a GOAL EVENT, and therefore EXTRA-INTENTIONAL whether represented (foreseen) or not.

(10) Did an event of agency occur?

COMMENTARY: Since the user has answered 'Yes' to the question 'Was this a case of rational agency?', then there was an event of agency. It remains to be determined whether it was the INTENTIONAL EVENT OF AGENCY, i.e. whether it was the proximate effect of the intention and whether it had the properties and relations intended of it.

(11) Did the goal event occur?

COMMENTARY: In this example, did the event of the Tirpitz sinking occur?

(12) ANSWER: Yes

(13) Did the goal event occur in the way it was intended?

COMMENTARY: I.e., did the intended cause of the goal event cause the goal event? In this example, did the firing of the torpedo cause the Tirpitz to sink?

(14) ANSWER: Yes

(15) Did the means event occur in the way it was intended?

COMMENTARY: I.e., did the intended cause of the means event cause the means event? In this example, did the pushing of the firing button cause the torpedo to be fired? (*NOTE*: This question will recur until an event of agency is identified.)

(16) ANSWER: Yes

(17) Did anything else of relevance occur in the agency stream? That is, any other significant events or properties?

(18) ANSWER: No

COMMENTARY: Answers to the above questions indicate that What Happened was:

$$\underset{E_3}{Ag+H} \rightarrow \underset{E_2}{G} \rightarrow \underset{E_1}{F}$$

The system places this representation of What Happened in square 2 of the Square of (Un)Intentionality for this case:

SQUARE OF (UN)INTENTIONALITY FOR THE OFFICER'S SINKING THE TIRPITZ

(1) INTENTION Ag+H ww → $\frac{G}{E_2}$ ww → $\frac{F}{E_1}$ Ev+ → E_3	(3) COMMENTARY Language which describes an object from square (2) in terms of the logical space of action: i.e., in terms which are relevant to explanation andappraisal of agency.
(2) WHAT HAPPENED $\frac{Ag+H}{E_3}$ → $\frac{G}{E_2}$ → $\frac{F}{E_1}$	(4) LOGICAL SPACE OF ACTION A representation of the object or property described in square (3).

2.4 The Ascription and the Ascribed (Squares 3 and 4)

COMMENTARY: The system will now complete the Square of (Un)Intentionality for this case. This will require that it choose, on the basis of the relationship between the content of squares 1 and 2, an object ascribable to the agent which will appear in square 4 and some appropriate action discourse for square 3 which refers to that object and ascribes it to the agent. The system will also indicate the intentional category of each element of What Happened. That is, square 4 may be thought of as an *intentional filter* for the contents of square 2.

(19) What explanatory object is ascribable to the agent?

The following is a completed Object Frame for the objects in this case:

PROPERTIES	EVENTS					
	E_1	E_2	E_3	E	E	E
INTENDED PRACTICAL PROPERTIES						
[AGENCY] [MEANS] [GOAL]						
AGENCY PROPERTY _____	___	___	H	___	___	___
MEANS PROPERTY _____	___	G	___	___	___	___
GOAL PROPERTY _____	F	___	___	___	___	___
ACTUAL PRACTICAL AND NON-PRACTICAL PROPERTIES						
EVENT OF AGENCY						
Causes _____	___	___	E_2	___	___	___
Is caused by _____	___	___	I	___	___	___
MEANS EVENT _____	___	___	___	___	___	___
Causes _____	___	E_1	___	___	___	___
Is caused by _____	___	E_3	___	___	___	___
GOAL EVENT _____	___	___	___	___	___	___
Causes _____	?	___	___	___	___	___
Is caused by _____	E_2	___	___	___	___	___
CONSEQUENCE _____	___	___	___	___	___	___
Causes _____	___	___	___	___	___	___
Is caused by _____	___	___	___	___	___	___
CONJOINING CAUSE _____	___	___	___	___	___	___
Causes _____	___	___	___	___	___	___
Is caused by _____	___	___	___	___	___	___
NON-CONJOINING _____	___	___	___	___	___	___
Causes _____	___	___	___	___	___	___
Is caused by _____	___	___	___	___	___	___

HELP

The practical explanatory objects from among which the system will choose are:

THE INTENTIONAL OBJECT: An object is intentional (an ACTION) *iff* it is the effect of an efficacious intention which is fully explained by the content of that intention, which it is just in case it has all and only those properties it was intended to have.

The Syntax of ACTION:

$$\frac{[\text{Ag}]\ [\text{MEANS}]\ [\text{GOAL}]\ [\text{Ev}^{+}]}{[\text{E}_x]}$$

UNINTENTIONAL OBJECTS: An object is unintentional *iff* it is the effect of an efficacious intention which is only partially explained by the content of that intention — the intention explains why that E_x was an Ag/E_x — and is only fully explained with the addition of an explanation of those unintended relations which it has instead of those it was intended to have. It is the function of the unintentional objects to explain the unintended relations of an Ag/E_x.

The Syntax of MISTAKE:

$$\frac{[\text{AGENCY}]\ [\text{NON-MEANS CAUSAL RELATION}]\ [\text{NON-GOAL CONSEQUENCE}]}{[\text{E}_x]}$$

The Syntax of ACCIDENT:

$$\frac{[\text{AGENCY}]\ \begin{bmatrix}\text{NON-MEANS}\\ \text{CAUSAL}\\ \text{RELATION}\end{bmatrix} + \begin{bmatrix}\text{UNFORESEEN}\\ \text{INTERVENING}\\ \text{CAUSE}\end{bmatrix}\ \begin{bmatrix}\text{NON-GOAL}\\ \text{CONSEQUENCE}\end{bmatrix}}{\text{E}_x}$$

THE EXTRA-INTENTIONAL: A property or object is Extra-intentional if it is causally downstream of an efficacious intention but was not intended, whether it was foreseen or not. INADVERTENCE is an explanatory object ascribed to an agent which contains essentially an extra-intentional property or object.

The Syntax of INADVERTENCE:

$$\frac{[\text{Ag}]\quad[\text{MEANS}]\quad[\text{GOAL}]\quad\left[\begin{array}{l}\text{EXTRA-INTENTIONAL}\\ \text{SIDE/END EFFECT}\end{array}\right]}{[E_x]}$$

For an object to be of one of the above types it must have all and only those properties the set of which is definitive of that type.

In ACTION *simpliciter*, there must be an intentional event of agency, Ag/E_x, which therefore has all the means and goal relations which were intended of it.

For all *explanatory objects* x, x is an action *iff* there is some y such that y is an intentional event of agency — i.e., y has the intended goal-functionality — and x is that intended goal-functionality of y.

In MISTAKE, there must be an unintentional event of agency with that unintended relation which it has instead of an intended means relation and which is explained by the falsity of the corresponding means-belief. This unintentional relation is definitive of mistake and makes the entire MISTAKE object unintentional. It is the function of a mistake to explain this type of unintentional relation of an event of agency.

For all *explanatory objects* y, y is a mistake *iff* there is some z such that z is an unintentional event of agency and:

(1) z lacks its intended goal-functionality,
(2) z has an unintended causal relation to some consequence due to a false belief about the causal properties of the event of agency, and
(3) y is that unintended functionality of z.

In ACCIDENT, there must be an unintentional event of agency with that unintended relation which it has instead of the intended means relation and which is explained by the unforeseen causal intervention of a causally contributory field event which has made the corresponding means belief false. This (unintentional) relational intervention is definitive of accident and makes the entire ACCIDENT object unintentional. It is the function of accident to explain this type of unintentional relation of an event of agency.

For all *explanatory objects x*, *x* is an accident *iff* there is some *y* such that *y* is an unintentional event of agency and

(1) *y* lacks its intended goal-functionality,
(2) *y* has an unintended causal relation to some consequence due to the causal interference of a conjoining field event, and
(3) *x* is that unintended functionality of *y*.

In INADVERTENCE, there must be an intentional event of agency with, in addition to what was intended, an extra-intentional property or relation to a side-effect or an end-effect. This extra-intentional addition is definitive of inadvertence and makes the entire INADVERTENCE object extra-intentional. It is the function of inadvertence to explain an extra-intentional property of an event of agency.

For all *explanatory objects z*, *z* is an inadvertence *iff* there is some *x* such that *x* is an intentional event of agency and

(1) *x* has its intended goal-functionality,
(2) *x* has an extra-intentional property or causal relation to an extra-intentional side- or end-effect, and
(3) *z* is the extra-intentionality of *x*.

COMMENTARY: The system displays and compares what was intended to happen and what actually happened (square 2).

WHAT WAS INTENDED:

$$\underset{E_3}{Ag+H} \longrightarrow \underset{E_2}{G} \longrightarrow \underset{E_1}{F}$$

WHAT HAPPENED:

$$\underset{E_3}{Ag+H} \longrightarrow \underset{E_2}{G} \longrightarrow \underset{E_1}{F}$$

(20) ANSWER: There is matching! Therefore, an ACTION is ascribable to the agent.

SQUARE OF INTENTIONALITY FOR THE OFFICER'S SINKING THE TIRPITZ

(1) INTENTION $\text{Ag+H ww} \rightarrow \underset{E_2}{G}\ \text{ww} \rightarrow \underset{E_1}{F}\ \text{Ev}^+$ $\xrightarrow[E_3]{}$	(3) ASCRIPTION The officer sank the Tirpitz.
(2) WHAT HAPPENED $\underset{E_3}{\text{Ag+H}} \rightarrow \underset{E_2}{G} \rightarrow \underset{E_1}{F}$	(4) THE ACTION $\underset{E_3}{\text{Ag+H}\ \underset{E_2}{G} \rightarrow \underset{E_1}{F}\ \text{Ev}^+}$

3. MISTAKE

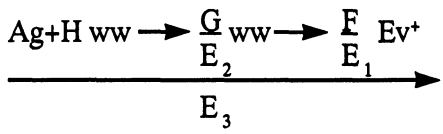
A naval officer aims and fires a torpedo at a ship he thinks is the Tirpitz and sinks the Bismark. Does he intentionally sink that ship?

3.1 Intention (Square 1)

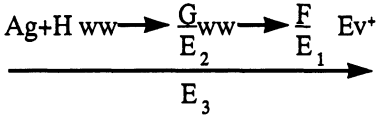
Suppose, for demonstration purposes, that Part 1 of the case analysis, in which the intention with which the agent acted is constructed, reveals the following about that intention:

The agent's goal event is F/E_1: the event of the Tirpitz sinking.
The agent's means event is G/E_2: the event of the torpedo firing.
The agent's event of agency is $Ag+H/E_3$: the event, proximately caused by his intention, of his finger moving thus and so against the firing button.

If so, the agent's intention was:



SQUARE OF (UN)INTENTIONALITY FOR THE OFFICER'S SINKING THE BISMARK

(1) INTENTION  	(3) COMMENTARY Language which describes an object from square (2) in terms of the logical space of action: i.e., in terms which are relevant to explanation and appraisal of agency.
(2) WHAT HAPPENED A representation of the agency stream — the event of agency and all relevant downstream events — together with all relevant field events.	(4) LOGICAL SPACE OF ACTION A representation of the object or property described in square (3).

3.2 What Happened (Square 2)

COMMENTARY: The system prompts the user for what actually happened in the world as a result of the agent's efficacious intention.

Thus the user is required to identify the actual event of agency caused for the case. Depending on the nature of the case, the agency stream may contain the *intentional*, which is the intended goal-functionality of the event of agency, and the *extra-intentional* which consists of everything causally downstream of the intention which was not intended. Or, the agency stream may be *unintentional*, namely, when

the intention fails to cause the intentional. In either case, the agency stream will include an area called 'The Domain of Negligence' whose members will be either extra-intentional or unintentional and for which the agent may be liable. The user will also identify any relevant field events, such as a causally contributory field event which enters the agency stream and frustrates the causing of the goal in cases of accident.

(1) Did an event of agency occur?

COMMENTARY: Since the user has answered 'Yes' to the question 'Was this a case of rational agency?', then there was an event of agency. It remains to be determined whether it was the INTENTIONAL EVENT OF AGENCY, i.e. whether it was the proximate effect of the intention and whether it had the properties and relations intended of it.

