Thinking before acting: intentions, logic, rational choice

Roy, O.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 6

Hybrid pragmatism, acceptances and norms on intentions

In the foregoing chapters I have explored how the volitive and the reasoning-centered commitment of intentions influence the deliberations of planning agents. As I mentioned in the Introduction, Section 1.2, intentions exert this influence because they are subject to four rationality constraints. They must be internally consistent, consistent with the beliefs, agglomerative and means-end coherent.

In this chapter I leave the formal theory and try to understand where these normative requirements come from. That is, I examine various ways to account for the fact that intentions are subject to these rationality constraints.

This issue has attracted much attention in recent years. Some authors have argued that the norms on intentions—which are usually thought of as practical norms—stem in fact from analogous norms of theoretical rationality associated with beliefs. This is the cognitivist view, championed for example by Harman [1976, 1986], Velleman [2003, 2005] and Setiya [2007, forthcoming]. Others, chiefly Bratman [2006b], have rather proposed an “agency approach”, which avoids this reduction of the practical to the theoretical by accounting for the norms on intentions solely in pragmatic terms.

Both sides have their pros and cons. But, with the notable exception of Wallace [2006, 2003a], very few authors have moved away from these two extremes. There is, however, sufficient conceptual space for manoeuvre between pure cognitivism and the agency approach. In this chapter I investigate how far one can get in accounting for the normative requirements on intentions by using such an “intermediate” approach. This approach, based on the concept of acceptances in deliberation, tries to derive the norms on intentions from similar norms on acceptances. As we shall see, this is an essentially mixed approach because acceptances are “hybrid” attitudes, definitely on the cognitive side but still responsive to practical concerns. For that reason, I call it hybrid pragmatism. I argue that it provides a reasonable compromise between cognitivism and the agency approach, in a way that does justice to both the practical and theoretical aspects of the
norms on intentions.

In Section 6.1 I return in greater detail to the four normative requirements on intentions that I presented in the Introduction. Section 6.2 is devoted to the cognitivist derivation of these requirements. The key sections of this chapter are 6.3 and 6.4, in which I introduce acceptances in deliberation and study how they can account for the norms on intentions.

In contrast to the preceding chapters, the investigation here is not formally driven. My goal is to provide the theory that I have developed so far with a philosophically solid basis. It will come clear as we move along, though, that hybrid pragmatism, with its focus on acceptances in deliberation, also introduces new issues onto the agenda for more formal enquiry. As such it helps us to understand better the functions of intentions in practical reasoning while opening up further research directions.

6.1 Constraints on rational intentions

In the Introduction (Section 1.2) I mentioned that rational intentions are required to be internally consistent, strongly belief-consistent, agglomerative and means-end coherent. Here I present these norms in more detail.

Means-end Coherence. A plan that aims to achieve some ends must contain intentions about necessary means\(^1\). More precisely, the agent must intend to do what he believes is necessary for him to intend to reach his end. Or, at least, he must plan to form the appropriate means-intentions later.

It is crucial that the “necessary means” are those that the agent has to intend to reach his end. Take for example an agent who intends to bring his oven up to a certain temperature, say to bake a cake. He might also know that this boils down to transforming electrical energy into heat. But he does not seem to be means-end incoherent if he does not have per se the intention to transform energy. To be sure, he has to intend to turn on the oven, which might require other actions such as turning some switches, but even though transforming energy is from a certain point of view a necessary means to achieve the required intention, it is not one that has to be directly intended for it to happen. The proviso on means-end coherence is precisely intended to cope with such cases.

The cases of means-end incoherence that I discuss feature a “gap” in the intention structure of the agent. That is, these are cases such that the agent intends an end \(E\), believes that to achieve \(E\) he must form an intention about some means \(M\), but does not have that intention. Of course, one can also conceive of stronger cases of means-end incoherent plans. An agent can intend an end \(E\),

\(^1\)Many authors have discussed the interpretation of this principle. The reader can consult Harman [1976], Brunero [forthcoming], Wallace [2006], Kolodny [2007], and Setiya [forthcoming].
believe that to achieve $E$ he must come to form intentions about some means $M$ but, instead, forms an intention that excludes his doing $M$. Such cases can in general be reduced to violations of one of the other requirements, and so I do not treat them as violations of means-end coherence.

**Strong Belief Consistency.** Means-end coherence crucially involves what the agent believes. But this connection between intentions and beliefs goes further. An intention should be feasible in a world where the agent’s beliefs are true. That is, a plan should not be impossible to realize, given what the agent believes. This is what Bratman [1987, p.31] calls the *strong belief consistency* requirement.  

**Internal Consistency.** Plans and intentions themselves have to be consistent. First, the content of an intention should be consistent. An agent should not intend impossible things, for example to do $A$ and not to do $A$. Let me call this the internal consistency of intentions. But it also seems reasonable to ask plans to be consistent as wholes. The intentions in a plan should not contradict each other, they should not preclude one another’s achievement. This can be called internal consistency of plans. Observe that a plan can be internally inconsistent even if each of its intentions is internally consistent. Internal inconsistency of plans arises out of the combination of the intentions it contains, a phenomenon that naturally brings us to the next requirement.

**Agglomerativity.** The original formulation of the principle of agglomerativity goes as follows:

> Given the role of intentions in coordination, there is rational pressure for the agent to put his intentions together into a larger intention.
> [Bratman, 1987, 134, my emphasis]

As I mentioned in the Introduction (Section 1.2), one can distinguish two readings of this principle. First, there is what I called *agglomerativity as closure*: an agent cannot rationally intend to $A$ and intend to $B$ unless he also intends both $A$ and $B$. What Bratman thus meant by “put intentions together” is, according to this interpretation, that they should close under conjunction.

For Yaffe [2004] this principle demands too much effort from agents with limited time and capacities. “It demands mental labor from agents that they have no need to perform, given their aims.” In other words, agglomerativity as closure requires worthless combinations of intentions. He illustrates his claim

---

2 Note that an agent does not have to believe that his plan is impossible to realize for it to be belief inconsistent. This is a stronger condition, to which I shall return soon. In Sections 6.2.4 and 6.4.4, however, I directly adopt this stronger condition.

3 The two interpretations of agglomerativity, and the quotations in the following paragraphs, come from Yaffe [2004, p.511-512].
with the following example. Suppose one has the intention to go to Los Angeles tomorrow and the intention to go to London a year from now. To him, it is not worth the effort to combine these two intentions into a single one. To have the merged intention neither contributes to the achievement of the individual ones nor does it help the agent to coordinate his own actions. So, according to Yaffe, agglomerativity as closure demands that one combines intentions that do not need to be put together. Since the combination itself takes a certain amount of time and effort, it should not be required of normal, i.e. resource-bounded, agents.

In view of this he proposes the following alternative interpretation of the principle, which I called *agglomerativity against potential irrationality*: it is irrational to intend \( A \) and to intend \( B \) if the intention to do both would be irrational according to some other norms of rationality for intentions. According to this second interpretation planning agents are no longer required to combine arbitrary intentions. Rather, they are required to do so only to the extent that this “makes conflicts evident to themselves, when there is doubt as to the rationality of the conjunctive intentions”. The conflicts mentioned here are conflicts “with other norms of rational intentions”. Given what I have said so far, this means that having two intentions is irrational if their combination would result in an intention whose content is contradictory or impossible to realize given the agent’s beliefs\(^4\).

Observe that, understood that way, agglomerativity parasitizes, so to speak, its rationality demands on these two other norms. In the case of internal consistency, for example, having the intention to do \( A \) and the intention to do \( B \) is irrational to the extent that having the intention to do \( A \) and \( B \) is internally inconsistent. Along the same lines, having the intention to do \( A \) and the intention to do \( B \), given the belief that \( A \) and \( B \) cannot be achieved together, is irrational to the extent that having the intention to do \( A \) and \( B \) is strongly belief-inconsistent.

It is thus no coincidence that the former case resembles what I have called above “internal inconsistency of plans”. Plans were called internally inconsistent precisely when a pair of their elements could not be achieved together\(^5\). In other words, internal consistency of plans is a particular case of agglomerativity against potential irrationality, preventing the agent from holding pairs of intentions which would violate internal consistency of intentions if they were put together.

It also worth noting that, unlike agglomerativity as closure, agglomerativity against potential irrationality does not require a systematic combination of inten-

\(^4\)Recall that I consider means-end incoherent plans as plans where the means-intentions are missing. But agglomerativity as potential irrationality is about pairs of intentions the agent already has. In view of that, a violation of that requirement that resorts on means-end coherence seems to involve what I have called strongly means-end incoherent plans. But, as I said, it seems that these cases are, in turn, reducible to violation of some other norms. For that reason I only consider cases where agglomerativity as potential irrationality leads to a violation of internal or strong belief consistency.

\(^5\)It is, in that respect, telling that Yaffe does not consider internal consistency of plans as a separate requirement.
6.2. Cognitivism

In this section I present the cognitivist view, according to which the norms just presented ultimately stem from norms of theoretical rationality associated with beliefs.

---

6I draw this appellation from Savage [1954]. My reservations with respect to Yaffe’s parallels the difficulty of explaining how agents set the “size” of their worlds, i.e. of their decision problems.

