Instrumental reasoning*

John Broome

1. Intention reasoning

Here is an example of practical reasoning:

       I am going to St Andrews   (A1)

and   To go to St Andrews, a necessary means is to get off at Leuchars,   (A2)

so   I shall get off at Leuchars.   (A3)

I mean (A1) to express an intention of yours to go to St Andrews rather than a belief that you are going to St Andrews. I mean (A2) to express a belief of yours, and I mean (A3) to express a decision you make.

Think of this as a process of reasoning you might actually run through. You might do so when you come to buy your ticket. You approach the ticket office intending to go to St Andrews; you form the belief that a necessary means is to get off at Leuchars when the ticket clerk tells you so; and then you go through the reasoning (A). It takes you from two of your existing states of mind, an intention and a belief, to a new state of mind, an intention to get off at Leuchars. To form an intention, at least as a result of reasoning like this, is to make a decision. So this reasoning concludes in a decision.

Making the states explicit by writing 'I' for 'you intend' and 'B' for 'you believe', we can describe your reasoning like this:

       I(I am going to St Andrews)   (B1)

and   B(To go to St Andrews, a necessary means is to get off at Leuchars)   (B2)

leads to   I(I shall get off at Leuchars).   (B3)

This is a description of your reasoning from the outside. It is not an inference. From the fact that you intend to go to St Andrews, and the fact that you believe a necessary means of doing so is to get off at Leuchars, we cannot infer that you intend to get off at Leuchars. You may not have that intention, for instance if you are irrational.

On the other hand, your reasoning seen from the inside is an operation on the contents of the states, and it *is* an inference. It is represented in (A).

I shall call reasoning that concludes in the forming of an intention 'intention reasoning'. Since it leads to a decision, it is practical reasoning. Reasoning could not be more practical than this; it could not get closer to action. It certainly could not conclude in a physical act itself, because it takes more than reasoning ability to bring about a physical act. It takes physical ability too. So intention reasoning is the paradigm of practical reasoning.

The reasoning process (B) can be categorized in another way. It is instrumental reasoning, which means it is concerned with taking an appropriate means to an end. If we want to understand practical reasoning, it is a good idea to start with instrumental reasoning, because it is generally held to be more straightforward than other sorts. Kant (1948, p. 80) says it 'requires no special discussion', though I hope this paper proves him wrong. It considers instrumental reasoning only.

Furthermore, it considers only a special case of instrumental reasoning: the case of reasoning to a necessary means. In another paper (Broome, 1998), I have extended the argument to means that are not necessary.

I have been begging the question of whether intention reasoning is truly reasoning at all. If it is to be the paradigm of practical reasoning, it must be reasoning. Indeed it must be correct reasoning. (I shall keep the term 'valid' for the content of reasoning, and use 'correct' for the reasoning process.) We could call almost any process of thought 'reasoning', but I shall generally use this term only for correct reasoning. Intuitively, (B) describes correct reasoning, but we need an explanation of why.

2. Kant and the correctness of intention reasoning

Kant (1948, p. 80) asks 'how we can conceive the necessitation of the will expressed by the imperative in setting us a task'. In the context of instrumental reasoning, I think he means, for instance: how does my intention of going to St Andrews make it necessary for me to intend to get off at Leuchars? (I read 'will' as 'intend'.) I have just said this has the necessity of correct reasoning. Intuitively at least, intention reasoning of a form like (B) is correct, and that is why I must intend a necessary means if I intend an end. So an explanation of how intention reasoning is correct will answer Kant's question, and an answer to Kant's question will explain how intention reasoning is correct.

In answering his question, Kant (1948, pp. 80–1) says, first:

Who wills the end, wills (so far as reason has a decisive influence on his actions) also the means which are indispensably necessary and in his power.

This remark needs a small correction: Kant should have said '. . . wills also what he *believes* to be the means which are indispensably necessary . . .' Once corrected, the remark is no doubt true. If you are rational, you cannot intend to go to St Andrews, and believe that getting off at Leuchars is a necessary means of getting there, without intending to get off at Leuchars. But this merely states what has to be explained. Why is it so? If (B) describes correct reasoning, then a rational person who intends to go to St Andrews, and believes that getting off at Leuchars is a necessary means of doing so, will be brought by this reasoning to intend to get off at Leuchars. Then Kant's remark will be true of her. But it does not help us understand why the reasoning is correct.

Kant (1948, p. 81) next says his remark is analytic, but again that does not take us far forward. If intention reasoning is correct, then it is part of the meaning of 'rational' that a rational person will be guided by it. 'Rational' means, in part, guided by correct reasoning. So Kant's remark will be true in virtue of the meaning of 'rational'. But we still need an explanation of how intention reasoning can be correct in the first place.