(2) Did the goal event occur?

COMMENTARY: In this example, did the event, E_1, of the Tirpitz sinking, F, occur?

(3) ANSWER: No

COMMENTARY: If the agent's goal event did not occur, then neither did the event of agency he intended occur, nor any of the intended means events. The agency stream caused will therefore be of the category *unintentional*. The system will now prompt the user to identify what did occur in the unintentional agency stream in place of what was intended. (The Unintentionality Question Sequence)

(4) What non-goal event occurred instead of the agent's goal event?

(5) ANSWER: The event which occurred instead of the goal event was:

the event of the sinking of the Bismark

COMMENTARY: The system assigns an event-name to this non-goal event and a property-name to its non-goal property:

Assign 'E_4' to the event.
Assign 'B' to the property: sinking of the Bismark.
Assign 'B/E_4' to the event, E_4, of the sinking of the Bismark, B.

The system can also assign a new event-name to the actual event of agency which was not, *ex hypothesi*, the event of agency intended.

Assign 'Ag/E_5' to the event, E_5, of agency, Ag.

These values are placed in the appropriate places in square 2:

WHAT HAPPENED:

$$\underset{E_5}{Ag} \longrightarrow \cdots \longrightarrow \underset{E_4}{B}$$

COMMENTARY: The user has indicated that a chain of events from an event of his agency to a non-goal event has occurred. It remains to identify the precise nature of that chain.

(6) Identify the agency stream: what events, if any, occurred between the event of agency and the non-goal event which explain that non-goal-functional property of the event of agency? Identify any non-means event(s) which occurred instead of the intended means event(s) and any relevant field event(s). Categorize the events as of one or more of the following types:

(a) the proximate effect of the event of agency,
(b) the proximate cause of the non-goal event,
(c) the proximate effect of the proximate effect of the event of agency, etc.,
(d) the proximate cause of the proximate cause of the goal event, etc.,
(e) a field event which conjoined with an event of type (a) — (d) to cause an event of type (a) — (d) or the non-goal event.

(7) ANSWER: The non-means event:

the torpedo firing at the Bismark

occurred and was:

(a) the proximate effect of the event of agency,
(b) the proximate cause of the non-goal event.

COMMENTARY: The system assigns an event-name to this non-means event and a property-name to its property:

Assign 'E_6' to the event.
Assign 'G' to the property: torpedo firing. ('G' has already been assigned to this property)
Assign 'G/E_6' to the event, E_6, of the torpedo firing, G.

These values are placed in the appropriate places in square 2:

WHAT HAPPENED:

$$\underset{E_5}{\underline{Ag}} \rightarrow \underset{E_6}{\underline{G}} \rightarrow \underset{E_4}{\underline{B}}$$

COMMENTARY: The above answer has completed the chain. It remains to complete the description of the event of agency.

(8) What type of body event was the event of agency Ag/E_5.

(9) ANSWER: The event of agency was

the agent's finger moving against the firing button

COMMENTARY: The system has assigned the property name 'H' to the property 'the agent's finger moving against the firing button.'

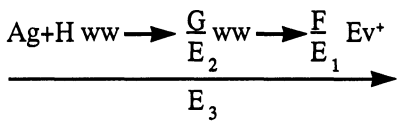
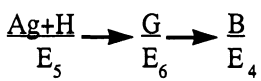
Assign 'H' to the event, E_5, of agency, Ag.

These values are placed in the appropriate places in square 2:

WHAT HAPPENED:

$$\underset{E_5}{\underline{Ag+H}} \rightarrow \underset{E_6}{\underline{G}} \rightarrow \underset{E_4}{\underline{B}}$$

SQUARE OF UNINTENTIONALITY FOR THE OFFICER'S SINKING THE BISMARK

(1) INTENTION  	(3) COMMENTARY Language which describes an object from square (2) in terms of the logical space of action: i.e., in terms which are relevant to explanation and appraisal of agency.
(2) WHAT HAPPENED  	(4) LOGICAL SPACE OF ACTION A representation of the object or property described in square (3).

3.3 The Ascription and the Ascribed (Squares 3 and 4)

COMMENTARY: The system will now complete the Square of (Un)Intentionality for this case. This will require that it choose, on the basis of the relationship between the content of squares 1 and 2, an object ascribable to the agent which will appear in square 4 and some appropriate action discourse for square 3 which refers to that object and ascribes it to the agent. The system will also indicate the intentional category of each element of What Happened. That is, square 4 may be thought of as an *intentional filter* for the contents of square 2.

The following is a completed Object Frame for the objects in this case:

PROPERTIES	EVENTS					
	E_1	E_2	E_3	E_4	E_5	E_6
INTENDED PRACTICAL PROPERTIES						
[AGENCY] [MEANS] [GOAL]						
AGENCY PROPERTY			Yes			
MEANS PROPERTY		Yes				
GOAL PROPERTY	F					
ACTUAL PRACTICAL AND NON-PRACTICAL PROPERTIES						
EVENT OF AGENCY					H	
Causes					E_6	
Is caused by					I	
MEANS EVENT						
Causes						
Is caused by						
GOAL EVENT						
Causes						
Is caused by						
CONSEQUENCE				B		G
Causes				?		E_4
Is caused by				E_6		E_5
CONJOINING CAUSE						
Causes						
Is caused by						
NON-CONJOINING						
Causes						
Is caused by						
NON-ACTUAL EVENT	Yes	Yes	Yes			

HELP

The practical explanatory objects from among which the system will choose are:

The Syntax of MISTAKE:

$$\frac{[\text{AGENCY}]\ [\text{NON-MEANS CAUSAL RELATION}]\ [\text{NON-GOAL CONSEQUENCE}]}{[E_x]}$$

The Syntax of ACCIDENT:

$$\frac{[\text{AGENCY}]\ \left[\begin{array}{c}\text{NON-MEANS}\\ \text{CAUSAL}\\ \text{RELATION}\end{array}\right] + \left[\begin{array}{c}\text{UNFORESEEN}\\ \text{INTERVENING}\\ \text{CAUSE}\end{array}\right]\ \left[\begin{array}{c}\text{NON-GOAL}\\ \text{CONSEQUENCE}\end{array}\right]}{E_x}$$

The Syntax of INADVERTENCE:

$$\frac{[\text{Ag}]\ [\text{MEANS}]\ [\text{GOAL}]\ \left[\begin{array}{c}\text{EXTRA-INTENTIONAL}\\ \text{SIDE/END EFFECT}\end{array}\right]}{[E_x]}$$

WHAT WAS INTENDED:

$$\frac{\text{Ag+H}}{E_3} \longrightarrow \frac{\text{G}}{E_2} \longrightarrow \frac{\text{F}}{E_1}$$

WHAT HAPPENED:

$$\frac{\text{Ag+H}}{E_5} \longrightarrow \frac{\text{G}}{E_6} \longrightarrow \frac{\text{B}}{E_4}$$

COMMENTARY: There is *mismatch* between what was intended and what happened. The mismatch is due either to:

(a) the failure of the event of agency to have the goal-functionality which it was falsely believed and therefore mis-intended to have,

(b) the unforeseen causal intervention of a goal-frustrating field event,
(c) the addition to the intentional of an extra-intentional property or side or end effect

In this case, the mismatch is of the form:
(a) the failure of the event of agency to have the goal-functionality which it was falsely believed and therefore mis-intended to have.

(10) Mistake is ascribable:

$$\frac{\text{[AGENCY] [NON-MEANS CAUSAL RELATION] [NON-GOAL CONSEQUENCE]}}{[E_x]}$$

SQUARE OF UNINTENTIONALITY FOR THE OFFICER'S MISTAKENLY SINKING THE BISMARK

(1) INTENTION $\text{Ag+H ww} \rightarrow \underset{E_2}{G}\ \text{ww} \rightarrow \underset{E_1}{F}\ Ev^{+}$ $\xrightarrow[E_3]{}$	(3) COMMENTARY The officer mistakenly sank the *Bismark*, by firing a torpedo at the ship he mistook for the *Tirpitz*, by moving his finger against the firing button.
(2) WHAT HAPPENED $\underset{E_5}{\text{Ag+H}} \rightarrow \underset{E_6}{G} \rightarrow \underset{E_4}{B}$	(4) THE MISTAKE $\dfrac{\text{Ag+H} \xrightarrow[E_6]{G} \underset{E_4}{B}}{E_5}$

4. ACCIDENT

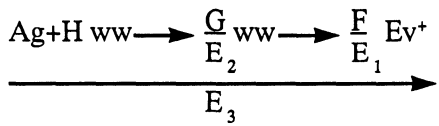
> You have a donkey, so have I, and they graze in the same field. The day comes when I conceive a dislike for mine. I go to shoot it, draw a bead on it, fire ... but as I do so, the beasts move, and to my horror yours falls.[3]

4.1 Intention (Square 1)

Suppose, for demonstration purposes, that Part 1 of the case analysis, in which the intention with which the agent acted is constructed, reveals the following about that intention:

> The agent's goal event was F/E_1: the event, E_1, of the death of his donkey, F.
>
> The agent's means event was G/E_2: the event, E_2, of the firing of his carefully aimed gun, G.
>
> The agent's event of agency was $Ag+H/E_3$: the event, E_3, of agent's finger moving thus and so, H, against the trigger of his carefully aimed gun.

If so, the agent's intention was:



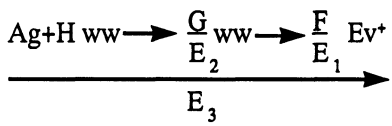
4.2 What Happened (Square 2)

COMMENTARY: The system prompts the user for what actually happened in the world as a result of the agent's efficacious intention. See *The Logical Spaces of Action* diagrams above.

Thus the user is required to identify the actual event of agency caused for the case. Depending on the nature of the case, the agency stream may contain the *intentional*, which is the intended goal-functionality of the event of agency, and the *extra-intentional* which consists of everything causally downstream of the intention which was not intended. Or, the agency stream may be *unintentional*, namely, when

the intention fails to cause the intentional. In either case, the agency stream will include an area called 'The Domain of Negligence' whose members will be either extra-intentional or unintentional and for which the agent may be liable. The user will also identify any relevant field events, such as a causally contributory field event which enters the agency stream and frustrates the causing of the goal in cases of accident.

SQUARE OF (UN)INTENTIONALITY FOR MY SHOOTING THE NEIGHBOUR'S DONKEY

(1) INTENTION	(3) COMMENTARY
	Language which describes an object from square (2) in terms of the logical space of action: i.e., in terms which are relevant to explanation and appraisal of agency.
(2) WHAT HAPPENED A representation of the agency stream — the event of agency and all relevant downstream events — together with all relevant field events.	(4) LOGICAL SPACE OF ACTION A representation of the object or property described in square (3).

(1) Did an event of agency occur?

COMMENTARY: Since the user has answered 'Yes' to the question 'Was this a case of rational agency?', then there was an event of agency. It remains to be determined whether it was the INTENTIONAL EVENT OF AGENCY, i.e. whether it was the proximate effect of the intention and whether it had the properties and relations intended of it.

(2) Did the goal event occur?

COMMENTARY: In this example, did the event, E_1, of the death of the agent's donkey, F, occur?

(3) ANSWER: No

COMMENTARY: If the agent's goal event did not occur, then neither did the event of agency he intended occur, nor any of the intended means events. The agency stream caused will therefore be of the category *unintentional*. The system will now prompt the user to identify what did occur in the unintentional agency stream in place of what was intended. (The Unintentionality Question Sequence)

(4) What non-goal event occurred instead of the agent's goal event?

(5) ANSWER: The event which occurred instead of the goal event was:

the event of the death of his neighbour's donkey

COMMENTARY: The system assigns an event-name to this non-goal-event and a property-name to its non-goal-property:

Assign 'E_4' to the event.

Assign 'D' to the property: death of the neighbour's donkey.

Assign 'D/E_4' to the event, E_4, of the death of the neighbour's donkey.

The system also assigns a new event-name to the actual event of agency which was not, *ex hypothesi*, the event of agency intended.