7There are of course interrelations between these various norms. See again Section 2.2.
Chapter 6. Hybrid pragmatism, acceptances and norms on intentions

The keystone of cognitivism is a postulated connection between intentions and beliefs. I present two ways to understand this connection. Most of this section is then devoted to seeing how well they support the derivation from the theoretical norms on beliefs to the practical norms on intentions. All of this, of course, rests on a particular philosophical theory of beliefs. In the previous chapters we could do most analyses without diving into such conceptual details. For the present chapter, though, they become crucial, especially in distinguishing between beliefs and acceptances in deliberation.

6.2.1 The functionalist view on beliefs

Many authors who have written on the relation between practical and theoretical rationality take good care to distinguish between probabilistic and flat-out [Bratman, 1987] or all-or-nothing [Harman, 1986] beliefs. They differ mainly in that the first, in contrast to the second, comes in various degrees. Probabilistic beliefs echo subjective probabilities or endogenous uncertainty, which I mentioned in the Introduction (Section 1.1.1). Flat-out beliefs, on the other hand, are attitudes that an agent either has or not. In that respect, they are close in behaviour to the qualitative representations of information that I used in most of the previous chapters.

Whether probabilistic or flat-out, the theory of beliefs that underlies most cognitivist accounts is functionalist. Just as intentions, beliefs are characterized through their “actual and potential, or typical, causal relations to sensory stimulations, behavior and other mental states” [Schwitzgebel, 2006].

A belief that \( p \) is thus viewed as an attitude of regarding \( p \) as true that:

1. dispose the subject to incorporate \( p \) into further practical and theoretical reasoning.

2. is “formed, revised and extinguished—or [...] for short [...] regulated for truth”, in the sense that it is responsive to evidence and reasoning.

3. is correct if and only if \( p \) is the case.

Once again following Schwitzgebel [2006], one can think of the first two conditions as respectively specifying the backward-looking and forward-looking conditions for an attitude to functionally count as a belief. I used this idea to describe

---

8It is, however, unclear whether flat-out beliefs match the subject matter of epistemic logic. For one thing, epistemic logic is often thought of as the study of sure beliefs, which correspond in turn to probabilistic beliefs of degree 1 or 0. But it is often stressed, e.g. by Bratman [1991], that flat-out beliefs are not reducible to beliefs with degree 0 or 1. I do not go into this distinction in detail, for it is orthogonal to my present concern. I simply proceed, unless explicitly stated, with the two notions of belief.

9This characterization is taken almost textually from Shah and Velleman [forthcoming, p.2-3].
6.2. Cognitivism

Intentions in the Introduction (Section 1.2). Intentions are *outputs* of practical reasoning (backward-looking) which typically also play a certain role as *inputs* to deliberation (forward looking).

Condition (2) is backward-looking in the sense that it points to processes that *affect* beliefs. They are created, revised or abandoned according to the input of new evidence or conclusions reached from theoretical reasoning. This is a *constitutive* claim about what beliefs are.

This should be contrasted with condition (3), which gives a criterion for correctness, and which is thus essentially *normative*. It states that beliefs are correct only to the extent that they are true, i.e. that they fit the world\(^\text{10}\). In the words of Austin [1953] and Searle [1983], beliefs have the “mind-to-world” direction of fit.

Conditions (2) and (3) sharply distinguish beliefs from *practical* attitudes like intentions. Intentions can be formed, revised and abandoned in response to changes in desires or preferences, and their correctness is more often than not a matter of instrumental rationality. Cases of beliefs formed or held for practical reasons are, on the other hand, not typical and are usually viewed as a form of wishful thinking. That is, these are pathological cases, i.e. incorrect ways to hold belief. For that reason, beliefs are often said to belong to the realm of *theoretical* rationality.

This is not to say, of course, that they do not take part in practical reasoning. Condition (1) makes this clear. It is “forward-looking” because it points to typical processes that take beliefs as *input*. Although there might be other forward-looking conditions that characterize beliefs, the disposition to be incorporated into practical and theoretical reasonings is the most important for our present concerns\(^\text{11}\).

### 6.2.2 Constraints on rational beliefs

I have already introduced a normative component in the functionalist definition of beliefs, but the cognitivist derivation of the practical norms on intentions does not explicitly use it. Instead, it appeals to three other normative requirements on beliefs, which I now present.

\(^{10}\)Here I follow Shah and Velleman [forthcoming] and include the normative claim in the definition of beliefs. This is by no means an uncontroversial practice, but it will prove useful in distinguishing beliefs from acceptances in deliberation.

\(^{11}\)It is worth noting that one even finds characterization of beliefs only in terms of potential input into practical reasoning. Holton [1994, p.68], for example, says that “your belief that a certain thing will happen is just the disposition that you acquire when you work the supposition that it will happen into your plans.” Similar remarks can be found in Schwitzgebel [2006] and Alonzo [forthcoming]. Note, furthermore, that this view of belief quite nicely matches the Bayesian approach to belief that is inherent in representation results in decision theory. See the references in the Introduction (Section 1.1.1) and Joyce [2004].
Internal Consistency of Beliefs. Just as with intentions, one can distinguish two senses of this requirement. Internal consistency within beliefs requires the content of beliefs to be consistent. An agent should not believe contradictions. Internal consistency between beliefs, on the other hand, asks the different beliefs of an agent to be consistent with each other. As was the case with intentions, internal consistency between beliefs follows from internal consistency within beliefs, together with agglomerativity, which I shall introduce shortly. For that reason, I use the plain “internal consistency of beliefs” to refer to internal consistency within beliefs.

Agglomerativity of Beliefs. Again, one can think of agglomerativity as closure under conjunction or as a safeguard against potential violations of other norms on beliefs. In the case of intentions, this second interpretation of the principle was proposed as an alternative to the first one, which Yaffe [2004] found too strong. Agglomerativity as closure is less controversial for beliefs. Here is Velleman [2003, p.18] on the subject:

Beliefs are agglomerative because they aim to fit the world, of which there is just one, in whose complete characterization the contents of all true beliefs are conjoined. The rational pressure to conjoin beliefs is a pressure to fuse them into a single characterization of the single world that all of them aim to fit.

As this last sentence suggests, there seems to be a relation between the standard of correctness for beliefs mentioned in Section 6.2.1 and the fact that they are agglomerative. Beliefs have to fit the world, and for that reason it seems that there is a rational requirement to agglomerate them.

One can obtain agglomerativity against potential irrationality of beliefs if they are closed under conjunction. This is so because, for beliefs, agglomerativity against potential irrationality really amounts to agglomerativity against potential inconsistency\textsuperscript{12}. An agent whose beliefs are closed under conjunction and who has “potentially inconsistent” beliefs would turn them directly into an internally inconsistent conjoined belief. For that reason, I take “agglomerativity of beliefs” to mean only agglomerativity as closure.

Explanatory Coherence. The last requirement, which I call “explanatory coherence” [Harman, 1986], comes from what Schwitzgebel [2006, Section 3.2] calls the “holistic” view on beliefs. In normative terms, it requires a rational agent to maintain “relations of immediate coherence or intelligibility” among his beliefs [Harman, 1986, p.75]\textsuperscript{13}. That is, given one particular belief of an agent,

\textsuperscript{12}In Section 6.1 I made a proviso regarding means-end coherence. The same applies to explanatory coherence, which I introduce next.

\textsuperscript{13}Holism about beliefs also has a constitutive counterpart. See again Schwitzgebel [2006, Section 3.2] for more explanations and references.
6.2. Cognitivism

one should be able to locate other beliefs that “explain” it\footnote{I leave open the question of what should be necessary or sufficient to count as an explanation here. See the illuminating discussion of Harman [1986] and the papers cited in the footnote on page 138. I return briefly to this issue in Section 6.3.}. Given that beliefs are responsive to evidence, this means that a belief that \( \phi \) should be backed, so to speak, by a belief about evidence in favour of \( \phi \).

6.2.3 Weak, intermediate and strong cognitivism

There are thus three norms on rational beliefs: internal consistency, agglomerativity and explanatory coherence. For the cognitivists the requirements on intentions that I presented in Section 6.1 can be derived from these requirements on beliefs. This derivation typically rests on the assumption that intentions “involve” beliefs. One can classify the various cognitivist derivations according to their view on how strong this involvement is. More precisely, Bratman [2006b] distinguishes three strengths of cognitivism, which I now present. Even though only the last two will be of interest hereafter, a close look at the first one will help to clarify issues.

The requirement of strong belief consistency is already an assumption on how intentions involve beliefs, but it is a negative involvement, so to speak. It states that having the intention to do \( A \) implies not believing that one will not do \( A \). Weak cognitivism holds something slightly stronger: if an agent intends to do \( A \) he must consider it possible that he will do \( A \)\footnote{This is stronger than strong belief consistency as long as we assume that not believing that not \( \phi \) is not equivalent to considering it possible that \( \phi \). Recall that this equivalence holds, for knowledge, in all the epistemic models that I used in the last chapters.}. In other words, to have the intention to do \( A \) implies believing that it is possible that one will do \( A \). This belief can be either flat-out, as advocated e.g. by Wallace [2003a, 2006], or probabilistic, as Chan [1999] takes it. In both cases, however, this belief is compatible with the belief that it is possible that one will not do \( A \), and both can be simultaneously and consistently incorporated into practical reasoning as partial beliefs. That is, the agent can work into his further deliberation the fact that he might and that he might not do \( A \).