Kant (1948, p, 81) finally says, in explanation:

For in my willing of an object as an effect, there is already conceived the causality of myself as an acting cause – that is, the use of means; and from the concept of willing an end the imperative merely extracts the concept of actions necessary to this end.

If you conceive of yourself as doing something, you must conceive of yourself as doing it *somehow* – that may be true. But when you conceive of yourself as doing something, you need not have any specific means in mind. So when you conceive of yourself as doing something, it cannot be part of your conception that you conceive of yourself as taking some specific means to do it. You could conceive of yourself as going to St Andrews without conceiving of yourself as getting off the train at Leuchars.

Kant acknowledges this, but he thinks it is not possible if you are rational and if you believe getting off at Leuchars is a necessary means of going to St Andrews. No doubt this is true. But it cannot be true just in virtue of the meaning of 'willing the end', as Kant suggests. It must also involve the meaning of 'rational'. It is true because, as a rational person, you are guided by correct intention reasoning. So we still need an explanation of how intention reasoning can be correct. But at this point Kant's explanation gives out.

I shall offer an explanation in the next section.

3. The correctness of intention reasoning and belief reasoning
To see why intention reasoning is correct, it will be helpful to compare it with a sort of reasoning that is better understood in philosophy. I mean theoretical reasoning or, as I shall call it 'belief reasoning'. All reasoning, conceived as a process, starts from existing states of mind and concludes in a new state of mind. By 'belief reasoning', I mean reasoning that

concludes in a belief. An example is:

	B(I am going to St Andrews)	(C1)
and	B(To go to St Andrews, a necessary means is to get off at Leuchars)	(C2)
leads to	B(I shall get off at Leuchars).	(C3)

Like (B), (C) describes a process of reasoning from the outside. It is not an inference. From (C1) and (C2), we cannot infer (C3). If you are irrational you might not.

Seen from the inside, the reasoning operates on the contents of your beliefs, like this:

	I am going to St Andrews	
and	To go to St Andrews, a necessary means is to get off at Leuchars,	
so	I shall get off at Leuchars.	

These three statements represent the contents of beliefs, and together they represent the content of the reasoning described in (C). I shall take the individual contents to be propositions. They constitute a valid inference. These three propositions stand in a particular relation to each other, a relation such that, if the first two are true, so is the third. This makes the inference valid, which in turn makes the process of reasoning correct. The reasoning is correct because its content is valid.

The validity of this content plays a part in other sorts of reasoning besides (C), because propositions do not need to be believed for them to play a part in reasoning. For example, the same content might feature in hypothetical reasoning, where you do not believe (C1) or (C2), but are working out what would be true if they were true. The same validity also plays a part in intention reasoning. The content of belief reasoning (C) is in fact exactly the same as (A), the content of intention reasoning (B). This content (A) is a valid inference. It turns out that this fact makes (B) correct just as it makes (C) correct.

The difference between (B) and (C) is not in their content but in the stance you take towards their content. In (C), your stance towards (A1), that you are going to St Andrews, is to take it as true. In (B) your stance is to set yourself to make it true. Because (A) is valid, if (A1) and (A2) are true, (A3) must be true. This is what makes belief reasoning (C) correct. Also because (A) is valid, if (A2) is true, then if you are to make (A1) true you must make (A3) true. This is what makes intention reasoning (B) correct. Both (B) and (C) appropriately track the transmission of truth through the valid inference (A). (B) tracks it in a truth-making way and (C) in a truth-taking way.

If David Hume (1978, Book 2, Part 3, Section 3) was right that reason is concerned only with truth, he should still have recognized that reasoning can transmit the truth-making stance as well as the truth-taking stance. It can transmit intention as well as belief. So reasoning can be practical.

Intention reasoning is parallel to belief reasoning in the way I have described, and both depend on the validity of the same content. Consequently, one might be tempted to think intention reasoning somehow has belief reasoning embedded in it. If you intend to go to St Andrews, you normally believe you are going there. So you are normally in a position to run through the belief reasoning (C), which will bring you to believe you will get off at Leuchars. One might be tempted to think this is part of the process of forming your intention to get off at Leuchars. But my explanation of intention reasoning shows that belief reasoning is actually not involved in this process; it merely has the same content. In any case, a belief that you will get off at Leuchars cannot by any rational process bring you to form the intention of doing so. Belief reasoning is parallel to intention reasoning, but not a part of it. The validity of the content plays its part in intention reasoning directly, and not by giving you a belief.

4. Restrictions on the notion of intention

If you believe the premises of a valid inference, then correct reasoning will bring you to take the conclusion as true; you will believe it. Similarly, if you intend some of the premises of a valid inference, and believe the others, then correct reasoning will bring you to set yourself to make the conclusion true. However, it does not always follow that you will intend the conclusion. Setting yourself to make something true is not necessarily intending it. Our notion of intention does not include all instances of setting to make true.