Assign 'Ag/E_5' to the event, E_5, of agency, Ag.

These values are placed in the appropriate places in square 2:

WHAT HAPPENED:

$$\underset{E_5}{Ag} \longrightarrow \cdots \longrightarrow \underset{E_4}{D}$$

COMMENTARY: The user has indicated that a chain of events from an event of his agency to a non-goal event has occurred. It remains to identify the precise nature of that chain.

(6) Identify the agency stream: what events, if any, occurred between the event of agency and the non-goal event which explain that non-goal-functional property of the event of agency. Identify any non-means event(s) which occurred instead of the intended means event(s) and any relevant field event(s). Categorize the events as of one or more of the following types:

(a) the proximate effect of the event of agency,

(b) the proximate cause of the non-goal event,

(c) the proximate effect of the proximate effect of the event of agency, etc.,

(d) the proximate cause of the proximate cause of the goal event, etc.,

(e) a field event which conjoined with an event of type (a) — (d) to cause an event of type (a) — (d) or the non-goal event.

(7) ANSWER: The event:

the firing of the agent's gun

occurred and was

(a) the proximate effect of the event of agency

The event:

the beasts moving

occurred and was

(e) a field event which conjoined with an event of type (a) to cause the non-goal event.

COMMENTARY: The system assigns event-names to these events and property-names to their properties:

Assign 'E_6' to the proximate effect of the event of agency.
Assign 'G' to the property: gun firing. ('G' has already been assigned to this property)

Assign 'G/E_6' to the event, E_6, of the gun firing, G.
Assign 'E_7' to the field event of the beasts moving.
Assign 'M' to the property: beasts moving
Assign 'M/E_7' to the event, E_7, of the beasts moving.

These values are placed in the appropriate places in square 2:

WHAT HAPPENED:

$$\underset{E_5}{Ag} \longrightarrow \underset{E_6}{G} + \underset{E_7}{M} \longrightarrow \underset{E_4}{D}$$

COMMENTARY: The above answer has completed the chain. It remains to complete the description of the event of agency.

(8) What type of body event was the event of agency Ag/E_5.

(9) ANSWER: The event of agency was

the agent's finger moving against the trigger of his gun

COMMENTARY: The system has assigned the property name 'H' to the property 'agent's finger moving against the trigger of his gun.'

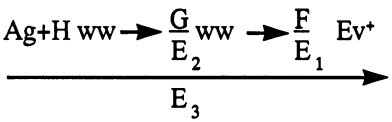
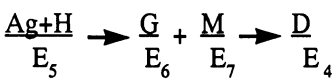
Assign 'H' to the event, E_5, of agency, Ag.

These values are placed in the appropriate places in square 2:

WHAT HAPPENED:

$$\underset{E_5}{Ag+H} \longrightarrow \underset{E_6}{G} + \underset{E_7}{M} \longrightarrow \underset{E_4}{D}$$

SQUARE OF UNINTENTIONALITY FOR MY SHOOTING THE NEIGHBOUR'S DONKEY

(1) INTENTION	(3) COMMENTARY
	Language which describes an object from square (2) in terms of the logical space of action: i.e., in terms which are relevant to explanation and appraisal of agency.
(2) WHAT HAPPENED	(4) LOGICAL SPACE OF ACTION
	A representation of the object or property described in square (3).

4.3 The Ascription and the Ascribed (Squares 3 and 4)

COMMENTARY: The system will now complete the Square of (Un)Intentionality for this case. This will require that it choose, on the basis of the relationship between the content of squares 1 and 2, an object ascribable to the agent which will appear in square 4 and some appropriate action discourse for square 3 which refers to that object and ascribes it to the agent. The system will also indicate the intentional category of each element of What Happened. That is, square 4 may be thought of as an *intentional filter* for the contents of square 2.

The following is a completed Object Frame for the objects in this case:

PROPERTIES	EVENTS						
	E_1	E_2	E_3	E_4	E_5	E_6	E_7
INTENDED PRACTICAL PROPERTIES							
[AGENCY] [MEANS] [GOAL]							
AGENCY PROPERTY			Yes				
MEANS PROPERTY		Yes					
GOAL PROPERTY	F						
ACTUAL PRACTICAL AND NON-PRACTICAL PROPERTIES							
EVENT OF AGENCY					H		
Causes					E_6		
Is caused by					I		
MEANS EVENT							
Causes							
Is caused by							
GOAL EVENT							
Causes							
Is caused by							
CONSEQUENCE				D		G	
Causes				?		E_4	
Is caused by				E_6E_7		E_5	
CONJOINING CAUSE							M
Causes							E_4
Is caused by							?
NON-CONJOINING							
Causes							
Is caused by							
NON-ACTUAL EVENT	Yes	Yes	Yes				

HELP

The practical explanatory objects from among which the system will choose are:

The Syntax of MISTAKE:

$$\frac{\text{[AGENCY] [NON-MEANS CAUSAL RELATION] [NON-GOAL CONSEQUENCE]}}{[E_x]}$$

The Syntax of ACCIDENT:

$$\frac{\text{[AGENCY]} \quad \left[\begin{array}{c}\text{NON-MEANS}\\ \text{CAUSAL}\\ \text{RELATION}\end{array}\right] \quad + \quad \left[\begin{array}{c}\text{UNFORESEEN}\\ \text{INTERVENING}\\ \text{CAUSE}\end{array}\right] \quad \left[\begin{array}{c}\text{NON-GOAL}\\ \text{CONSEQUENCE}\end{array}\right]}{E_x}$$

The Syntax of INADVERTENCE:

$$\frac{\text{[Ag]} \quad \text{[MEANS]} \quad \text{[GOAL]} \quad \left[\begin{array}{c}\text{EXTRA-INTENTIONAL}\\ \text{SIDE/END EFFECT}\end{array}\right]}{[E_x]}$$

WHAT WAS INTENDED:

$$\underset{E_3}{\text{Ag+H}} \longrightarrow \underset{E_2}{\text{G}} \longrightarrow \underset{E_1}{\text{F}}$$

WHAT HAPPENED:

$$\underset{E_5}{\text{Ag+H}} \longrightarrow \underset{E_6}{\text{G}} + \underset{E_7}{\text{M}} \longrightarrow \underset{E_4}{\text{D}}$$

COMMENTARY: There is *mismatch* between what was intended and what happened. The mismatch is due either to:

(a) the failure of the event of agency to have the goal-functionality which it was falsely believed and therefore mis-intended to have,

(b) the unforeseen causal intervention of a goal-frustrating field event,

(c) the addition to the intentional of an extra-intentional property or side or end effect

In this case, the mismatch is of the form:

(b) the unforeseen causal intervention of a goal-frustrating field event.

(10) Accident is ascribable:

$$\frac{[\text{AGENCY}] \quad [\text{NON-MEANS CAUSAL RELATION}] \; + \; [\text{UNFORESEEN INTERVENING CAUSE}] \quad [\text{NON-GOAL CONSEQUENCE}]}{E_x}$$

SQUARE OF UNINTENTIONALITY FOR MY ACCIDENTALLY SHOOTING THE NEIGHBOUR'S DONKEY

(1) INTENTION $\text{Ag+H ww} \rightarrow \underset{E_2}{\underline{G}}\text{ww} \rightarrow \underset{E_1}{\underline{F}}\ Ev^+$ $\xrightarrow[E_3]{}$	(3) COMMENTARY I accidentally shot my neighbour's donkey.
(2) WHAT HAPPENED $\underset{E_5}{\underline{\text{Ag+H}}} \rightarrow \underset{E_6}{\underline{G}} + \underset{E_7}{\underline{M}} \rightarrow \underset{E_4}{\underline{D}}$	(4) THE ACCIDENT $\dfrac{\text{Ag+H} \rightarrow \underset{E_6}{\underline{G}} + \underset{E_7}{\underline{M}} \rightarrow \underset{E_4}{\underline{D}}}{E_5}$

5. INADVERTENCE

You have a donkey, so have I, and they graze in the same field. The day comes when I conceive a dislike for mine. I go to shoot it, draw a bead on it, fire: the brute falls in its tracks. Unfortunately, those tracks run through the garden and his fall crushes the cabbages.

5.1 Intention (Square 1)

Suppose, for demonstration purposes, that Part 1 of the case analysis, in which the intention with which the agent acted is constructed, reveals the following about that intention:

> The agent's goal event was F/E_1: the event, E_1, of the death of his donkey, F.
> The agent's means event was G/E_2: the event, E_2, of the firing of his carefully aimed gun, G.
> The agent's event of agency was $Ag+H/E_3$: the event, E_3, of agent's finger moving thus and so, H, against the trigger of his carefully aimed gun.

If so, the agent's intention was:

$$\text{Ag+ H ww} \longrightarrow \underset{E_2}{G}\ \text{ww} \longrightarrow \underset{E_1}{F}\ \ Ev^{+}$$
$$\xrightarrow[E_3]{\qquad\qquad\qquad\qquad}$$

SQUARE OF (UN)INTENTIONALITY FOR MY CRUSHING THE CABBAGES

(1) INTENTION $\text{Ag+H ww} \longrightarrow \underset{E_2}{G}\ \text{ww} \longrightarrow \underset{E_1}{F}\ \text{Ev+}$ $\xrightarrow[E_3]{\qquad\qquad\qquad}$	(3) COMMENTARY Language which describes an object from square (2) in terms of the logical space of action: i.e., in terms which are relevant to explanation and appraisal of agency.
(2) WHAT HAPPENED A representation of the agency stream — the event of agency and all relevant downstream events — together with all relevant field events.	(4) LOGICAL SPACE OF ACTION A representation of the object or property described in square (3).

5.2 What Happened (Square 2)

COMMENTARY: The system prompts the user for what actually happened in the world as a result of the agent's efficacious intention. See *The Logical Spaces of Action* diagrams above.

Thus the user is required to identify the actual event of agency caused. Depending on the nature of the case, the agency stream may contain the *intentional*, which is the intended goal-functionality of the event of agency, and the *extra-intentional* which consists of everything outside or causally downstream of the intentional which was not intended. Or, the agency stream may be *unintentional*, namely, when the intention fails to cause the intentional. In either case, the agency stream will include an area called 'The Domain of Negligence' whose members will be either extra-intentional or unintentional and for which the agent may be liable. The user will also identify any relevant field events, such as a causally contributory field event which enters the agency stream and frustrates the causing of the goal in cases of accident.

(1) Did an event of agency occur?

COMMENTARY: Since the user has answered 'Yes' to the question 'Was this a case of rational agency?', then there was an event of agency. It remains to be determined whether it was the INTENTIONAL EVENT OF AGENCY, i.e. whether it was the proximate effect of the intention and whether it had the properties and relations intended of it.

(2) Did the goal event occur?

COMMENTARY: In this example, did the event, E_1, of the death of the agent's donkey, F, occur?

(3) ANSWER: Yes

(4) Did the goal event occur in the way it was intended?

COMMENTARY: I.e., did the intended cause of the goal event cause the goal event? In this example, did the, E_2, firing of the gun cause the death of the donkey.

(5) ANSWER: Yes

(6) Did the means event occur in the way it was intended?

COMMENTARY: I.e., did the intended cause of the means event cause the means event? In this example, did the event E_3 of the moving of the finger against the trigger of the gun, H, cause the means event, E_2, of the gun firing, G?

(7) ANSWER: Yes

(8) Did any significant extra-intentional properties or events occur in the agency stream which may be relevant to the agent's appraisal? Select from the following menu:

(a) the event which was the side-effect of the event of agency
(b) the event which was the side-effect of a means event
(c) the event which was the end-effect of the goal event
(d) the extra-intentional property of an agency stream event

(9) ANSWER: The event

the crushing of the cabbages

occurred and was: (c) the event which was the end effect of the goal event.

COMMENTARY: The system assigns an event-name to this non-goal event and a property-name to its non-goal-property:

Assign 'E_4' to the event.
Assign 'C' to the property: crushing the cabbages
Assign 'C/E_4' to the event, E_4, of the crushing of the cabbages.