I call the second type of cognitivism intermediate. It holds that having the intention to do \( A \) implies believing that one will do \( A \). This, of course, entails weak cognitivism, but not the other way around. As I take it, the key idea underlying intermediate cognitivism is that intending to do \( A \) implies, first and foremost, being disposed to work in flat-out the fact that one will do \( A \) in further planing and, second, that this assumption is regulated by truth and is responsive to evidence. That is, it seems that intermediate cognitivism is not compatible with incorporating the fact that one might not do \( A \) in practical reasoning, once one intends to do \( A \). Or at least this is what I take intermediate cognitivism to mean: intending to do \( A \) implies believing that \( A \) in the sense of being disposed...
to use the fact that one will do $A$ in deliberation, in a way that is regulated by truth and is responsive to evidence.

Except for Harman [1986], very few authors have directly argued for the intermediate cognitivist view. The most popular cognitivist standpoint is rather stronger. According to e.g. Harman [1976], Velleman [2005] and Setiya [2007], intending to do $A$ is the same as having a special kind of belief that one will do $A$. I call this strong cognitivism. Again, the kind of belief involved here seems to be essentially flat-out, at least as far as integration into practical reasoning is concerned. Strong cognitivism thus holds that to have the intention to do $A$ is nothing else than being disposed to work in the fact that you will do $A$ in your practical reasoning, in a way that is responsive to evidence and regulated by truth\footnote{The authors mentioned above usually add special conditions to the way these intention-as-beliefs are responsible to evidence and regulated by truth. I leave these details aside.}.

6.2.4 Rational intentions from rational beliefs

I now show how cognitivists derive the norms on intentions from the norms on beliefs. I look at each of the derivations in some detail, for they will serve as a landmark for the “hybrid” derivation that I present later.

Internal and strong belief consistency of intentions. These two norms are the easiest to derive for the cognitivist. Let me look first at internal consistency of intentions. Suppose one intends something contradictory, say to do $A$ and not to do $A$. By the intermediate cognitivist assumption, one then must believe that he will do $A$ and that he will not do $A$, which violates internal consistency of beliefs. Putting back this contrapositive argument in its normal direction, it shows that internal consistency of beliefs and the intermediate cognitivist assumption together imply internal consistency of intentions.

The argument for strong belief consistency goes along the same lines. An intention that is not feasible given the agent’s background beliefs will generate a new belief that is, by assumption, inconsistent with this background. But then, using agglomerativity, one gets a belief that is internally inconsistent\footnote{Observe that one could have argued directly from internal consistency between beliefs.}.  

Means-end from explanatory coherence. To derive means-end coherence of intentions from explanatory coherence of beliefs, cognitivists usually make one more assumption about what counts as evidence for an agent that he will act in a certain way. As I have mentioned many times now, beliefs are taken to be responsive to evidence. But in the case of beliefs about what one will do, it seems that we have a special source of evidence\footnote{I am using “evidence” rather sloppily here. For in-depth discussions about the epistemology of agency, see Anscombe [1957], Faley [2000] and Velleman [2005].}, namely the intentions themselves.
Cognitivism derive means-end coherence of intentions from explanatory coherence of beliefs as follows. Suppose that one has means-end incoherent intentions. He intends the end \( E \) but he does not intend any of what he believes are means \( M \) that he must come to intend to achieve \( E \). What is more, he does not intend to form such means-intentions later. By the intermediate cognitivist assumption, this means that this agent must believe that he will do \( E \). But since he also believes that he will do \( E \) only if he does something in \( M \), he should believe that he will do some \( M^{19} \). But since he does not intend to do anything in \( M \), and he does not believe that he has such intentions, it seems that he lacks the required evidences to back up this new belief, so to speak. In other words, the beliefs of the agent are explanatory incoherent. It thus seems that explanatory coherence, together with the fact that beliefs about intentions count as evidence for beliefs about what an agent will do, imply means-end coherence.

This derivation is unfortunately not sound. It slips too quickly from one not having the required means-intentions to one not believing that one has these intentions. As pointed out by Brunero [forthcoming], Bratman [2006b] and Wallace [2006, 2003a]), the latter can come without the former. Here is Bratman [idem]:

Suppose I intend \( E \) and know that \( E \) requires both \( M \) and that I intend \( M \). If I still do not intend \( M \) my intentions suffer from means-end incoherence. But suppose that, while I in fact do not intend \( M \), I nevertheless falsely believe that I intend \( M \). So my beliefs are that \( E \), that \( E \) requires both \( M \) and that I intend \( M \), that I intend \( M \), and that \( M \). There is no incoherence (though there is falsity) in this structure of beliefs. So means-end coherence is not belief coherence.

Bratman points out that an agent can falsely come to believe that he intends something and so he has coherent beliefs, but nevertheless incoherent intentions. To carry the above derivation through, cognitivists thus need not only to assume that the intentions are means-end coherent, but also that the agent is not mistaken about what he intends.

This, according to Wallace [2003a, p.21], cannot be assumed without some additional rationality constraints: “theoretical constraints on rational beliefs can get you as far as the belief that you intend to do [something]; to go beyond that, to a rational requirement that you form the [required means-intentions], we need an additional principle [...]”, a principle that is independent of explanatory coherence of beliefs.

The principle that Wallace has in mind is a pragmatic one. He holds that to have true beliefs about one’s own intentions is, “in deliberative contexts where [means-end coherence] is relevant, [...] an executive virtue, to be included among...”

\(^{19}\)This step assumes, of course, something like an inference closure principle for beliefs.
the traits and capacities that make us, in general, effective in the pursuit of our
goals” or “is a strategy that enhances our ability to realize the broader aims
that are given with our nature as deliberating agents” [Wallace, 2006, p.119, my
emphasis]. That is, having a false belief about one’s own intentions, when that
belief features in the cognitive background of deliberation, seems to threaten the
overall prospect of reaching one’s ends. Bluntly, being wrong about what you
intend is, in the long run, not good.

This is an important shift in Wallace’s argument, because he thereby steps
outside the pure cognitivist enterprise. In his view, it is not “plausible to suppose
that the whole field of instrumental rationality can be explained in terms of the
requirements of coherence in beliefs” [idem]. This recourse to the principle of
practical rationality will also be very important later on. I argue in Section 6.4
that this is precisely the kind of hybrid justification that comes out of using
acceptances instead of beliefs to derive the norms on intentions.

This is looking too far ahead, however: for now what matters is to observe that
Bratman [2006b, p.13] has offered a counterexample to the thesis that mistaken
beliefs about one’s own intentions are irrational. He pointed out that “to reflect
carefully on all that one intends […] is an activity that takes time and uses other
resources, and one may well have better things to do.” Bratman considers, for
example, an agent who believes that he intends to go shopping on Thursday while,
in fact, he intends to go on Friday. A week before, it seemed perfectly rational
for this agent not to reflect on the accuracy of this belief because “other matters
are more pressing right now” [idem].

Bratman’s example is rightly aimed at showing that it is not always irrational,
practically speaking, to have false beliefs about one’s intentions. But Wallace
seems to have two ways around it.

First, he crucially uses the notion of relevant deliberations. It is notable that,
in Bratman’s counter-example, the agent is correct about the fact that he intends
to go shopping, but he is mistaken as to when he intends to go. Now, the essence
of the case is that it does not seem irrational to have an incorrect belief as long as
its accuracy is not relevant to further deliberations. To see this, suppose that the
agent intends to go shopping in order to buy some camping gear that he needs
for a hiking trip the week after. The fact that he is mistaken about when he
actually intends to buy the gear does not threaten his hiking plans. On the other
hand, this mistaken belief becomes more problematic if he is to plan his Thursday
evening. Whether or not it is irrational to have such mistaken beliefs depends
on the context of deliberation, and this is precisely what Wallace seems to point
out when he focuses on “deliberative contexts where [means-end coherence] is
relevant […]”.

Second, Wallace also holds that being mistaken about what one intends is
irrational because it threatens the overall prospect of achieving one’s own inten-
tions. In other words, not all instances of such mistaken beliefs have to mess up
the agent. Rather, such beliefs are irrational because they embody, so to speak,
a bad deliberation habit. In this context, is not so devastating that there exist cases of false beliefs about one’s own intentions which are not irrational.

These two replies to what we may call the “problem of false beliefs” will also be very important later, because they apply even more naturally with acceptances in deliberation. But for now the reader should bear in mind that cognitivism has difficulty in accounting for means-end coherence. There seems to be a way to derive this norm on intentions from explanatory coherence of beliefs, if one is willing to use an additional pragmatic principle. But, as we shall see presently, means-end coherence is not the only norm that resists a purely cognitivist derivation.

**Agglomerativity.** Let us look first at agglomerativity against potential irrationality. As this requirement imports part of his rationality demands from other norms on intentions, it should not come as a surprise that cognitivism can explain it to the extent that it can explain the others. There are, more precisely, two cases to consider. Suppose first that an agent has the intention to do A and the intention to do B and that the combination of these into the intention to do A and B would be internally inconsistent. Clearly, this pair of intentions would be irrational because the combined intention would generate an internally inconsistent belief, given the intermediate cognitivist assumption. An entirely similar argument covers the case where the combination of A and B would generate a strongly belief-inconsistent intention. So cognitivism can easily explain agglomerativity as potential irrationality, simply because it can explain internal and strong belief consistency.

The case of agglomerativity as closure is more problematic. Strong cognitivism can of course explain it, given agglomerativity of beliefs. After all, to hold that beliefs are agglomerative is just the same as to hold that intentions are also agglomerative, once one views intentions as a special kind of beliefs.