Take this putative case of intention reasoning, for example:

        I(I am going to St Andrews)

and   B(If I am going to St Andrews, I shall be kept awake by seagulls)

leads to   I(I shall be kept awake by seagulls).

The content of this reasoning is valid, but the reasoning itself is not an intuitively correct piece of intention reasoning. Intuitively, it seems you need not intend to be kept awake by seagulls, even though you recognize this will be a consequence of what you do intend. True enough, because you intend to go to St Andrews, you must adopt the stance of truth-making towards the proposition that you will be kept awake by seagulls. To fulfil your intention, you must make this proposition true. But we would not say you intend to be kept awake by seagulls.

This is partly because of a conventional restriction we normally impose on our use of 'intend': we normally only say you intend something if it is an act of yours. Since being kept awake is not an act, we would not normally say you intend it. If this feature of normal usage was all there was to it, the restriction would be unimportant from the point of view of practical reason, because from that point of view we are only interested in acts. Furthermore, our normal usage might be dispensable and have no philosophical significance. Indeed, we do sometimes dispense with it: we would indeed say you intend to be kept awake by seagulls if that was your purpose in going to St Andrews.

But there may also be a restriction that applies even to acts. According to the 'doctrine of double effect', you may intend to do some act, and believe that some other act of yours will be a consequence of doing it, yet not intend to do the other act. For instance, according to the doctrine of double effect, the following is not correct intention reasoning, even though its content is valid:

        I(I shall save the woman)

and   B(If I save the woman, I shall kill her unborn child)

leads to   I(I shall kill her unborn child).

According to the doctrine, you need not intend to kill the child even though you recognize you will do so as a consequence of what you intend to do. You must intend to do it only if you believe it is a means to what you intend to do. If the doctrine of double effect is right, it is a further restriction on the correctness of intention reasoning: intention reasoning is correct only if it is premised on a belief of the particular form that one act is a necessary means to another. We could not dismiss this restriction as merely a dispensable feature of our normal usage, because if the doctrine is right, it is morally significant.

I do not wish to argue about the doctrine of double effect. That is why I made my example satisfy its special restriction. (A2) contains the strong modality 'necessary means', rather than a simple material implication. Consequently, there should be no doubts about the correctness of (B). It might also be correct with a weaker conditional statement in place of (A2), but I do not wish to argue about that.

In summary, not all putative intention reasoning that has a valid content is correct. To make the reasoning correct, some further constraints must be satisfied. But when these constraints are satisfied, intention reasoning is genuinely correct reasoning, and it provides a

clear paradigm of practical reasoning.

5. Reasoning is not reason-giving
Because intention reasoning is reasoning, we may say it is normatively guided; it is an application of reason. Reason guides you from intending an end to intending a necessary means. But intention reasoning is normative in no other way. For one thing, its premises are not normative and nor is its conclusion; neither is about what you ought to do or have a reason to do. (By 'a reason' I mean a *pro tanto* ought. If you have a reason to do something, that means you ought to do it unless you also have a contrary reason not to.)

Furthermore intention reasoning is not ought-giving nor even reason-giving; that is what I shall argue in this section and the next two. What do I mean? In my example, intention reasoning takes your from your intention to go to St Andrews to an intention to get off at Leuchars. But it does not determine that you ought to get off at Leuchars, or even that you have a reason to get off at Leuchars. I could equally well say that the intention on which the reasoning is premised is not reason-giving. Your intention to go to St Andrews gives you no reason to get off at Leuchars. That is what I shall argue.

Reasoning in general is not ought-giving or reason-giving. (I shall mention an exception later.) Once again, it is easiest to see this by looking at the more familiar example of belief reasoning. Suppose you believe some proposition *P* from which *Q* follows by a valid inference. It does not follow that you ought to believe *Q*, nor that you have a reason to believe *Q*. (This is so even if the inference is immediate; it has nothing to do with the complexity of the inference.) This section and the next defend this claim. Section 7 returns to intention reasoning.

From one point of view, a defence scarcely seems needed because the conclusion is obvious. For instance, suppose you ought not to believe *P*, though you do. Then it is surely obvious that it may not be the case that you ought to believe *Q* or that you have a reason to believe *Q*.

Still, here is a simple defence in case it helps. *P* itself follows from *P* by a valid inference. But obviously, from the fact that you believe *P* it cannot follow that you ought to believe *P*, or have a reason to believe *P*. Beliefs do not automatically justify themselves. So it cannot be a general principle that if you believe *P* you ought to believe its consequences, or that you have a reason to do so.