COMMENTARY: Answers to the above questions indicate that What Happened was:

$$\frac{Ag+H}{E_3} \longrightarrow \frac{G}{E_2} \longrightarrow \frac{F}{E_1} \longrightarrow \frac{C}{E_4}$$

The system places this representation of What Happened in square 2 of the Square of (Un)Intentionality for this case:

SQUARE OF (UN)INTENTIONALITY FOR MY CRUSHING THE CABBAGES

(1) INTENTION Ag+H ww → $\underset{E_2}{G}$ ww → $\underset{E_1}{F}$ Ev+ E_3	(3) COMMENTARY Language which describes an object from square (2) in terms of the logical space of action: i.e., in terms which are relevant to explanation and appraisal of agency.
(2) WHAT HAPPENED $\underset{E_3}{Ag+H}$ → $\underset{E_2}{G}$ → $\underset{E_1}{F}$ → $\underset{E_4}{C}$	(4) LOGICAL SPACE OF ACTION A representation of the object or property described in square (3).

5.3 The Ascription and the Ascribed (Squares 3 and 4)
COMMENTARY: The system will now complete the Square of (Un)Intentionality for this case. This will require that it choose, on the basis of the relationship between the content of squares 1 and 2, an object ascribable to the agent which will appear in square 4 and some appropriate action discourse for square 3 which refers to that object and ascribes it to the agent. The system will also indicate the intentional category of each element of What Happened. That is, square 4 may be thought of as an *intentional filter* for the contents of square 2.

The following is a completed Object Frame for the objects in this case:

PROPERTIES	EVENTS					
	E_1	E_2	E_3	E_4	E	E
INTENDED PRACTICAL PROPERTIES						
[AGENCY] [MEANS] [GOAL]						
AGENCY PROPERTY			Yes			
MEANS PROPERTY		Yes				
GOAL PROPERTY	F					
ACTUAL PRACTICAL AND NON-PRACTICAL PROPERTIES						
EVENT OF AGENCY			H			
Causes			E_2			
Is caused by			I			
MEANS EVENT		G				
Causes		E_1				
Is caused by		E_3				
GOAL EVENT	F					
Causes	E_4					
Is caused by	E_2					
CONSEQUENCE				C		
Causes				?		
Is caused by				E_1		
CONJOINING CAUSE						
Causes						
Is caused by						
NON-CONJOINING						
Causes						
Is caused by						
NON-ACTUAL EVENT						

HELP

The practical explanatory objects from among which the system will choose are:

The Syntax of MISTAKE:

$$\frac{[\text{AGENCY}]\ [\text{NON-MEANS CAUSAL RELATION}]\ [\text{NON-GOAL CONSEQUENCE}]}{[E_x]}$$

The Syntax of ACCIDENT:

$$\frac{[\text{AGENCY}]\ \left[\begin{array}{c}\text{NON-MEANS}\\ \text{CAUSAL}\\ \text{RELATION}\end{array}\right] + \left[\begin{array}{c}\text{UNFORESEEN}\\ \text{INTERVENING}\\ \text{CAUSE}\end{array}\right]\ \left[\begin{array}{c}\text{NON-GOAL}\\ \text{CONSEQUENCE}\end{array}\right]}{E_x}$$

TThe Syntax of INADVERTENCE:

$$\frac{[\text{Ag}]\ [\text{MEANS}]\ [\text{GOAL}]\ \left[\begin{array}{c}\text{EXTRA-INTENTIONAL}\\ \text{SIDE/END EFFECT}\end{array}\right]}{[E_x]}$$

WHAT WAS INTENDED:

$$\frac{\text{Ag+H}}{E_3} \longrightarrow \frac{\text{G}}{E_2} \longrightarrow \frac{\text{F}}{E_1}$$

WHAT HAPPENED:

$$\frac{\text{Ag+H}}{E_3} \longrightarrow \frac{\text{G}}{E_2} \longrightarrow \frac{\text{F}}{E_1} \longrightarrow \frac{\text{C}}{E_4}$$

COMMENTARY: There is *mismatch* between what was intended and what happened. The mismatch is due either to:

(a) the failure of the event of agency to have the goal-functionality which it was falsely believed and therefore mis-intended to have,

(b) the unforeseen causal intervention of a goal-frustrating field event,
(c) the addition to the intentional of an extra-intentional property or side or end effect

In this case, the mismatch is of the form:

(c) the addition to the intentional of an extra-intentional property or side or end effect

(10) Inadvertence is ascribable:

$$\frac{[\text{Ag}]\ [\text{MEANS}]\ [\text{GOAL}]\ \left[\begin{array}{c}\text{EXTRA-INTENTIONAL}\\ \text{SIDE/END EFFECT}\end{array}\right]}{[E_x]}$$

THE SQUARE OF EXTRA-INTENTIONALITY FOR MY INADVERTENTLY CRUSHING THE CABBAGES

(1) INTENTION $\text{Ag+H ww} \xrightarrow[E_3]{} \underset{E_2}{\text{G}}\ \text{ww} \rightarrow \underset{E_1}{\text{F}}\ \text{Ev}^{+}$	(3) COMMENTARY I inadvertently crushed the cabbages
(2) WHAT HAPPENED $\underset{E_3}{\text{Ag+H}} \rightarrow \underset{E_2}{\text{G}} \rightarrow \underset{E_1}{\text{F}} \rightarrow \underset{E_4}{\text{C}}$	(4) THE INADVERTENCE $\text{Ag+H} \xrightarrow[E_3]{\underset{E_2}{\text{G}}\ \ \underset{E_1}{\text{F}}} \underset{E_4}{\text{C}}$

APPENDIX 1

NOTATION

1. EVENTS

Our notation for events, worldly and mental, and their properties, has the following structure:

[Description]
[Event Name]

1.1 Examples

(1) The event E_1 with property F.

$$\frac{F}{E_1}$$

Such an event would be represented *within* the text as 'F/E_1'.

(2) The event E_3 with the relational property *cause of event* E_2.

$$\frac{\longrightarrow E_2}{E_3}$$

(3) The event E_4 with the relational property *cause of event* E_1 *which has property F.*

$$\frac{\longrightarrow \underline{F}\ E_1}{E_4}$$

(4) The event of agency, E_3, which has the relational property *cause of the event* E_1, *which is F, by cause of* E_2, *which is G.*

Or, the event, E_3, of agency, Ag, which causes G/E_2, which causes F/E_1.

$$\frac{Ag \xrightarrow[E_2]{G} \underline{F}\ E_1}{E_3}$$

2. THE CAUSAL STATES OF AGENCY

Our notation for the cognitive and causal states/events of practical rational agency has the following structure:

[Description]
—-[Attitude function]—-
[Name of the causal object]

2.1 The Active Attitudes

(1) The mental state which is causal, →, of the event, E_1, which is F.

$$\underline{F}$$
$$E_1$$

Such a state would be represented within the text as '–F/E_1→'.

(2) The mental state which is causal, →, of the event, E_3, of agency, Ag, which would cause, ww→, the event, E_1, which is F.

$$\text{Ag ww} \rightarrow \underline{F}$$
$$E_1$$
$$\xrightarrow[E_3]{\qquad}$$

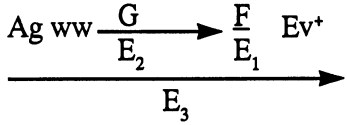
This is the desire to F.

(3) The mental state which is causal of the event, E_3, of agency, Ag, which would cause, ww→, the event, E_2, which is G, which would cause, ww→, the event, E_1, which is F.

$$\text{Ag ww} \xrightarrow[E_2]{G} \underline{F}$$
$$E_1$$
$$\xrightarrow[E_3]{\qquad}$$

This is the desire to F, by G-ing, etc..

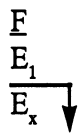
(4) The intention to F, which is causal with respect to an event, E_3, of agency, Ag, which would cause, ww→, the goal event, E_1, which is F, by causing the means event, E_2, which is G, and which causing is judged to be desirable, Ev^+.



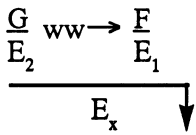
That is, an intention which is causal with respect to the goal-functionality of the event of agency, E_3: that is, the cognitive and causal state which is causal with respect to the causal properties of its proximate effect in virtue of which it has causal relations to the goal.

2.2 The Epistemic Attitudes

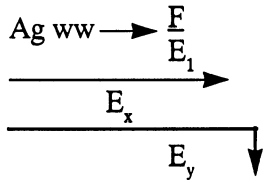
(1) The belief that the, E_1, cat is on the mat, F.



(2) The belief that the event, E_2, which is G, would cause the event, E_1, which is F.



(3) The belief that the agent desires to F.



APPENDIX 2

PART 1. OBJECTS IN THE LOGICAL SPACE OF ACTION

1. EVENT OF AGENCY

A. For all events *x*, *x* is an event of agency *iff*:
   (1) The causal condition: *x* is the *proximate effect* of an intention;
   (2) The intentional condition:
      (a) *x* was intended to be the *proximate effect* of the intention.[1]
      (b) *x* was intended either:
         (i) to be the goal event, or
         (ii) to have all the intended causal properties sufficient for the goal event; i.e., it was intended to have goal-functionality.[2]

B. For all events of agency *y*, *y* is an *intentional* event of agency *iff y* has *all* the properties intended of it: i.e., *iff y* has the intended goal-functionality.

C. For all events of agency *z*, *z* is an *unintentional* event of agency *iff z* fails to have *all*[3] the properties intended of it: i.e., *iff z* fails to have the intended goal-functionality. This failure has two broad forms. Either:
   (1) Mistake: *z lacks* goal-functionality due to a false belief, and has instead either a non-goal property or causal relations to some non-goal consequence.
   (2) Accident: *z lacks* goal-functionality due to an unforeseen causally conjoining field event and has as a result either a non-goal property or causal relations to some non-goal consequence.

2. AGENCY STREAM EVENT

For all events x, x is an agency stream event *iff x* is causally downstream from an efficacious intention.

3. MEANS EVENT

For all events *y*, *y* is a means event *iff y* is a member of the agency stream and y has the intended causal relation to a goal event.

4. GOAL EVENT

For all events *z*, *z* is a goal event *iff z* is a member of the agency stream, *z* has the goal-property and *z* has the intended effect relations from an intention, either proximately, or *via* an event of agency and/or a means event.

5. ACTION

For all *explanatory objects x*, *x* is an *action iff* there is some *y* such that *y* is an intentional event of agency — i.e., *y* has the goal-functionality intended of it — and *x* is that intended goal-functionality of *y*.

6. MISTAKE

For all explanatory *objects y*, *y* is a mistake *iff* there is some *z* such that *z* is an unintentional event of agency and:

(1) *z* lacks its intended goal-functionality,

(2) *z* has an unintended causal relation to some consequence due to a false belief about the causal properties of the event of agency, and

(3) *y* is that unintended functionality of *z*.

7. ACCIDENT

For all explanatory objects *x*, *x* is an accident *iff* there is some *y* such that *y* is an unintentional event of agency and

(1) *y* lacks its intended goal-functionality,

(2) *y* has an unintended causal relation to some consequence due to the causal interference of a conjoining field event, and

(3) *x* is that unintended functionality of *y*.

8. INADVERTENCE

For all explanatory objects *z*, *z* is an inadvertence *iff* there is some *x* such that *x* is an intentional event of agency and

(1) *x* has its intended goal-functionality,
(2) *x* has an extra-intentional property or causal relation to an extra-intentional side- or end-effect, and
(3) *z* is the extra-intentionality of *x*.

9. CONSEQUENCE

For all events *y*, *y* is a consequence *iff*:

(1) The causal condition: *y* is a member of the agency stream, and
(2) The non-intentional condition: *y* was not intended: that is, while it may have been foreseen, *y* was not represented in the *practical* content of the intention which initiated the string of events which make up the agency stream.

10. FIELD EVENT

For all events *z*, *z* is a field event *iff z* is not a member of the agency stream and:

(1) *z* is a conjoining cause which, together with an agency stream event, causes a consequence which is an agency stream event; or,
(2) *z* is a non-conjoining field event which has no causal relations to a member of the agency stream.