But observe that the argument does not go so straightforwardly for intermediate cognitivism, even if one assumes agglomerativity of beliefs. Suppose an agent has the intention to do A and the intention to do B, but not the intention to do A and B. From the assumption underlying intermediate cognitivism, together with belief agglomerativity, we can conclude that the agent believes that he will do A and B. But what is supposed to be wrong with having this belief while not having the corresponding intention? Note that he does not have the intention not to do A and B, which would simply make his beliefs internally inconsistent.

One possible way out would be to use explanatory coherence again. One would thus say that the belief that one will do A and B wants an explanation, which the belief that one has the intention to do A and B would provide. But why can the beliefs that one intends to do A and the belief that one intends to do B not together provide the required explanation? If they do, then we are still missing

---

20 Note that one could also argue directly from these two intentions, the intermediate cognitivist assumption and agglomerativity of beliefs.
a intermediate cognitivist argument for agglomerativity.

What is more, as we have seen for means-end coherence, the recourse to explanatory coherence of beliefs makes one vulnerable to the problem of false beliefs about intentions. In the present context, the problem can be rephrased as follows. If an agent can be mistaken about his own intentions, i.e., can have false beliefs about what he intends, then his belief that he will do $A$ and $B$ can be explained by the (false) belief that he intends to do both.

In view of all this, it seems that intermediate cognitivism can explain internal consistency, belief consistency and agglomerativity as potential irrationality. It has difficulties with means-end coherence, because of the problem of false beliefs about one’s own intentions. Strong cognitivism, however, explains agglomerativity as closure.

6.2.5 A general concern against cognitivism

The shortcomings of the cognitivist derivations of agglomerativity and means-end coherence are in themselves enough to motivate the search for an alternative. But Bratman [1987] has famously cast doubts on the basic assumption that intending to do $A$ implies believing that one will do $A$. According to him, this assumption rules out some plausible cases of agents who apparently have the intention to do $A$ while not believing that they will do $A$. He explains:

[In [Bratman, 1987]], it seemed to me plausible that I might, for example, intend to stop at the bookstore on the way home even though I know that, once I get on my bike I tend to go on automatic pilot, and so even though I do not, strictly speaking, believe that I will stop (though I do not believe that I will not stop). So I thought it best not to tie the theory of intention and planning to such a strong belief condition. [Bratman, 2006b, p.3]

The “strong belief condition” that Bratman writes about here is what I have called intermediate cognitivism. Given what I said in Section 6.2.3, what is at stake is whether the agent believes flat-out that he will stop at the bookstore. For a cognitivist, if this agent really does not have such a flat-out belief then we cannot say that he has the intention to stop at the bookstore. Bratman, on the other hand, thinks that the agent can genuinely intend to do so even if he does not have this flat-out belief.

Recall that an agent counts as having such a belief if he regards the fact that he will do $A$ as true in a way that:

1. disposes him to incorporate the fact that he will do $A$ into further practical and theoretical reasoning.
2. is regulated by evidence in favour of him going to do A.

3. is correct if and only if he will in fact do A.

It seems to me that the only thing to deny is the fact that the agent, from having the intention to do A, is automatically disposed to work the assumption that he will do A into his theoretical reasoning. To be sure, one cannot deny that the agent regards the fact that he will do A as “true”. This agent is disposed to incorporate this fact into practical reasoning. What is more, as we have seen in Section 6.2.4, having the intention to do A can be seen as an evidence for the fact that one will do A, especially in the context where the intention at hand is future-directed. There is not yet a “fact of the matter” that can settle the correctness of this attitude. Given all this, the only thing left to deny is the disposition to incorporate “I will do A” into theoretical reasoning. This reading is supported, I think, by the original formulation of the absent-minded cyclist example:

I might intend now to stop at the bookstore on the way home while knowing of my tendency towards absentmindedness—especially once I get on my bike and go into “automatic pilot.” If I were to reflect on the matter I would be agnostic about my stopping there, for I know I may well forget. It is not that I believe that I will not stop; I just do not believe I will. [Bratman, 1987, p.37, my emphasis]

The reflection that is mentioned here seems to be essentially truth-oriented, in the sense that what is at stake is whether the agent will, in fact, stop at the bookstore. In other words, the agent deliberates about the truth of “I will stop at the bookstore” and not about, for example, whether he should stop or how he would make it happen.

In view of all this, I will take the general concern about cognitivism to be the following. Intending to do A seems to come with an attitude of regarding “I will do A” as true, but this attitude is not quite a belief. In particular, it does not dispose the agent to incorporate this fact into theoretical reasoning. The overall aim of this chapter is indeed to see whether the normative requirements on intentions can be derived if one thinks of this attitude not as a belief but rather as an acceptance in deliberation.

\[21\] Witness the discussion in Bratman [1987, p.37-39].

\[22\] It is interesting to note, before going further, that Bratman [2006b] does not take this route. As noted in the introduction of this chapter, he proposes a justification of the norms on intention that altogether bypasses the recourse to cognitive-like attitudes, whether beliefs or acceptances. For him norms on intentions stem from the general “aim to achieve what is intended” and of the “projected unity of agency”. Thus the appellation “agency approach” to describe Bratman’s approach.
6.3 Acceptances in deliberation

In this section I first introduce the concept of “acceptance in deliberations”, which bases the alternative derivation of the norms of intentions. After that I briefly present the companion idea of “adjusted background of deliberation”, which will be important hereafter, when I look at the normative requirements that apply to acceptances.

6.3.1 Accepting or taking for granted in deliberation

In the context of practical reasoning, some authors\textsuperscript{23} observed that there is more than beliefs in the “cognitive background of deliberation”. We sometimes intentionally accept or take for granted some facts about the world, even though we neither believe them with degree 1 nor flat-out. These phenomena are called “acceptances in a context”. In what follows I am specifically concerned with the role of these attitudes in practical reasoning. For that reason I refer to them as acceptances in deliberation, or simply as acceptances. I should mention that Shah and Velleman [forthcoming] talk about acceptances as the general attitude of regarding something as true. For them, beliefs, assumptions and even images are all specific kinds of acceptances. As we shall soon see, the acceptances that I am concerned with, acceptances in deliberation, are a specific kind in this general category of epistemic attitudes.

Acceptances in deliberations are not only regulated by truth and responsive to evidence, but also also “regulated for practice” and responsive to “pragmatic considerations” [Alonzo, forthcoming]. This difference is best illustrated by an example.

In planning my day—a June day in Palo Alto—I simply take it for granted that it will not rain even though I am not certain about this. If I were instead figuring out at what odds I would accept a monetary bet from you on the weather I would not simply take it for granted that it will not rain. But in my present circumstances taking this for granted simplifies my planning in a way that is useful, given my limited resources for reasoning. [Bratman, 1991, p.22]

Even though the agent, in this case, incorporates the fact that it will not rain into his practical reasoning, he does it in a peculiar way. Observe first that he

\textsuperscript{23}Chiefly Bratman [1991]. For a congenial but nevertheless different characterization of acceptances, see Grice [1971], [Cohen, 1989] and [Engel, 1998]. Holton [1994] has also studied related phenomena in the context of thrusting relations. Harman [1976, p.438] already spoke of “taking for granted” in relations with intentions, but the appellation “acceptances” seems to come from Williams [1973]. Note that some authors—e.g. Holton [item] and Alonzo [forthcoming]—use “reliance” instead of acceptance to talk about what appears to be the same attitudes. Engel [1998, p.148] finally remarks that one finds discussion of “pragmatic beliefs” already in Kant, in a way that is very close to what will be described as acceptances.
6.3. Acceptances in deliberation

does not plan with the idea that the chances of rain are extremely low, even though one could argue that this is what he really believes. He plans with the plain assumption that it will not rain. Observe too that what really triggers the incorporation of this fact into the agent’s planning is a pragmatic concern. It simplifies deliberation. These seem to be the key features of acceptances. They can be at variance with the agent’s beliefs and are responsive to pragmatic considerations.

In what follows I shall thus take an acceptance that $p$ in a given deliberation to be an attitude of regarding $p$ as true that:

1. Incorporates $p$ into that very deliberation.

2. Is regulated by either the truth of $p$ or the pragmatic context of a given deliberation, in a way that is responsive to evidence or practical concerns.

3. If both $p$ is not true and the agent is not better off by accepting $p$, then the acceptance that $p$ is not correct.

Acceptances thus share with beliefs the part of the “forward-looking” (see Section 6.2.1) functional characterization. They are both states that are crucially incorporated into practical reasoning. But I do not take acceptances in deliberation as featuring typically in theoretical reasonings, even though some authors, e.g. Stalnaker [1984], have studied closely related phenomena in that context. I take them to be specifically tailored for practical contexts.

Acceptances and beliefs differ in the way they are regulated, i.e. formed, changed or discarded. Condition (2) states that acceptances can be regulated both by evidence and practical concerns. Similarly, condition (3) states that practical concerns also come into play to assess the correctness of acceptances. Here I deliberately leave unanswered the question whether there is a systematic relation between the evidential and practical inputs, for example whether one has priority on the other. For what follows, it will be enough to know that acceptances are responsive to both.

I think that to provide sufficient conditions for correctness one would have to be more precise about this relation, especially to handle properly the relation between bracketed beliefs and acceptances (see below). For that reason (3) only specifies necessary conditions for correctness.

Such a more precise stance could go as follows. Following Alonzo [forthcoming], I find it plausible to think that the pragmatic considerations that justify acceptances are constrained by evidence. That is, it seems that truth somehow has precedence over practice in terms of the correctness of acceptances. One could propose to make this precedence more explicit, namely in lexicographical terms. Condition (3) would then state that an acceptance that $p$ is ultimately correct if $p$ is the case. But in case the truth of $p$ is not (yet) a settled matter then, and only then, can practical considerations enter the assessment of whether accepting that $p$ is correct. This will happen when $p$ is about something in the future, as in the example above, and especially about future actions. As observed again by Alonzo [forthcoming], this could help to distinguish acceptances from simple wishful thinking. Acceptances should respond to evidence in a way that wishful thinking does not. If there is any way that the later can be justified, it seems that this will be solely by practical concerns.
Observe that, in the last example, it is the acceptance in itself that seems to be practically useful. The practical consequences of the fact that it will rain do not really influence how one assesses the acceptance. This is what condition (3) makes precise: the pragmatic considerations should bear on the acceptance itself, rather than on its content. Pragmatic arguments that justify accepting that \( p \) are arguments that show that the agent can be better off by incorporating \( p \) into his practical reasoning, and not necessarily that the agent would be better off if \( p \) were the case. These can go together, but they need not\(^{25}\).

The simple fact that practical considerations regulate and take part of the standard of correctness for acceptances suffices to distinguish them from beliefs. But this difference also shows in the context-dependency of these two attitudes. Bratman [1991] has strongly emphasized that the rationality of acceptances depends in general on the context of deliberation. An agent may rationally take some facts for granted in some deliberations while not taking them for granted in others. This is not the case for beliefs, either probabilistic or flat-out. Whether a belief is justified depends on the evidences that support it, not on the specific deliberation in which it is applied. To appreciate this difference, look again at Bratman’s example. While planning his day, it seems perfectly justified for the agent to take it for granted that it is not going to rain. But it is also justified for him to abandon this acceptance when he has to decide how much he would bet on that very fact, even though the matter remains unsettled in both cases. To put it differently, it would not be rational for the agent to change his belief about the weather from one deliberation to the other, unless he receives new information in the meantime. But the change in acceptance seems perfectly justified. This, [Bratman, 1999, chap.1] argues, shows that acceptances are not the same as beliefs.

It should be noted that an agent can decide, to a certain extent, which facts he takes for granted. Following Bratman [1991, p.29], we can distinguish two ways in which an agent can take a fact \( A \) for granted: *positing* and *bracketing*. The first occurs when the agent decides to take \( A \) for granted even though the agent is uncertain about it. This is the case I had in mind in most of the previous discussion. \( A \) gets temporarily promoted, so to speak, from mere possibility to “hard” fact. Bracketing occurs, on the other hand, when an agent believes that \( A \) but decides not to work this assumption into his deliberation. Unfortunately, Bratman is not very explicit about bracketing. It seems to concern mainly flat-out beliefs or those with degree 0 or 1. In these cases, one can imagine that the agent decides to plan as if the possibility of not-\( A \) was serious, even though he does not believe so. For flat-out beliefs, note that this crucially uses the fact that the agent

\(^{25}\)Take for example a chess player who takes for granted that his opponent will play very well, without having any conclusive evidence for that. We can easily imagine that this acceptance makes the player better off. He will plan several moves ahead and play more cautiously, which might make him more likely to win. But observe that if, in fact, the opponent does play well, then the chances of winning for the player seem to be actually diminished.
can decide what he takes for granted. Flat-out beliefs were indeed characterized as disposing the agent to include their contents into practical reasoning. By bracketing the agent explicitly overrides this disposition.

### 6.3.2 The adjusted background of deliberation

At the beginning of this section I introduced acceptances by way of the \textit{(adjusted) cognitive background of deliberation}. Bratman [1991, p.29] put forward this concept to stress that the information invoked in practical reasoning is different from the collection of the agent’s beliefs, which he calls the \textit{default} cognitive background. The cognitive background of deliberation is “adjusted” precisely because it contains acceptances resulting from positing or bracketing elements of the default background.

I remarked earlier that acceptances and beliefs have a similar forward-looking functional characterization, at least as far as practical reasoning is concerned. They are attitudes that feature in the adjusted cognitive background of deliberation. The adjusted cognitive background of deliberation thus differs from the “default” background in that it features contents of acceptances and of beliefs.

The remarks at the end of the previous section suggest that agents can somehow build the adjusted background of deliberation by positing and bracketing some facts. But agents with limited time and resources can only make a small number of such explicit decisions\textsuperscript{26}. Should we take, then, the adjusted cognitive background to contain only the facts that are explicitly taken for granted, or should it also include other elements of the default background? According to Bratman [1991, p.30, my emphasis], in a particular deliberation, “if one has a relevant all-or-none, context-independent belief that $p$, and this belief is not bracketed, then one accepts $p$ in that context. And similarly concerning a context-independent degree of confidence of 1.” Observe that, without the emphasized notion of relevance, the adjusted cognitive background becomes rather large. It includes what is explicitly taken for granted and all the default background beliefs that can be added to it—consistently, as we shall see. The notion of relevant beliefs is aimed at limiting the size of the adjusted background. I return on the notion of relevance in Section 6.4.3. It is enough for now to see that the adjusted background might not decide on every single fact.

### 6.3.3 Requirements on the background of deliberation

There seem to be norms of rationality that apply to the constituents of the adjusted cognitive background, regardless of whether they are acceptances or beliefs. As we shall see, these general norms on the adjusted background mirror the norms

\textsuperscript{26}The ideal agents that I have studied in the previous chapters do not, of course, have such constraints. I come back to this in Section 6.5, the conclusion of the present chapter.
on beliefs that I presented in Section 6.2.2. This can be seen as a consequence of the fact that beliefs and acceptances play essentially the same role in deliberation. Paraphrasing Velleman [2003, p.18], one can say that the adjusted cognitive background aims at picturing the world, of which there is just one, where the various courses of action considered shall be pursued. This can be seen as the starting point of an argument for the various norms that I am about to present. But I will not go into such a justification. My goal here is rather to present the norms on the adjusted background, from which follow most norms on acceptance, and to show how can one derive the norms on intentions from them.

Closure under logical operations. I first assume that the cognitive background of deliberations should be “logically” closed. I do not go into much detail about which logical rules of inference I mean here. For what follows I “only” suppose that the adjusted background should be closed under classical implication and conjunction. This means that if I accept in a given deliberation both that by doing $A$ I also do $B$, and that I will do $A$, then I should also incorporate in my deliberation the fact that I will do $B$. Similarly, what an agent includes in the background should agglomerate. If $\phi$ and $\psi$ feature in the adjusted background of a deliberation, then their conjunction should, too.

Obviously, agglomerativity of acceptances follows from this general closure requirement. If, in a particular deliberation, an agent takes $\phi$ for granted and takes also $\psi$ for granted, he should take their conjunction for granted. But the agent is no more required to hold to this agglomerated acceptance in other contexts than he is with respect to the conjuncts. I take the same to apply when one of the conjuncts comes from a context-independent belief and the other from an acceptance. In this case the conclusion is also restricted to the given deliberation. Agents are required to carry agglomerated cognitive attitudes from one deliberation to another only when this operation can be done in the default background, as for example when two beliefs agglomerate\(^{27}\).

To follow up on the remark at the end of Section 6.3.2, it is worth noticing that this closure requirement makes the adjusted cognitive background considerably larger. It does not, however, makes it complete in the sense of deciding on all facts.

Internal consistency. Beliefs were required to be internally consistent, and I assume that this is also the case in general for the elements of the adjusted cognitive background. Here I have in mind internal consistency in a strict sense: an agent should not include plain contradictions into the background of his deliberation. Internal consistency of acceptances is, of course, a special case of this general norm of consistency.

\(^{27}\)This applies more generally to all “conclusions” that are reached via the closure requirement.
6.3. Acceptances in deliberation

Another important consequence of internal consistency, together with the logical closure requirement, is a context-dependent form of belief consistency for acceptances. What is taken for granted should be consistent with the “believed” facts that are imported into the adjusted background. Otherwise, by the closure requirement, one would quickly obtain a new internally inconsistent acceptance. The strength of this new requirement, of course, depends on which beliefs are actually carried in the adjusted background. But even without being specific about which beliefs should be imported, this belief consistency requirement precludes an agent from bracketing flat-out beliefs that are explicitly invoked in practical reasoning. This seems to be in line with condition (3) of the definition of acceptances. An agent who is convinced that $\phi$ and is ready to use it in a particular context cannot rationally take for granted that $\phi$ is not the case in that same context.

Observe that one may require a stronger form of belief-consistency for acceptances, namely that they should be consistent with all non-bracketed flat-out and probabilistic beliefs of degree 0 or 1, regardless of whether they feature in the adjusted background or not. This requirement does not follow from internal consistency and closure of the adjusted background, but I do not need it to carry out the derivation of belief consistency of intentions, and so I leave it open whether one should include it in a theory of acceptances.

Explanatory coherence. Explanatory coherence of beliefs is mirrored in their standard of correctness (see Section 6.2.2). Agents were required to hold beliefs about evidence, but the adjusted background of deliberation also contains acceptance, the standard of correctness of which is different from that of beliefs. I shall thus use the following generalization of explanatory coherence. For a given element of the adjusted cognitive background, one should be able to find another element in the background that underpins its correctness.