PART 2. AGENCY STREAM PROFILES AND THEIR CORRESPONDING EXPLANATORY OBJECTS

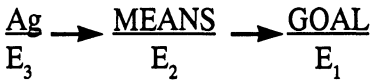
1. ACTION

In action the agency stream has the following profile: the intended causal relations of the event of agency occur and so cause the goal event.

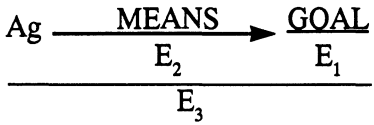
The intended agency stream:

$$\underset{E_3}{Ag} \rightarrow \underset{E_2}{MEANS} \rightarrow \underset{E_1}{GOAL}$$

The actual agency stream:



The action:

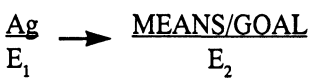


2. MISTAKE

In cases of mistake the agency stream has either of the two following profiles:

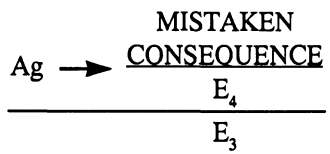
(1) The intended causal relation between the event of agency and the means event or goal event fails to occur (and some other causal relation does occur).

The Intended agency stream:



The actual agency stream:

The mistake:



(2) The intended causal relation(s) between the means event and the goal event fails to occur and some other causal relation does occur.

The intended agency stream:

$$\underset{E_1}{\text{Ag}} \longrightarrow \underset{E_2}{\text{MEANS}} \longrightarrow \underset{E_3}{\text{GOAL}}$$

The actual agency stream:

$$\underset{E_4}{\text{Ag}} \longrightarrow \underset{E_5}{\text{NON-MEANS}} \longrightarrow \underset{E_6}{\text{CONSEQUENCE}}$$

The mistake:

$$\frac{\text{Ag} \xrightarrow[E_5]{\text{NON-MEANS}} \underset{E_6}{\text{MISTAKEN CONSEQUENCE}}}{E_4}$$

3. ACCIDENT

In cases of accident the agency stream has either of the two following profiles:

(1) The intended causal relation(s) between the event of agency and the means event or the goal event fails to occur because of causal interference of a field event.

The intended agency stream:

$$\underset{E_1}{\text{Ag}} \longrightarrow \underset{E_2}{\text{MEANS/GOAL}}$$

The actual agency stream:

$$\underset{E_3}{\text{Ag}} + \underset{E_4}{\text{FIELD EVENT}} \longrightarrow \underset{E_5}{\text{CONSEQUENCE}}$$

The accident:

$$\frac{\text{Ag} \;+\; \underset{E_4}{\text{FIELD EVENT}} \longrightarrow \underset{E_5}{\text{ACCIDENTAL CONSEQUENCE}}}{E_3}$$

(2) The intended causal relation(s) between the means event and the goal event does not occur because of the causal interference of a field event.

The intended agency stream:

$$\underset{E_1}{\text{Ag}} \longrightarrow \underset{E_2}{\text{MEANS}} \longrightarrow \underset{E_3}{\text{GOAL}}$$

The actual agency stream:[4]

$$\underset{E_4}{\text{Ag}} \longrightarrow \underset{E_5}{\text{NON-MEANS}} + \underset{E_6}{\text{FIELD EVENT}} \longrightarrow \underset{E_7}{\text{CONSEQUENCE}}$$

The accident:

$$\frac{\text{Ag} \longrightarrow \underset{E_5}{\text{NON-MEANS}} + \underset{E_6}{\text{FIELD EVENT}} \longrightarrow \underset{E_7}{\text{ACCIDENTAL CONSEQUENCE}}}{E_4}$$

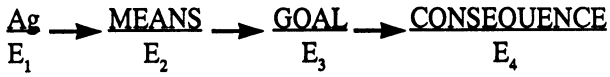
4. INADVERTENCE

In inadvertence the agency stream has the following profile: the event of agency has all the intended properties and relations but the object ascribed, rather than being limited to only what was intended, has extra-intentional properties and/or relations as well.

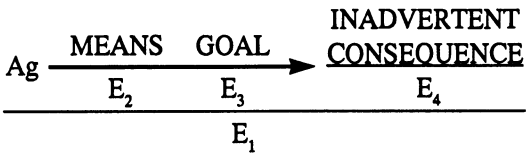
The intended agency stream:

$$\underset{E_1}{\text{Ag}} \longrightarrow \underset{E_2}{\text{MEANS}} \longrightarrow \underset{E_3}{\text{GOAL}}$$

The actual agency stream:



The inadvertence:



NOTES

CHAPTER ONE

1 There is a dispute about the nature of this relation. The *agency theorists*, such as R. Chisholm and R. Taylor, maintain that the relation is "agency causation" — a relation in which an agent stands to an effect; and the *causal theorists*, almost everyone else, maintain that it is garden variety event causation. This dispute reflects a metaphysical difference but not a denial that some such relation is necessary for action. See R. Chisholm, 'The Agent as Cause,' *Action Theory*, M. Brand and D. Walton, eds. (Dordrecht: D. Reidel, 1976) and *Person and Object* (La Salle, Illinois: Open Court, 1976); R. Taylor, *Action and Purpose* (Englewood Cliffs, N.J.: Prentice-Hall, 1966) and *Metaphysics* (Englewood Cliffs, N.J.: Prentice-Hall, 1974).

2 All causal theorists accept this tri-partition.

3 We deal with two versions of the Effect Theory: the Single Effect Theory and the Chain Effect Theory.

4 This view has, unfortunately, come to be called 'The Causal Theory'; unfortunate, since all of its major competitors are causal theories. Proponents of the Effect Theory include: Hobbes, *Leviathan*; Locke, *Essay Concerning Human Understanding*; Hume, *A Treatise of Human Nature* and *Enquiry Concerning Human Understanding*; D. Davidson, *Essays on Actions and Events* (*EAE*) (Oxford: Clarendon Press, 1980); W. Sellars, 'Thought and Action' in *Freedom and Determinism*, K. Lehrer, ed. (New York: Random House, 1966), and 'Actions and Events,' *Nous* 7, (1973); A. Goldman, *A Theory of Human Action* (New York: Prentice Hall, 1970) and 'The Volitional Theory Revisited,' in *Action Theory*, M. Brand and D. Walton, eds.; H-N. Castañeda, *Thinking and Doing* (Dordrecht: D. Reidel, 1975) and 'The Doing of Thinking: Intending and Willing,' in *Action and Responsibility*, M. Bradie and M. Brand, eds. (Bowling Green, Ohio: Bowling Green State University Press, 1980); M. Brand, *Intending and Acting* (Cambridge: MIT Press, 1984); and M. Bratman *Intentions, Plans, and Practical Reason* (Cambridge: Harvard University Press, 1987).

5 The Cause Theory has been held by the following: H.A. Prichard, *Moral Obligation* (Oxford: Clarendon Press, 1949), ("to act is really to will something", 190); B. Aune, *Reason and Action* (Dordrecht: D. Reidel, 1979); J. Hornsby, *Actions* (London: Routledge and Kegan Paul, 1980), (actions = mental tryings); B. O'Shaughnessy, *The Will* (Cambridge: Cambridge University Press,

1980), (actions = attempts). C. Ginet in *On Action* (Cambridge: Cambridge University Press, 1990) develops a theory which is a hybrid of the Cause Theory and the Relational Theory according to which there are mental actions and there are complex actions which include volitions and their effects as proper parts: "Simple actions are all mental actions and include, most importantly, volitions." (73).

6 According to the Relational Theory of Action the causal mental event and the caused bodily event are proper parts of the action. Colin McGinn, in *The Character of Mind* (Oxford: Oxford University Press, 1982) calls this an "ontologically and conceptually hybrid" view of action. See also J. Hornsby, 'Bodily Movements, Actions, and Mental Epistemology,' in *Midwest Studies in Philosophy* **10** (1986). The Relational Theory has been held by J.S. Mill, *A System of Logic* (Book I, chapter iii, sec. 5), J.J. Thomson, *Acts and Other Events* (Ithaca, New York: Cornell University Press, 1977), F. Dretske, *Explaining Behavior: Reasons in a World of Causes* (Cambridge: MIT Press, 1988), and J. Searle, *Intentionality: An Essay in the Philosophy of Mind* (Cambridge: Cambridge University Press, 1983).

7 For an argument for the ineliminability of common sense psychology see Paul Grice, 'Method in Philosophical Psychology (From the Banal to the Bizarre),' in *Proceedings and Addresses of the American Philosophical Association* **xlviii** (Nov. 1975): 23-53. The Humeans argue for a bipartition of practical rationality into desire and belief: they include David Lewis, 'Desire as Belief,' *Mind* **xcvii** (July, 1988): 323-332; Michael Smith, 'The Humean Theory of Motivation,' *Mind* **xcvi** (Jan., 1987): 36-61; Lloyd Humberstone, 'Wanting as Believing,' *Canadian Journal of Philosophy* (1987): 49-62; John Collins, 'Belief, Desire, and Revision,' *Mind* **xcvii** (July, 1988): 333-342.

8 D. Davidson, 'Intending,' *EAE*, 83-102; M. Brand, *Intending and Acting* (Cambridge: MIT Press, 1984).

9 W. Sellars, 'Thought and Action' and 'Fatalism and Determinism,' in *Freedom and Determinism*, K. Lehrer ed., and 'Objectivity, Intersubjectivity, and the Moral Point of View,' in his *Science and Metaphysics* (London: Routledge and Kegan Paul: 1967); and H-N. Castañeda, *Thinking and Doing.*

10 The Anti-Humeans: T. Nagel, *The Possibility of Altruism* (Princeton, N.J.: Princeton University Press, 1970), Part II; J. McDowell, 'Are Moral Requirements Hypothetical Imperatives?,' *Proceedings of the Aristotelian Society Supplementary Volume* (1978); M. Platts, *Ways of Meaning* (London: Routledge and Kegan Paul, 1979), chap. 10.

11 D. Davidson, *EAE, passim*; M. Bratman, *Intention, Plans, and Practical Reason* (*IPPR*), chap. 8; M. Brand, *Intending and Acting*, 31-2, 37, 39, etc.; and C. Ginet, *On Action*, esp. 75-78. See chapter 4, sec. 1, below, for more on this point.

12 We do.

13 Whatever the *desire is belief* proponents mean, it cannot be that the one function is equivalent to the other.

14 This is common ground to all who accept the Davidsonian Effect Theory of Action.

15 Let '*A*' range over agents, let '*P*' range over states of affairs, and let '*F*' range over types of behaviour.

16 (1) through (8) are analyzed in chapters 2 and 3. (9) through (12) and (17) through (19) are analyzed in chapter 4. We give accounts of (13) through (16) later in this chapter.

17 For a useful account of this property of action see Paul Grice, 'Actions and Events,' *Pacific Philosophical Quarterly* **67** (1986): 1-35.

18 There are other Effect Theorists, such as J.J. Thomson, *Acts and Other Events*, who, unlike Davidson, see an action as consisting not solely in the bodily event the agent intentionally causes, but in the entire chain of events from bodily event to the goal event. The information imparted with an action ascription on this theory would include the Davidsonian's information as well as the information that the action consisted of a certain causal chain of events.

19 This description is the one which satisfies Davidson's *criterion of action*. See his 'Agency' in *EAE*. It need not, of course, be the one used in the action ascription.

20 Hornsby's term. See her *Actions*, 1.

21 See Searle's notion of *intention in action* in his *Intentionality*, esp. 94.

22 To accept mental causation is to accept that the explanatory relation holds between mental states of the agent and physical states of the agent's body. If one believes in psychokinesis (of other sorts), then the explanatory relation will hold between mental states of the agent and states of the physical world other than the agent's.

23 A self-explanatory object is one which contains essentially the information about its cause and which therefore explains the effect itself. More on this later.

24 Appraisal of an action is transparently an appraisal of the agent.

25 "*...actus non facit reum, nisi mens sit rea.*"

26 In chapters 2 and 3 we present a theory of Practical Rational Agency and therein an account of the scope and nature of the morally relevant information.