For elements of the adjusted background that come from relevant beliefs, this new requirement boils down to the one presented above. Things are different for acceptances, though. Recall that the truth of their content or the fact that they make the agent better is necessary for their correctness. In terms of explanatory coherence of the adjusted background, it means that, for a given acceptance, one should be able to find other elements in the background that provide either evidence in favour of what is taken for granted or facts about the practical circumstances that motivate this very acceptance.

The cognitivist argument from explanatory to means-end coherence requires a specific assumption on beliefs about what one will do. Namely, these are to be explained by beliefs about what one intends. Now, the same will be required of elements of the adjusted background. I assume that if an agent accepts in a deliberation that he will do $A$ he should also accept that he intends to $A$, or at least that he will later form the intention to $A$.
From the three general norms on the adjusted background thus follow agglomerativity, internal consistency, a form of belief consistency and explanatory coherence of acceptances. It should be stressed that these norms are all context-dependent. Acceptances, unlike beliefs, are not required to be consistent nor to agglomerate across contexts. I now investigate whether, with these characteristics to hand, one can explain internal consistency, belief consistency and means-end coherence of intentions without assuming that they involve beliefs.

6.4 Hybrid Pragmatism

Cognitivism is so called because it aims at showing that the norms of practical rationality associated with intentions come, in fact, from the norms of theoretical rationality associated with beliefs. Similarly, I have called Bratman’s approach an “agency approach” because it tries to find an explanation for the norms of practical rationality in general structures of agency. Here I want to see whether one can derive the norms on intentions from norms on acceptances that mirror the norms of theoretical rationality for beliefs. But this derivation is based on the idea that intentions involve acceptances, and I find that the most fitting overall justification for this requirement is a pragmatic one. So even though the approach I investigate here has an important cognitivist component, it is also firmly pragmatist. For that reason I call it hybrid pragmatism.

In this section I first present the assumption that drives hybrid cognitivism, namely that intentions “involve” acceptance. I then go on to say a few words on the notion of relevance for a particular deliberation, and come back to the absent-minded cyclist case. The core of the section is the last part, in which I finally look at how the norms on intentions can be derived from the norms on acceptance.

6.4.1 Intentions and acceptances

The key idea underlying the explanation of the requirements on intentions is the following:

(1) Having the intention to do \(A\) implies, in deliberations where this intention is relevant, accepting that one will do \(A\).

This is indeed very close to the intermediate cognitivist assumption, with the crucial difference that beliefs are replaced by acceptances. One can find many statements that get close to (1) in the literature, notably in [Harman, 1976, p.438], [Bratman, 1999, p.32] and [Wallace, 2006, postscript to chap.5]. The changes from beliefs to acceptances, however, introduces a complication that the

\[28\]See the note on page 151.
notion of relevance for a deliberation tries to handle. The problem is that it does not seem realistic to assume that one can explicitly accept a very large number of facts when one deliberates\(^{29}\). The notion of relevance aims precisely at avoiding such an “overload” of acceptances. If, for example, I intend now to write my PhD thesis, there are many deliberations where I do not have to take that fact for granted. For most of my daily decisions this is just not relevant. But, as we shall see, relevance also plays a crucial role in the derivation of the requirements on intentions.

6.4.2 The absent-minded cyclist revisited

Before going further it is worth stressing that hybrid pragmatism is not bound to accept that the absent-minded cyclist described in Section 6.2.5 does not really intend to stop at the bookstore.

Recall the key features of this case. The agent intends to stop by the bookstore but he is uncertain whether he will in fact do so, because he knows that he tends to go on “automatic pilot” once on his bike. According to (1), he must come to accept that he will stop at the bookstore in further deliberations where this intention is relevant. But one can still deny that the agent would also assume that he will stop in theoretical reasonings. In short, one can still deny that the agent believes that he will stop at the bookstore.

6.4.3 Relevance for a deliberation

I have already mentioned twice the notion of relevance for a deliberation: in the Section 6.4.1, where it applied to intentions with respect to deliberations, and in Section 6.3.2, where it constrained the beliefs that have to be incorporated in the adjusted cognitive background.

These two uses of relevance regulate what should appear in the adjusted cognitive background. Indeed, in (1) it limits the intention-based acceptances that have to be incorporated in the adjusted background. Similarly, to say that beliefs, if not posited or bracketed, count as acceptances in deliberations where they are relevant is also to constrain what has to feature in the adjusted background. So, even though I have twice before mentioned the idea of relevance, relevance of beliefs once and relevance of intentions once, it served the same purpose in both cases.

I spell out the notion of relevance using the deliberately vague notion of “reasonably direct influence”. I state that a belief is relevant for a deliberation if the outcome of this deliberation depends in a reasonably direct way on the truth of the belief. In other words, if the content of the belief can make a reasonable difference in the deliberation, I state that this belief is relevant. The case is similar

\(^{29}\)This, again, is not a problem for ideal agents. More on this in conclusion of this chapter.
for intentions. I state that an intention is relevant for a deliberation whenever the fact that one would execute the intention under consideration can in a reasonably direct way influence the outcome of the deliberation. The obvious case is when one takes into account the fact that the agent will accomplish what he intends influences the payoff structure of a deliberation as in, for example, situations of “overloading” [Pollack, 1991]. But it may also be that some options in a given deliberation enhance or diminish the feasibility of a pre-existing intention. In that case the influence is less direct. The intention is relevant because the agent might want to take into account the consequences of his choice for the intentions that he has already settled on.

I do not think that it is necessary to get very precise about this notion of “reasonably direct influence”. I put it forward only because it seems too strong to suppose, for example, that all beliefs whose content can have the slightest influence on the outcome of a deliberation are relevant. That would, I think, make too many beliefs relevant, and the same for intentions. On the other hand, there are cases where intentions or beliefs are obviously relevant for a deliberation; for example, when choosing one option in a deliberation would make it impossible to achieve some intention. In such a case it is clear that the deliberation influences the prospect of achieving the intention. The correct notion of relevance for deliberations probably lies somewhere between these two extremes. But I do not think it is useful, nor easy, to draw a sharp distinction between what is relevant to a deliberation and what is not.

This is not to say that this notion of relevance is unimportant for what follows. Quite the contrary. On the one hand, it embodies a strong concern for the limited capacities of non-idealized planning agents by keeping down the number of facts that have to be taken into account during deliberation. But it nevertheless forces, so to speak, some acceptances to feature in deliberations. This provides hybrid pragmatism with a natural shield against the problem of false beliefs.

6.4.4 Rational intentions from rational acceptances

This section and the next are the keystones of this chapter. I look at how well hybrid pragmatism supports the derivation of internal consistency of intentions, strong belief consistency, means-end coherence and agglomerativity.

**Internal consistency.** This requirement is the easiest to derive. The argument is brief, and essentially parallels the one presented in Section 6.2.4. By (1), an internally inconsistent intention generates an internally inconsistent acceptance. Putting this contrapositive argument in the other direction, this means that internal consistency of acceptances implies internal consistency of intentions.

---

30See e.g. [Horty and Pollack, 2001].
31I should warn the reader that, in what follows, I often just use “influence”, living implicit the “reasonably direct” proviso.
Strong belief consistency. The argument for belief consistency is more involved because one has to deal with the notion of relevance for deliberation. The difficulty lies in the fact that intentions have to be realizable in a world that corresponds to the default cognitive background. For the intermediate cognitivist this is not a problem because its main assumption goes directly from intentions to beliefs, which “live” in the default background, so to speak. But hybrid pragmatism stays at the level of the adjusted cognitive background of relevant deliberations. To carry the derivation through in this framework, one must show that, in case of strong belief inconsistency, the contradictory belief-intention pairs are somehow jointly relevant in certain deliberations. But once this is shown, the argument is more or less the same as in Section 6.2.4.

Suppose an agent has strongly belief-inconsistent intentions. He has the intention to do $A$ but $A$ cannot be realized in a world where his (default) background beliefs turn out to be true. Here I assume that this is the same as his believing that $A$ cannot be done. Observe that this belief should be included in the adjusted cognitive background of a deliberation if this deliberation can in a reasonably direct way be influenced by the fact that $A$ cannot be done. Take any such deliberation. Since doing $A$ implies that $A$ can be done, which is just the negation of something we assumed is relevant for this deliberation, we get that the fact that the agent will do $A$ can also influence—through one of its consequences—the upshot of that deliberation. But this is just saying that the intention to do $A$ is also relevant for that deliberation. This means, by (1), that the agent should also include in the cognitive background of that deliberation the fact that he will do $A$. But this new acceptance is belief inconsistent\(^{32}\).

Agglomerativity. Let me once again begin with agglomerativity against potential irrationality. As for cognitivism, hybrid pragmatism derives it automatically given that it can explain internal and strong belief consistency. Again, there are two cases to consider. Since they are essentially similar, I only sketch the argument for agglomerativity against potential violation of belief consistency. Suppose that an agent has the intention to do $A$ and the intention to do $B$ which, were they conjoined in a single intention, would not be achievable, given his beliefs. We know from the previous section that this hypothetically conjoined intention

\(^{32}\)The reader should also bear in mind that internal consistency and agglomerativity of the adjusted cognitive background are at work here. They were used to derive belief consistency of acceptances. It should also be noted that the argument could also have proceeded as follows. Take any deliberation on means to achieve the intention to do $A$. It is clear that the fact that $A$ cannot be done is relevant to such deliberation, and so that it should feature in the adjusted cognitive background. Given (1), one thus directly obtains a violation of beliefs consistency of acceptances. This argument is not as general as it should be, if one grants that some intentions never require any further deliberations on means. I honestly doubt that there are such intentions. But in the absence of an explicit argument for that claim, I think that hybrid pragmatism should retain the more abstract derivation presented above.
would generate a violation of the belief consistency of acceptances, which is just what was needed.

Agglomerativity as closure, however, is more difficult to derive. Acceptances are indeed agglomerative within relevant backgrounds of deliberations. But here this restriction to the adjusted background is a blight rather than an asset. For suppose that an agent intends to do $A$ and intends to do $B$. Why should he intend to do $A$ and $B$? To use agglomerativity of acceptances, we would have to make sure that these two intentions are relevant to at least one common deliberation. But since we are considering arbitrary intentions, I do not see why this should be so.

We have seen in Section 6.2.4 that strong cognitivism can, however, easily justify this requirement. Maybe a stronger form of hybrid pragmatism, characterized as follows, could work.

(2) An agent intends to do $A$ if and only if he accepts, in relevant deliberations, that he will do $A$.

At first sight, such strong hybrid pragmatism is tempting, because acceptances are functionally very close to intentions. But just like its cognitivist analogue, (2) comes with independent problems. Namely, it makes it difficult to distinguish unintended side effects from intended consequences of actions, as argued in [Bratman, 2006b, p.18-20]. For that reason I think that one should resist the temptation to accept (2) and rather stick to (1). This, of course, means that hybrid pragmatism still falls short of an explanation of agglomerativity as closure of intentions. Given that this principle can itself be questioned, as we saw in Section 6.1, this might not be a very serious problem.

It should be noted, however, that if one adopts a principle of agglomerativity as closure restricted to plans, relevance in deliberation might come back into force and provide what we need to push agglomerativity as closure. For one would no longer be dealing with arbitrary intentions, but rather with intentions that are part of a single plan. This could help to find the common relevant deliberations that are absent in the general case.

I shall not go in that direction here, though. It would require me to specify the adequate level of agglomeration, and I have for now no precise idea of how one could do this. It is telling, however, that the only argument I sketched for agglomerativity as closure with a restricted scope was, in essence, pragmatic (see Section 6.1). I mentioned that, at the level of plans, to systematically agglomerate intentions might help personal coordination in extensive decision problems. If this argumentative strategy turns out to be fruitful, it would in itself lobby in favour of the “hybrid” character of the derivation using acceptances.

**Means-end coherence.** The main ingredient in the derivation of means-end coherence of intentions is explanatory coherence of acceptances. But, just like
cognitivism, using means-end coherence makes hybrid pragmatism vulnerable to the problem of false beliefs. I shall return to this at the end of the section. There is, in the meantime, another complication that the hybrid pragmatist derivation has to overcome.

This complication comes once again from the fact that (1) only requires the agent to take things for granted in the adjusted background of relevant deliberations, while the belief that “triggers” means-end incoherence is primarily a denizen of the default background. The main step of the derivation is thus to show that the means-end incoherent belief-intention pair can be relevant to at least some common deliberations. Just as for belief consistency, once this is shown the argument is fairly similar to the one in Section 6.2.4.

Suppose that an agent intends to do $E$ and believes that he will achieve $E$ only if he does ($M_1$ or $M_2$ or... or $M_n$). Take a deliberation the upshot of which can affect the prospect of achieving $E$. As noted in Section 6.3.3, this is a case where the fact that the agent will do $E$ can influence the upshot of that deliberation. The agent might want to take into account the effect of his decision on his background plans. This means, first, that this intention is relevant for that context, and so by (1) that he should include in the adjusted background the fact that he will do $E$. Observe, however, that the upshot of the deliberation can affect the prospect of doing $E$ only by affecting the prospect of doing $M_1$ or $M_2$ or ... or $M_n$, at least according to what the agent believes. That is, the agent might want to take into account the effect of his decision on the feasibility of each of these means. This makes the fact that doing $E$ implies doing one of these $M$ is also an important input into that deliberation, which is just to say that it is also relevant here. But then the adjusted background features both the facts that the agent will do $E$ and that doing $E$ implies doing one of the $M$s. By the closure principle for the adjusted background, the agent must come to accept that he will do $M_1$ or that he will do $M_2$ or... or that he will do $M_n$. Finally, by applying explanatory coherence one can conclude that the agent should also accept in that background that he intends one of these means or, at least, that he will come to intend one of them later.

As Wallace puts it in the context of the cognitivist derivation, this is as far as explanatory coherence of acceptances can get the derivation. But this is not quite as far as one needs, unless one can show that the agent should not take for granted that he has the required means-intentions without, in fact, having these intentions. In other words, mistaken acceptances about one’s own intentions seem to threaten the hybrid pragmatist derivation just as it did the cognitivist one.

As we saw in Section 6.2.4, Wallace has proposed a solution that seemed to cope with the problem. This solution, however, involves a step outside the pure cognitivist perspective because it crucially invokes a principle of practical rationality. But hybrid pragmatism is built on the notion of acceptances in deliberations, which are sensitive to practical concerns. For that reason, I think it can naturally claim Wallace’s reply to the problem of false belief, in a way that
might be even more suited to his “hybrid” nature.

Recall that the essence of Wallace’s reply was that, in relevant deliberations, it is independently irrational, for general pragmatic reasons, to hold false beliefs about one’s own intentions. Mutatis mutandis, one can say that it is independently irrational, for general pragmatic reasons, to mistakenly take for granted one’s own intentions in deliberations where the latter would be relevant.

I have already explained in Section 6.2.4 why Wallace’s emphasis on relevant contexts blocks counterexamples like the one proposed by Bratman. The explanation is readily applicable to acceptances, and is arguably even more convincing in these terms.\(^{33}\)

But to push through the derivation of means-end coherence one needs not only to show that this particular counterexample could be handled; one needs to show that, in general, mistaken acceptances about one’s own intentions are irrational. This was the object of Wallace’s recourse to a general principle of practical rationality. Hybrid pragmatism is, once again, especially suited to the incorporation of such a principle. Acceptances are after all responsive to practical concerns.

The reader should observe, however, that Wallace’s general type of practical concerns is not the same as the one that features in the characterization of acceptances (Section 6.3.1). The latter is specific to deliberation contexts, while the former are “to be included among the traits and capacities that make us, in general, effective in the pursuit of our goals”. In fact, there is a principle of this sort that can be used to provide a “wide scope” justification of (1), which I shall look at in the next section.\(^{34}\)

Before I do that, though, let me summarize where things stand now. Hybrid pragmatism can easily explain internal consistency of intentions and, after taking care of complications regarding relevance in deliberation, strong belief consistency as well as agglomerativity against potential irrationality. Just like intermediate cognitivism, however, it fails to explain agglomerativity as closure. It can also explain means-end coherence, provided that we can secure the irrationality of

\(^{33}\)This reply also takes care of another potential counterexample. Acceptances are, as we saw, both responsive to evidence and practical concerns. Now, suppose that an agent has, in the default background, a mistaken belief about his own intention, which is supported by some misleading evidence. Would it not then be rational of him to incorporate this fact in relevant deliberation? The answer, I think, should be “no”, because such deliberations are again precisely deliberations where the fact that the agent has the intention is relevant.

\(^{34}\)To the extent that the last step of the argument succeeds, it shows how suited to hybrid pragmatism is Wallace’s attempt to save the cognitivist derivation of means-end coherence. Except for the switch from acceptances to beliefs, he spelled out all the key steps of the argument. He saw the importance of looking at deliberation where one’s own intentions are relevant, and connected it to pragmatic consideration. In fact, his suggestion is much more at odds with the orthodox cognitivist line than with hybrid pragmatism. At the end of the day, one may wonder why he did not directly speak in terms of acceptances.
mistaken acceptances about one’s own intentions. This requires a more general pragmatic argument, which I now explore.

### 6.4.5 Acceptances and intentions in deliberation

The hybrid pragmatist derivation of the norms on intentions rests on two assumptions that still have to be justified: the implication (1) from intention to acceptances and the connection between intention and acceptances about intentions. In this section I first sketch how one can provide a two-step argument for (1). At the core is a constitutive claim in favour of (1), which is cloaked by a “wide scope” pragmatic justification. The ideas driving this second step are quite unifying, allowing one also to secure the derivation of means-end coherence. I finally look back at the constitutive claim, in the light of a more general worry about the hybrid pragmatist enterprise.

A key ingredient in the theory of practical reasoning for planning agents that I have developed in this thesis is that intentions should not only be seen as outputs of deliberation, but also as important inputs. This influence of previously adopted intentions and plans on deliberation is twofold. They trigger deliberation about the means by which they will be achieved and they rule out of consideration options with which they are incompatible.

The planning theory, however, says little about how they do so, and it seems that acceptances help to understand this connection. More precisely, the implication (1) from intentions to acceptances, together with the various norms on the latter, provide the required connection between intentions and deliberation.

Let me first look at the filtration of options that comes from intentions. How is that supposed to happen? Bratman [1991, p.32] proposed that “what is required of an option in deliberation is that one’s so acting be consistent with one’s intentions and beliefs, and that one accept in that context that one will so act if one so decides”. One can say, more generally, that options in deliberations have to be consistent with the adjusted cognitive background. But then, if we assume (1), the options to be considered in a deliberations will have to be consistent with the agent doing what he intends. Conversely, (1) gives us a way to see why intentions rule out inadmissible options. They do so via acceptances.