27 See Table 1 'Grading the Theories' below.

28 Prichard's *willing*, Hornsby's *mental trying*, Ginet's *basic mental actions*, and O'Shaughnessy's *attempts*.

29 D. Davidson, M. Brand, and M. Bratman, among others.

30 J.L. Austin, 'A Plea for Excuses,' in *Philosophical Papers* (Oxford: Clarendon Press, 1961), and J. Feinberg, 'Action and Responsibility,' in *Philosophy in America*, M. Black, ed. (Ithaca, New York: Cornell University Press, 1965), 134-160.

31 J.S. Mill, *A System of Logic* (Book I, chapter iii, sec. 5), J. Searle, *Intentionality*, and perhaps J.J. Thomson, *Acts and Other Events.*

32 See J. Feinberg, 'Action and Responsibility,' and Davidson's discussion of *the accordion effect* in 'Agency,' *EAE.*

33 J.J. Thomson, *Acts and Other Events.* This is one possible reading of her view. The other is the Relational Theory.

34 See the literature on events, especially Davidson's essays in *EAE*: 'The Logical Form of Action Sentences,' 'Causal Relations,' 'The Individuation of Events,' 'Events as Particulars,' and 'Eternal *vs.* Ephemeral Events.'

35 See D. Davidson, the chief Single Effect Theorist, in *EAE*, esp. 'Causal Relations,' 'Events as Particulars,' 'The Logical Form of Action Sentences,' and 'Agency.'

36 J.J. Thomson's theory of events in *Acts and Other Events.*

37 See J. Hornsby, *Actions*, 1.

38 An event of agency or agency-event is an intended proximate effect of an intention. See Glossary.

39 See P. Grice, 'Actions and Events.'

40 Although we do take the Davidsonians to have given one essential property to actions which distinguishes them from mere events, it is possible that this theory takes an action to be just an event under a description, namely, 'caused by an intention which truly represented it'. Thus, actions would have no intrinsic properties beyond those which belong to all events. But it is not at all clear why, except for pragmatic reasons, Davidson would allow events under a certain description to constitute an acceptable class of objects, which is what he claims for actions, but not allow other equally important descriptions such as 'intentional' to equally denominate a class of objects. Nevertheless, Davidson's language indicates that he holds the one essential property view of action, i.e., just that property which distinguishes them among events. We accept that view of his. This is not unfair to him since our argument here would oppose equally other such possible views of his.

41 See Paul Grice's *Studies In The Way Of Words* (Cambridge: Harvard University Press, 1989), especially his essays 'Meaning,' and 'Utterer's Meaning, Sentence-Meaning, and Word-Meaning.'

42 J.L. Austin, *Sense and Sensibilia* (Oxford: Clarendon Press, 1962), 74.

43 Including M. Bratman and M. Brand, etc.. See also sec. 1, above, and chapter 4, sec. 1, below.

44 While this matter is germane and comprehensible here, it may be more persuasive if encountered after chapters 2 and 3. Perhaps interested readers will defer their reading of this argument and return to it later.

45 Michael Bratman, 'Two Faces of Intention,' *IPPR*, chapter 8.

46 We use 'efficacious intention' as equivalent to 'intention in action'— Searle's term, in *Intentionality*, and to 'present directed intention' — Bratman's term, in *IPPR*.

47 Bratman, *IPPR*, 113.

48 Ibid., 114.

49 Ibid., 113.

50 Ibid.

CHAPTER TWO

1 See the *unintentionalities* mistake and accident, and the *extra-intentionality* inadvertence, in chapter 5.

2 Terms which are causally related within the processes of practical rationality will also be rationally related: causes will be reasons for their effects.

3 *Pace* Thomson: see J. Hornsby's review of *Acts And Other Events* in *The Journal of Philosophy* **78** (1981): 234-243.

4 See his 'Actions, Reasons and Causes' and 'Agency', in *EAE*, where his *criterion of action* requires that the bodily event caused be intended under at least one true description. This, for us, is an *event of agency* $=_{df}$ the intended proximate effect of an efficacious intention. See GLOSSARY.

5 We do not intend, by this term, what others have: namely, a *sui generis* causation. We mean only what is necessary for agents to be cognitively causal.

6 Brand argues that the objects of the attitudes are properties. What he calls "The Traditional View" takes these objects to be propositions. See Brand's *Intending and Acting*, especially chapter 4, 'The Object of the Attitudes,' 85-118.

7 The intentional object of an active attitude is its causal object: that which it represents and, *ceteris paribus*, causes.

8 They stand in the *about* relation to their cause.

9 Myles Brand discusses two general methods for distinguishing among the attitudes. According to "object assimilationism" the attitudes all take objects of the same type and differ in virtue of their differing attitudes to their objects. Conversely, "attitude assimilationism" is the view that the attitudes of the various mental states are all of a single type and differ, rather, in virtue of taking type-distinct objects. See Brand, *Intending and Acting*, 85 ff.. In our model, the mental states differ in virtue of their causal function and the structure of their content.

10 We take it that the propositional attitudes are states of an agent with cognitive content and a causal function. The class of such states which we call the *active attitudes* includes desires for, that and to, intentions, wishes, hopes and compulsions. They are distinguishable from the *epistemic propositional attitudes* by their causal role and the nature of their cognitive content. The active attitudes alone represent what they cause. The active attitudes are themselves differentiable by their distinct causal functions and their functional positions in the process of intention formation and action. These distinctions will become clear as we proceed.

11 We keep our use of symbols to an absolute minimum and if the reader will bear with us their limited use will make his and our work easier. See APPENDIX 1 for a tutorial on our notation.

12 For typographical reasons, we adapt, where expedient, the mathematician's convention of using the diagonal line '/' in place of the horizontal line to express a fraction.

13 It will become apparent that the notion of an agent's causal state is capable of doing all and only the work of an attitude.

14 Strictly represented, this notation should be:

$$\frac{F}{E_1} \longrightarrow E_m$$

Reading from bottom to top, this is the mental event, E_m, which is causal, $\rightarrow$, with respect to the represented event, E_1, under the description, F. With this understood, we proceed with the simpler $-F/E_1\rightarrow$.

15 A state whose intentional object — what it represents and would, *ceteris paribus*, cause — is that state of affairs: the embarrassment of the government.

16 The intentional object of a cognitive and efficacious state we take to be that which it represents and may cause.

17 While all of the fully practical states of the process we are preparing to describe are conclusory, this would not prevent an agent from being "parachuted into" such a state: that is, finding oneself in a desire-to state, e.g., due to a non-rational process.

[18] The belief that causation is necessary to make actual the non-actual but causable.

[19] 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 of 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.

[20] Perhaps, however, that is exactly what desires do — pass on information into the world. Perhaps both cognitive and non-cognitive items, we and the world, are informationally linked in more directions than are obvious.

[21] The *desire-as-belief* theorists include T. Nagel, *The Possibility of Altruism*, Part II; J. McDowell, 'Are Moral Requirements Hypothetical Imperatives?'; M. Platts, *Ways of Meaning*, chap. 10.

[22] J.L. Austin, 'A Plea for Excuses,' *Philosophical Papers*, 130.

[23] 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*.

[24] "Think of the sky!" Although we are not practitioners of Zen, the above process surely has sometimes not seemed entirely unattractive to authors.

[25] See chapter 3, sec. 8.

[26] See John Searle in both *Intentionality* and *Speech Acts*: *An Essay in the Philosophy of Language* (Cambridge: Cambridge University Press, 1969).

[27] These causal roles may be distinct only in their position in the process. See chapter 4, sec. 6: 'Theoretical Smoothness'.

CHAPTER THREE

1 If not, see Appendix 1 'Notation'.

2 Recall that a desire is the resultant state of a proto-desire's rationalization by the belief in causation. See chap. 3, sec. 3.

3 Let 'B_i' represent the belief complex of (1) the belief that causation is necessary and sufficient for every non-actual and causable event, and (2) the belief that, in this example, the embarrassment of the government is non-actual and causable.

4 The proto-practical state $=_{df}$ a desire-that. See GLOSSARY.

5 A weaker condition for the rational entry into practical reasoning would be, not that the system believe that a causing of the motivant is possible, but only that it not believe that such a causing is impossible. We have chosen the stronger condition.

6 A parallel attenuation of belief states is possible: believing, doubting, considering, entertaining, wondering whether, etc., are distinguishable by a corresponding attenuation of assent. To claim assent is to report on an epistemological causal state with respect to some cognitive content.

7 Or again, as above, the weaker condition that the system does not believe that the causing is impossible.

8 Later in this chapter in the section entitled 'MODEL II' we offer an alternative and better engineered model of practical rational agency whose explanation and advantages are more readily seen against the articulation of this more intuitively attractive one.

9 This is the first point at which it would be true to say of the agent that his goal is the embarrassment of the government. Prior to the desire-to stage, the embarrassment functions merely as a motive for practical speculation: the agent desires that there be such a happening but not that it be a happening he initiates. The latter state puts the embarrassment of the government into relation with his agency and that is what makes it his goal.

10 AUTO-*PRAGMA*.

11 'Effects' and not 'affects' since agents make actual, i.e., cause, their intended possible worlds.

12 We represent the essential normative content of an intention *is desirable* with 'Ev^+', our symbol for *positive evaluation.*

13 See M. Bratman, 'Two Faces of Intention,' *The Philosophical Review* **93** (1984): 375-405; and *IPPR*.

14 A means event is one the causal relation of which stands between a goal event and an event of agency. The means itself is the goal-functionality of an event of agency: the causal properties of an event of agency to cause the goal.

15 Davidson rejects the second version too but does not offer the first. See his 'Agency,' in *EAE*. For interest's sake, note that in the question "What did you do?" 'did' functions only as an auxiliary (Jennifer Hornsby made, we believe, the same point about 'be') so that it could be replaced by the past tense of 'do' and the question becomes "What was it that you did?" in which both 'did' and 'it' may be treated as they were in "You did it".

16 See Bratman, *IPPR*.

17 We examine the practical syllogism with an intention premise in the next section.

18 Apologies to Castañeda.

19 There is, of course, a large body of work on this problem, which G. Harman has called the problem of the *extent of intention*. See his 'The Extent of Intention,' *Social Theory and Practice* **9**, 2-3 (Summer 1983). There are those — the Holists — who hold that not only are means intended just as goals are, but that all foreseen aspects of one's action are also intended. R.B. Brandt, in *A Theory Of The Good And The Right* (Oxford: Clarendon Press, 1979) argues for a version of Holism. It is also one of the main views of intention in the law: See, for example, R.A. Duff, 'Intentions Legal and Philosophical,' *Oxford Journal of Legal Studies* **9**, 1 (1989); John Finnis, 'Intentions and Side-Effects' (draft, 1988); J.C. Smith, 'A Note On Intention,' *The Criminal Law Review* (1990); The Law Reform Commission of Canada, 'Report No. 31, Recodifying Criminal Law,' (1987); The British Law Commission for the Criminal Code, nos. 89, 143, 177.

20 D. Davidson, 'Intending,' *EAE*, 83-103; M. Bratman, 'Intention and Means-End Reasoning,' *The Philosophical Review* **90** (1981): 252-265; 'Taking Plans Seriously,' *Social Theory and Practice* **9** (1983): 271-287; 'Two Faces of Intention,' (1984).

21 See M. Bratman, *IPPR*.

22 Brand, following Sellars, who finds Davidson lacking on this matter, calls the latter question, "possibly *the* fundamental problem ... in philosophical action theory." M. Brand, *Intending and Acting*, 33.

23 See A. Danto, 'Basic Actions,' *American Philosophical Quarterly* **2** (1965): 141-148; D. Davidson, 'Agency,' *EAE*, 43-62.

24 See M. Brand, *Intending And Acting*, chap. 1.

25 The active attitudes are also the states which in concert with beliefs are causally functional with respect to the processional aspects of practical reasoning — the change from state to state.

26 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,' *EAE*, 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.

27 See especially M. Bratman, *IPPR*.

28 See the dispute between the Humean two function theorists and the anti-Humean single function theorists mentioned in chap. 1 above.