The situation is analogous with respect to the standard of relevance with respect to means that intentions impose. Recall the reasoning that I pictured in Section 6.4.4. I started with an intention to do \( E \) and the belief that to do \( E \) one must do \( M_1 \) or \( M_2 \) or... or \( M_n \). From these I concluded, after a few intermediate steps which crucially involved (1), that one should accept that he will do \( M_1 \) or that he will do \( M_2 \) or... or that he will do \( M_n \). But, in the absence of a mistaken acceptance that the agent already has the corresponding intention, recognizing that this intention is absent may well trigger means-reasoning. In other words, becoming aware of this threat of means-end incoherence may trigger deliberation about means. Note that this is not to say that the agent should form
the intentions required to resolve means-end incoherence. For now the claim is only constitutive. With (1), intentions about ends can find their way into the adjusted background and set off deliberation about means.

So (1) can be seen as a way to understand how intentions influence practical reasoning. As I just mentioned, this is a constitutive claim. The next step is to ask why the intentions of planning agents should play this role. In other words, is there a normative reason to think that planning agents should take their previously adopted intentions into account while deliberating? Here I think that the most promising answer to this question is the general pragmatist approach proposed by Bratman [2006b] and Wallace [2006].

Intentions generally “aim at coordinated control of action that achieve what is intended” [Bratman, 2006b, p.33, emphasis in original] or, in other words, intentions are “among the traits and capacities that makes us, in general, effective in the pursuit of our goals” [Wallace, 2006, p.118]. Forming intentions is, in short, a generally useful tool to achieve our goals. Many features of intentions make them useful. They are “conduct controlling”, relatively resistant to reconsideration and, what is important for hybrid pragmatism, they influence further deliberations.

This normative claim indeed very well suits the hybrid nature of (1). If the reason why intentions should play a role in practical reasoning is a general, pragmatic one, it is natural to think that they will do so in a way that is also responsive to practical concerns, i.e. via acceptances. But one should bear in mind that it is a “wide scope”, normative claim. In the words of Broome [1999], the “should” is not detachable. If an agent intends to do A it does not follow that he automatically has good, practical reasons to take it for granted that he will do A, especially if he did not have good reasons to form that intention to start with. Rather, the argument aims at showing that there are general pragmatic reasons for planning agents to comply with (1).

This general pragmatic claim reaches further than the justification of (1). It also allows us to overcome the problem of mistaken acceptances. In exactly the same fashion as Wallace, one can say that to mistakenly accept that one intends something in relevant deliberation is irrational because it threatens the general efficiency of planning agents to reach their goal. Again, bluntly, mistaken acceptances can mess up the agent in the long run. With this in hand, hybrid pragmatism has a way to explain the means-end coherence requirement for intentions, in a way that fits the overall justification of the others requirements. As such, it seems to provide a more unified argumentation than Wallace’s sup-

---

35Observe that one is not bound to take this route. Since (1) is taken here as a constitutive claim, one could as well try to cloak it with a more cognitively-oriented normative argument. I choose the pragmatist one because it seems to me the more plausible.

36In quasi-formal terms, the argument aims at showing that “pragmatic-should (Intention A → Acceptance that A)” and not that “Intention A → pragmatic-should (Acceptance that A)”.

37In fact, one could argue that they can mess up even more than mistaken beliefs, because of the crucial context-dependency of acceptances.
implemented cognitivism\textsuperscript{38}. In this case one could wonder why one suddenly needs to appeal to practical concerns to derive means-end coherence, while this is not the case for belief and internal consistency. But if one uses acceptances, practical concerns are there all along.

This raises a more general question about the motivations behind the hybrid pragmatist view. One attractive feature of cognitivism, to the extent that it succeeds, is its clear stance: the practical norms on intentions can be explained only in terms of theoretical norms on beliefs. The same can be said about Bratman’s agency approach: the practical norms on intentions can be explained only in practical terms\textsuperscript{39}. In comparison, hybrid pragmatism sits in a rather grey area, invoking both pragmatic and theoretical-like requirements.

I think one can better appreciate the relevance of such a mixed approach by looking at the constitutive claim mentioned at the beginning of this section. Intentions are playing a role in practical reasoning via their influence on the cognitive background of deliberation, which is under some pressure to represent accurately the world. From that point of view, it seems that the norms on intentions fall in two categories. On the one hand, the derivations of agglomerativity against potential irrationality, of internal consistency and of strong belief consistency rest on the requirements of closure under logical operation and of internal consistency for the adjusted background of deliberation. These norms obviously mirror norms of theoretical rationality. This should not come as a surprise. To paraphrase again Velleman [2003], the adjusted cognitive background seems to aim at picturing the world, of which there is just one, where the various courses of action considered shall be pursued. It plays in practical reasoning a role similar to the role of the default background in theoretical reasoning. Inasmuch as the argument in Section 6.4.4 is correct, the norms of agglomerativity against potential irrationality, of internal consistency and of strong belief consistency ultimately derive from this role of the adjusted background. This gives a strong “cognitive” character to these norms, by which they somehow differ from means-end coherence. The derivation of this norm on intention indeed crucially rests on practical concerns, especially to avoid the problem of false beliefs or of mistaken acceptances. As such, means-end coherence appears to be, unlike the three other norms just mentioned, a genuinely practical norm on intentions.

This general distinction between the practically- and the cognitively-oriented norms on intentions help to understand why “pure” cognitivism as so much difficulty with means-end coherence, and why it seems forced in the end to fall back on pragmatic concerns. To explain this practically-oriented norm one needs at least some pragmatist component. But, on the other hand, a pure agency approach seems to leave behind the fact that some norms on intentions have a

\textsuperscript{38}The epithet comes from Bratman [2006b, p.8].

\textsuperscript{39}One should take care about where the “only” is located. Bratman does not seems to claim that the practical norms can only be explained in practical terms.
important cognitive character. The theoretical-like constraints that they embody are perhaps better explained in their relations with the cognitive background of deliberation, which is precisely what hybrid cognitivism does by using acceptances.

6.5 Conclusion

Hybrid pragmatism is thus a plausible philosophical approach to the norms of consistency and coherence which apply to intentions. It not only helps us to understand where these norms come from, but also allows us to explain how intentions influence deliberation, via the notion of acceptances. That this approach definitely stands in between cognitivism and Bratman’s agency approach can also be seen as an asset. It provides an account that does justice to the influence of intentions on both the pragmatic and the cognitive component of practical reasoning. In other words, it provides a unifying background to the role of intentions in rational decision making, something that was arguably missing in the last chapters.

To conclude this chapter I would like to address one final question, bearing to resource-boundedness. As I explained in Chapter 2, I have carried the formal investigations under strong assumptions about the capacities of agents: all issues pertaining to resource-boundedness were ignored. Resource-boundedness played, however, an important role in the discussion of hybrid pragmatism. A crucial asset of acceptances is that they simplify deliberation, by for instance provisionally settling issues that the agents is uncertain about. Resource-boundedness also came back through the idea of relevance for deliberation, in Section 6.4.3, and during the discussion of the problem of false beliefs (Sections 6.2.4 and 6.4.4). Does it mean that hybrid pragmatism is a suitable theory of how intentions influence practical reasoning for limited agents, but not for the type of agents that I studied in Chapters 2 to 5?

I think not, for the following reasons. First, in the discussion around the idea of relevance for deliberation and around the problem of false beliefs, resource-boundedness rather raised difficulties for the hybrid pragmatist justification. It has, for instance, forced to look carefully at what agents should accept in deliberation, precisely for the reason that resource-bounded agents can only take for granted a limited amount of facts. Since this problem does not arise for non-resource-bounded agents, the hybrid pragmatist derivation of the norms on intention seems even more straightforward under the idealizations I made in Chapters 2 to 5. Second, the idea that acceptances are of greater importance for agents with limited capacities does not undermine their explanatory power in the general case. They still provide a simple way to understand how intentions influence practical

---

40I am grateful to Michael Bratman (p.c.) for pointing me out this potential weakness in the connection between the present chapter and the foregoings.
6.5. Conclusion

reasoning—the “constitutive claim” of Section 6.4.5—an issue that has been very little addressed in the philosophical literature. In other words, acceptances in deliberation and hybrid pragmatism complete the picture of how future-directed intentions constraint practical reasoning, whether for ideal or resource-bounded agents. As such, the philosophical theory developed in this chapter seems indeed to encompass both the case of agents with limited capacities and the ideal agents which have been studied in Chapters 2 to 5.

Of course, hybrid pragmatism also opens up questions about intention-based practical reasoning from a formal perspective. How is one to include acceptances in the formal theory of deliberations that I proposed? How will they be distinguished from other kinds of “informational” states, like knowledge and belief? How, in these models, can one account for the fact that acceptances are also responsive to pragmatic considerations? These are crucial questions that a formal enquiry would surely help answer. Furthermore, one can hope that bringing acceptances into the formal picture will also unveil new issues about them, just as in the previous chapters game and decision-theoretic models unveiled new issues about intentions. All in all, hybrid pragmatism is very well suited as the final theme for this thesis. It rounds up issues that were carried all the way, and opens up exiting new ones.