29 D. Davidson, 'Intending,' in *EAE*.

30 See "desire-as-belief" view, chapter 1.

31 This echoes Hume's *dictum* that there is no action without passion, or motivation. See D. Hume, *A Treatise of Human Nature*, Book II, Part III, S. III.

32 Can we amend a given desire without the sheer *donné* of an antecedent desire which itself cannot be amended without some further such *donné*; but won't the process of amendment thus be *not* in the control of reason alone? Is some useful combination of the two models possible? Model III is a possibility: in this model the desire function emerges only from a fully rationalized belief function. That is, from an all out judgement or belief that F-ing is desirable, an intention to F emerges. This emergent view comes closest to the desire-as-belief thesis, but for any thesis about rational agency both functions are necessary.

33 But not, perhaps, of non-cognitive structure.

34 See Harry Frankfurt, 'Freedom of the Will and the Concept of a Person,' *Journal of Philosophy* **68** (Jan. 1971): 5-20.

CHAPTER FOUR

1 See M. Brand, *Intending and Acting*, 31-32, for a commitment to, but no argument for, the negation of this proposition.

2 See chapter 1 above and S.C. Coval and J.C. Smith, *Law and its Presuppositions: Actions, Agents and Rules*, chapter 1, as well as S.C. Coval and T. Forrest 'Which Word Wears the Trousers?, *Mind* **76** (1967): 73-82, and S.C. Coval and D.D. Todd 'Adjusters and Sense-Data,' *American Philosophical Quarterly* **9** (1972): 107-112.

3 The trio of terms which correspond to those which operate in action, *viz.*, 'did it' (the standard case), 'did it unintentionally' (the deviation from the standard), and 'did it intentionally' (the rebuttal of the deviation and the setting back to standard), are, for 'reality' and 'freedom' respectively, 'is an A', 'is an apparent but false or counterfeit A', and 'is a real A'; 'did it', 'did it under duress', and 'did it freely'.

4 Is this Davidson's early, but later neglected, view from 'Actions, Reasons, and Causes,' that 'intention' is a syncategorematic term?

5 Davidson's for example.

6 It should be possible to construct a table, which, however, we here spare the reader, whose y axis consisted of normal, over-weak and over-strong active function, and whose x axis consisted of normal, over-weak and over-strong information function. The information function should be divided into goal-functional and evaluative information. This table would yield 27 discrete types of case which can occur within the process of practical reasoning. It would be an interesting exercise to try to put common names to all these cases, some of which we have done above.

7 A fuller account of *trying* occurs in chapter 1.

8 See Davidson's 'Causal Relations' and 'Agency' in *EAE*.

9 See chapter 1 above for more on this matter.

10 This is somewhat the same as saying that he *really* called upon his neighbour.

11 The absence of this entails mistake or accident depending upon whether the absence of true belief is replaced with a false belief or no belief. Evidence for this follows shortly.

12 If the representation is unjustified, we have a case of serendipity; if the representation or causal belief is that the justification of the nomic connection is doubtful we have a case of trying which will be either an unsuccessful or successful attempt according to whether the nomic belief is false or true. In each of the above cases the act is non-standard and contains an argument for exculpation or "excreditation".

13 He was, that is, in a state which was causal with respect to the object represented in (2) below.

14 As with any use of causation, the notion of effectiveness has to be treated so that causation does not occur through some further unintended medium. The intention has to be the *direct* cause of what is intended. Brand and Davidson raise interesting cases of *antecedential waywardness* which, because of an intervening causal medium, require a tightening of the causal relation. Brand speaks of *proximate causes* in this connexion.

15 See chapter 5 for a fuller account of the Squares.

16 Again, this step is required for a rational practical system, not for a practical system which is not.

17 If the agent is not justified in his means or identificatory beliefs or in his evaluative beliefs but nevertheless causes what he intends, then we qualify the intentionality of the act ascribed to him as *serendipitous*. A question raised by condition (6) is whether, if his *evaluative* beliefs are false, the case would be treated exactly as if his causal beliefs were false, namely, as mistake and unintentional.

18 In our symbols for intention which follow, we omit the notation 'Ev^{+}' for *is desirable*, the essential normative content of an intention.

19 D. Davidson, 'Agency,' *EAE*, 46.

20 Why not "intentionable"? Because that would contain a redundancy.

21 See 'The Practical Syllogism: Intention,' in chapter 3 above, for the way in which the identity of intentions are affected by false beliefs.

22 "... sinking the Bismark is his action, for that action is identical with his attempt to sink the ship he took to be the Tirpitz, which is intentional.", 'Agency,' in *EAE*, 46.

23 See 'Intending' in chapter 3 above, and M. Bratman, 'Two Faces of Intention.' Could this inconsistency be made even sharper by having the agent believe that turning this knob has the causal relations he wants to the sinking of the Tirpitz when in actuality it stands in causal relations to the prevention of the sinking of the Tirpitz. For Davidson, the officer would have intended, by substitution, the prevention of the Tirpitz's sinking and also have intended its sinking.

24 !

25 J.L. Austin, 'A Plea for Excuses.'

26 Normative appraisals of agency include moral, legal, and aesthetic appraisals, among others.

27 What is foreseeable and what unforeseeable will be determined by one's theory of *the reasonable man*.

28 Emphasis ours. M. Bratman, *IPPR*, 113, 167.

29 Ibid., passim, and especially chapter 8: "My version of the Single Phenomenon View loosens the connection between what is intended and what is done intentionally: it sees what is intended as a fact about the agent's mind which need not reflect all the complexities of our scheme for classifying actions as intentional", 124.

CHAPTER FIVE

1 *PRAGMA* will contain a library of cases such as the following which users may choose to have analyzed.

2 In the interest of brevity, we will here omit the steps of filling in the Object Frame until the complete Frame can be shown at the end of the analysis.

3 From Austin's example in 'A Plea for Excuses.'

APPENDIX 2

1 The description 'this proximate effect of this efficacious intention' from the practical content of its proximately causing efficacious intention is *true* of it. Or, equivalently, it has that property/causal relation of being the intended proximate effect of the efficacious intention.

2 That is, an event is an event of agency *iff* it satisfies Davidson's *criterion of action*! What is an action for Davidson is an event of agency for us. See his 'Agency'.

3 Not 'all *and only*' because every event of agency fails to have only its intended properties and relations. Notice that this makes the event of agency in cases of inadvertence not unintentional: only in cases of mistake and accident is the corresponding event of agency unintentional.

4 The non-means event, E_5, is best described, in such cases, as 'the event which would have been means had it not been for the causal interference of the conjoining field event'. That is, E_5 is a *contrary-to-fact* means, or a *counterfactual* means.

GLOSSARY

ACCIDENT: a practical explanatory object which is the non-goal-functionality of an event of agency caused by the unforeseen causal intervention of a FIELD EVENT. The causal intervention has made the corresponding means belief false, and therefore rendered the event of agency non-goal-functional. ACCIDENT is one form of UNINTENTIONALITY - the other is MISTAKE - which make the entire AGENCY STREAM unintentional.

ACTION: a practical explanatory object which is the intended GOAL-FUNCTIONALITY of an EVENT OF AGENCY; an action is all and only the intended properties of an EVENT OF AGENCY which enable it to be the first event in the intended causal pathway to the agent's GOAL.

ACTIVE ATTITUDE: an informational and causal function which takes information from beliefs (an EPISTEMIC ATTITUDE) and can cause what it represents. See DESIRE; INTENTION; PROTO ACTIVE ATTITUDE.

AGENCY PROPERTY: the property of an event in virtue of which it is an EVENT OF AGENCY; namely, its being the proximate effect of an agent's EFFICACIOUS INTENTION.

AGENCY STREAM: the chain of events which is causally downstream of an EFFICACIOUS INTENTION. Events not in the AGENCY STREAM are FIELD EVENTS.

BELIEF: a rationalized informational and causal function (an EPISTEMIC ATTITUDE) which is an information-conductor in PRACTICAL REASONING.

BODY EVENT: an event of an agent's body.

CONSEQUENCE: an event which is an extra-intentional effect of an event of agency, a means event, or a goal event but being neither of these. See also: EXTRA-INTENTIONAL; SIDE-EFFECT; END-EFFECT.

DESIRE: a practical ACTIVE ATTITUDE whose object is the coming to be of a GOAL EVENT; an informational and casual function whose causal role is to cause what it represents. See ACTIVE ATTITUDE, DESIRE-THAT and DESIRE-TO.

DESIRE-THAT: a desire whose object is the coming to be of a motivating state of affairs.

DESIRE-TO: a desire whose object is an EVENT OF AGENCY with causal relations to a GOAL EVENT.

EFFICACIOUS INTENTION: an informational and causal function of the *genus* ACTIVE ATTITUDE whose goal-functional information has been sufficiently "vetted" within the process of practical rationality to achieve efficacy. Information has been sufficiently "vetted" for rational efficacy when its object has been judged desirable and is believed a sufficient description of the goal-functionality of the event of agency. See INTENTION.

END-EFFECT: an event which is a CONSEQUENCE of a GOAL EVENT and is therefore EXTRA-INTENTIONAL. See also SIDE-EFFECT.

EPISTEMIC ATTITUDE: the *genus* of informational and causal functions which includes BELIEF, doubt, musings, speculations, wonderings, assumptions, presumptions, etc., which can take on information from events, etc., and which can pass on information to ACTIVE ATTITUDES and other states.

EVENT OF AGENCY: the PROXIMATE EFFECT of an agent's EFFICACIOUS INTENTION which was represented in that intention by at least one true description of it. The minimal descriptive content for an event of agency is the demonstrative content *this event* which is a component of every efficacious intention. For agents like ourselves, an event of agency will be either a body event or a mental event. An EVENT OF AGENCY is equivalent to a Davidsonian action.

EXTRA-INTENTIONAL: those properties and relations of an AGENCY STREAM (and its field) which were not intended. The EXTRA-INTENTIONAL is equal to the extra-goal-functional.

FIELD EVENT: an event which is not part of an AGENCY STREAM. Field events are of two sorts: those which contribute causally to the agency stream and those which do not. Of those field events which do contribute causally to the agency stream, some will be goal-frustrating, some will be goal-contributing, and others will be neutral in this respect.

GOAL EVENT: an event with a GOAL PROPERTY whose causing is the object of a DESIRE-TO or INTENTION.

GOAL-FUNCTIONALITY: the intended causal properties of an EVENT OF AGENCY in virtue of which it is the first event in a chain of events to the GOAL.

GOAL PROPERTY: a relational or non-relational property of an event the causing of which is the object of a DESIRE-TO or INTENTION, and which was inherited by the DESIRE-TO or INTENTION from its PROTO ATTITUDE ancestor.

GOAL: a motivating event or property the causing of which is the object of a DESIRE-TO or INTENTION.

INADVERTENCE: a practical explanatory object which is an event of agency with, in addition to the intentional, an extra-intentional property or causal relation to a side-effect or an end-effect. This extra-intentional addition is definitive of inadvertence and makes the entire object extra-intentional. It is the function of inadvertence to explain this type of extra-intentional property of an event of agency.

INTENDED PROPERTY: a GOAL-FUNCTIONAL property of an EVENT OF AGENCY.

INTENTION: a causal and informational state of the type ACTIVE ATTITUDE whose function is causal (potentially or actually) with respect to the object it represents in its content, that object having been judged to be, on the whole, desirable.

INTENTIONAL EVENT OF AGENCY: an EVENT OF AGENCY which has all the properties and relations it was intended to have, i.e., those goal-functional properties represented in the content of its EFFICACIOUS INTENTION.

MEANS: the intended GOAL FUNCTIONALITY of an event; the intended GOAL-causing property of an event.

MEANS EVENT: an event with intended causal relations from the agent's EFFICACIOUS INTENTION and to the agent's GOAL EVENT. An event with a GOAL-FUNCTIONAL causal relation.

MISTAKE: a practical explanatory object which is the non-goal-functionality of an event of agency which it has instead of the intended MEANS properties and which is explained by the falsity of the corresponding means BELIEF. These properties are definitive of mistake and makes the entire object UNINTENTIONAL. It is the function of MISTAKE to explain its unintentional properties.

NON-INTENTIONAL: those events and their properties which are not in any AGENCY STREAM.

PRACTICAL RATIONAL AGENCY: the causing of an event of agency by a state of an epistemological, teleological and effective system which has been rationalized by the process of practical reasoning.

PRACTICAL RATIONAL AGENT: a system with the properties of PRACTICAL RATIONAL AGENCY.

PRACTICAL REASONING: the process of intention formation in which an event of agency with GOAL-FUNCTIONAL properties is identified and evaluated.

PROTO ACTIVE ATTITUDE: a non-practical but ACTIVE ATTITUDE whose causal object is only a represented property of an event rather than its causing.

PROXIMATE EFFECT: a "direct" effect; an effect not separated from its cause by intervening events; the presumption that there is a conception of causation which allows an INTENTION to be capable of causing an ACTION in the way intended.

REPRESENTATION: the informational content of an attitude.

SIDE-EFFECT: an event which is a collateral consequence of an EVENT OF AGENCY, a MEANS EVENT, or a GOAL EVENT,

and therefore EXTRA-INTENTIONAL whether represented (foreseen) or not.

UNINTENTIONAL: any properties of events in an AGENCY STREAM which are not the GOAL-FUNCTIONAL properties.

UNINTENTIONAL EVENT OF AGENCY: an EVENT OF AGENCY which does not have all the GOAL-FUNCTIONAL properties it was intended to have.

BIBLIOGRAPHY

Anscombe, G.E.M. (1957). *Intention*. Oxford: Blackwell.

Audi, Robert (1973). 'Intending.' *The Journal of Philosophy* **70**: 387-403.

—— (1989). *Practical Reasoning*. London: Routledge and Kegan Paul.

Aune, Bruce (1979). *Reason and Action*. Dordrecht: D. Reidel.

Austin, J.L. (1961). 'A Plea for Excuses.' *Philosophical Papers*, ed. J.O. Urmson and G.J. Warnock. Oxford: Oxford University Press.

—— (1962). *Sense and Sensibilia*. Oxford: Clarendon Press.

Bentham, Jeremy, (1982). *An Introduction to the Principles of Morals and Legislation*. London: Methuen.

Bishop, John (1989). *Natural Agency*. Cambridge: Cambridge University Press.

Black, Max, ed. (1965). *Philosophy in America*. Ithaca, New York: Cornell University Press.

Bradie, M. and Myles Brand, eds. (1980). *Action and Responsibility*. Bowling Green, Ohio: Bowling Green State University Press.

Brand, Myles and D. Walton, eds. (1976). *Action Theory*. Dordrecht, Holland: D. Reidel.

Brand, Myles (1984). *Intending and Acting*. Cambridge, Mass.: MIT Press.

Brandt, R.B. (1979). *A Theory Of The Good And The Right*. Oxford: Clarendon Press.

Bratman, Michael (1981). 'Intention and Means-End Reasoning.' *The Philosophical Review* **90**: 252-265.

—— (1983). 'Taking Plans Seriously.' *Social Theory and Practice* **9**: 271-287.

—— (1984). 'Two Faces of Intention.' *The Philosophical Review* **93**: 375-405.

—— (1987). *Intentions, Plans, and Practical Reasoning* (*IPPR*). Cambridge: Harvard University Press.

British Law Commission for the Criminal Code. Reports 89, 143, and 177.

Castañeda, Hector-Neri (1975). *Thinking and Doing*. Dordrecht, Holland: D. Reidel.

—— (1980). 'The Doing of Thinking: Intending and Willing' in *Action and Responsibility*, M. Bradie and M. Brand, eds., Bowling Green, Ohio: Bowling Green State University Press.

Chisholm, R. (1976). 'The Agent as Cause' in *Action Theory*, M. Brand and D. Walton, eds. Dordrecht: D. Reidel.

—— (1976). *Person and Object.* La Salle, Illinois: Open Court.

Collins, John (1988). 'Belief, Desire, and Revision,' *Mind* **XCVII**: 333-342.

Coval, S.C. and T. Forrest (1963). 'Which Word Wears the Trousers.' *Mind* **LXXVI**: 73-82.

Coval, S.C. and D.D. Todd (1972). 'Adjusters and Sense-Data.' *American Philosophical Quarterly* **9**: 107-112.

Coval, S.C. and J.C. Smith (1986). *Law and Its Presuppositions: Actions, Agents and Rules.* London: Routledge and Kegan Paul.

Danto, Arthur C. (1965). 'Basic Actions.' *American Philosophical Quarterly* **2**: 141-148.

Davidson, Donald (1963-78). *Essays on Actions and Events* (*EAE*). Oxford: Clarendon Press.

Dretske, Fred (1986). *Explaining Behavior*: Reasons in a World of Causes. Cambridge, Mass.: MIT Press.

Duff, R.A. (1989). 'Intentions Legal and Philosophical.' *Oxford Journal of Legal Studies* **9**: 76-94.

Feinberg, J. (1965). 'Action and Responsibility' in *Philosophy in America,* Max Black ed. Ithaca, N.Y.: Cornell University Press.

Finnis, John (1988). 'Intention and Side-Effects' (draft).

Frankfurt, Harry (1971). 'Freedom of the Will and the Concept of a Person.' *Journal of Philosophy* **68**: 5-20.

Ginet, Carl (1990). *On Action.* Cambridge: Cambridge University Press.

Goldman, Alvin (1970). *A Theory of Human Action.* New Jersey: Prentice-Hall.

—— (1976). 'The Volitional Theory Revisited' in *Action Theory,* M. Brand and D. Walton, eds. Dordrecht, Holland: D Reidel.

Grice, H.P. (1971). 'Intention and Uncertainty.' *Proceedings of the British Academy* **57**: 263-279.

—— (1975). 'Method in Philosophical Psychology (From the Banal to the Bizarre)' in *Proceedings and Addresses of the American Philosophical Association* **XLVIII**: 23-53.

—— (1986). 'Actions and Events.' *Pacific Philosophical Quarterly* **67**: 1-35.

—— (1989). *Studies In The Way Of Words.* Cambridge: Harvard University Press.

Harman, Gilbert (1983). 'Rational Action and the Extent of Intentions.' *Social Theory and Practice* **9**: 123-141.

Hobbes, Thomas (1651). *Leviathan.*

Hornsby, Jennifer (1980). *Actions.* London: Routledge and Kegan Paul.

—— (1981). Review of *Acts and Other Events*, by J. J. Thomson. In *The Journal of Philosophy* **78**: 234-243.

—— (1986). 'Bodily Movements, Actions, and Mental Epistemology.' *Midwest Studies in Philosophy* **10**.

Humberstone, Lloyd (1987). 'Wanting as Believing.' *Canadian Journal of Philosophy*: 49-62.

Hume, David (1739). *A Treatise of Human Nature.*

—— (1748). *Enquiry Concerning Human Understanding.*

LePore, Ernest and Brian McLaughlin, eds. (1985). *Actions and Events: Perspectives on the Philosophy of Donald Davidson.* Oxford: Basil Blackwell Ltd.

Law Reform Commission of Canada, (1987). 'Report No. 31: Recodifying Criminal Law'.

Lehrer, Keith, ed. (1966). *Freedom and Determinism.* New York: Random House.

Lewis, David (1988). 'Desire as Belief.' *Mind* **XCVII**: 323-32.

Locke, John (1690). *Essay Concerning Human Understanding.*

McDowell, J (1978). 'Are Moral Requirements Hypothetical Imperatives?' in *Proceedings of the Aristotelian Society Supplementary Volume.*

McGinn, Colin (1982). *The Character of Mind.* Oxford: Oxford University Press.

Mill, J.S. (1872). *A System of Logic.*

Nagel, Thomas (1970). *The Possibility of Altruism.* Princeton, N.J.: Princeton University Press.

O'Shaughnessy, Brian (1980). *The Will.* Cambridge: Cambridge University Press.

Platts, M. (1979). *Ways of Meaning.* London: Routledge and Kegan Paul.

Prichard, H.A. (1949). *Moral Obligation.* Oxford: Clarendon Press.

Raz, Joseph ed. (1978). *Practical Reasoning.* Oxford: Oxford University Press.

Searle, John R. (1969). *Speech Acts: An Essay in the Philosophy of Language.* Cambridge: Cambridge University Press.

—— (1979). 'The Intentionality of Intention and Action.' *Inquiry* **22**: 253-280.

—— (1983). *Intentionality: An Essay in the Philosophy of Mind.* Cambridge: Cambridge University Press.

Sellars, Wilfrid (1966). 'Thought and Action' in *Freedom and Determinism*, K. Lehrer, ed. New York: Random House.

—— (1966). 'Fatalism and Determinism,' in *Freedom and Determinism*, K. Lehrer, ed. New York: Random House.

—— (1967). 'Objectivity, Intersubjectivity, and the Moral Point of View,' in his *Science and Metaphysics.* London: Routledge and Kegan Paul.

—— (1973). 'Actions and Events.' *Nous* **7**: 179-202.

Sidgwick, Henry, (1962). *The Methods of Ethics*. Chicago: University of Chicago Press.

Smith, J.C. (1990). 'A Note On Intention.' *The Criminal Law Review*: 85-91.

Smith, Michael (1987). 'The Humean Theory of Motivation.' *Mind* **XCVI**: 36-61.

Taylor, R (1966). *Action and Purpose*. Englewood Cliffs, N.J.: Prentice-Hall.

—— (1974). *Metaphysics*. Englewood Cliffs, N.J.: Prentice-Hall.

Thalberg, Irving (1972). *Enigmas of Agency*. New York: Humanities Press.

Thomson, J.J. (1977). *Acts and Other Events*. Ithaca, N.Y.: Cornell University Press.

Wilson, George M, (1980). *The Intentionality of Human Action*. Amsterdam: North Holland.

—— (1985). 'Davidson on Intentional Action,' in *Actions and Events: Perspectives on the Philosophy of Donald Davidson*, Ernest LePore and Brian McLaughlin, eds. Oxford: Basil Blackwell Ltd.

INDEX

STUDIES IN COGNITIVE SYSTEMS

Series Editor: James H. Fetzer, *University of Minnesota, Duluth*

1. J. H. Fetzer (ed.): *Aspects of Artificial Intelligence.* 1988
ISBN 1-55608-037-9; Pb 1-55608-038-7
2. J. Kulas, J.H. Fetzer and T.L. Rankin (eds.): *Philosophy, Language, and Artificial Intelligence.* Resources for Processing Natural Language. 1988
ISBN 1-55608-073-5
3. D.J. Cole, J.H. Fetzer and T.L. Rankin (eds.): *Philosophy, Mind and Cognitive Inquiry.* Resources for Understanding Mental Processes. 1990
ISBN 0-7923-0427-6
4. J.H. Fetzer: *Artificial Intelligence: Its Scope and Limits.* 1990
ISBN 0-7923-0505-1; Pb 0-7923-0548-5
5. H.E. Kyburg, Jr., R.P. Loui and G.N. Carlson (eds.): *Knowledge Representation and Defeasible Reasoning.* 1990 ISBN 0-7923-0677-5
6. J.H. Fetzer (ed.): *Epistemology and Cognition.* 1991
ISBN 0-7923-0892-1
7. E.C. Way: *Knowledge Representation and Metaphor.* 1991
ISBN 0-7923-1005-5
8. J. Dinsmore: *Partitioned Representations.* A Study in Mental Representation, Language Understanding and Linguistic Structure. 1991
ISBN 0-7923-1348-8
9. T. Horgan and J. Tienson (eds.): *Connectionism and the Philosophy of Mind.* 1991 ISBN 0-7923-1482-4
10. J.A. Michon and A. Akyürek (eds.): *Soar: A Cognitive Architecture in Perspective.* 1992 ISBN 0-7923-1660-6
11. S.C. Coval and P.G. Campbell: *Agency in Action.* The Practical Rational Agency Machine. 1992 ISBN 0-7923-1661-4

KLUWER ACADEMIC PUBLISHERS – DORDRECHT / BOSTON / LONDON