However, I need to respond to a plausible contrary thought: if you believe *P*, then surely in some sense or other that gives you a reason to believe its consequence *Q*, when *Q* is different from *P* itself. No doubt, we would not say believing *P* gives you a reason to believe *P* itself, because you do not need a reason for that; you already believe it. But if you are going to believe a consequence of *P* that is distinct from *P*, you do need a reason for that, and surely believing *P* gives you one, in some sense or other.

Perhaps this thought does not apply to every consequence of *P* that is distinct from *P*. For instance, suppose you believe the earth is flat. Perhaps this gives you no reason to believe that either the earth is flat or the moon is made of green cheese, even though this disjunction is a consequence of your belief and distinct from it. But perhaps this is because, once again, you do not need a reason to believe the disjunction, since perhaps in some way you already believe it by believing one of the disjuncts. However, if you believe the earth is flat, then surely in some sense that gives you a reason to believe the *very* distinct consequence that there is no horizon.

In response to this contrary thought, I am willing to agree that *in some sense or other* your belief that the earth is flat gives you a reason to believe there is no horizon. In general, if you

believe *P*, in some sense or other this gives you a reason to believe the consequence *Q*, if *Q* is significantly different from *P* itself. But this sense is an unsatisfactory one. Expressed more precisely, the position is that a particular relation holds between your believing *P* and your believing *Q*: one *rationally requires* the other, as I shall put it. In symbols:

$$\text{B}(P) \text{ O}{\rightarrow} \text{B}(Q), \tag{D}$$

where O→ is this relation of rationally requiring. The formula (D) attaches rationality to the relation between the beliefs.

In belief reasoning, the relation of rationally requiring mirrors the relation of entailment that holds between propositions. If one proposition entails another, believing the first rationally requires believing the second.

6. Rational requirement

If we needed to formalize a rational requirement, we could adopt the notion of a 'conditional obligation' from deontic logic. (See the survey in Åqvist (1984).) The analysis of conditional obligations is by no means settled. (For a recent analysis, see Makinson (1998).) It might turn out, for instance, that (D) is best understood as:

$$\text{O}(\text{B}(P) \rightarrow \text{B}(Q))$$

where 'O' stands for 'ought' and '→' is the material conditional. But the important question for our purposes is whether (D) implies

$$\text{B}(P) \rightarrow \text{R}(\text{B}(Q)), \tag{E}$$

where '→' is the material conditional, and 'R' stands for 'have a reason to'.

In (E), the rationality is attached to the consequent rather than the connective, as it is in (D). This means the consequent can be detached by modus ponens. From B(*P*) and (E), we can infer R(B(*P*)). If you believe *P*, then (E) says you have a reason to believe *Q*. As it stands, (D) does not allow detachment of that sort. It says that believing *P* gives you a reason to believe *Q* in a sense, but not in a sense that allows us to say, literally, that if you believe *P* you have a reason to believe *Q*. The reason is always relative to your belief in *P*. Unless it can be detached, it does not determine what you actually have a reason to believe. So unless the reason can be detached, the sense in which believing *P* gives you a reason to believe *Q* is decidedly misleading. Detachment requires (E), so we must know whether (E) follows from (D).

The following argument shows it does not. Suppose you believe both *P* and ¬*Q*. These are inconsistent beliefs, since *Q* follows from *P*, but of course you might have inconsistent beliefs. Parallel to (D), we have that

$$\text{B}(\neg Q) \text{ O}{\rightarrow} \text{B}(\neg P)$$

Your belief in *P* requires you to believe *Q*, and your belief in ¬*Q* requires you to believe ¬*P*. These relative requirements make good sense; it is part of your inconsistent condition that some of your beliefs require you to have beliefs that contradict others of your beliefs. But suppose it was possible to infer (E) from (D). Then we would be able to infer that you have a reason to believe *Q* and similarly that you have a reason to believe ¬*P*, both nonrelatively. If you have a reason to believe ¬*P*, evidently you have a reason not to believe *P*. So you are supposed to have a reason to believe *Q*, but this reason is supposed to be derived from your belief in *P*, which you are supposed to have a reason not to have. This is incredible.

If you do not find it immediately incredible, imagine that, apart from their connection with each other, you have no reason to believe *P* or ¬*P*, or *Q* or ¬*Q*. Then if your belief in *P* gives you a nonrelative reason to believe *Q*, you ought to believe *Q*, because there is no contrary reason. Likewise, if your belief in ¬*Q* gives you a reason not to believe *P*, you ought not to believe *P*. So the claim is that you ought to believe *Q*, solely because of a belief you ought

not to have. This really is incredible.

A possible reply is to say that a belief only generates a nonrelative reason if it is consistent with your other beliefs. That is to say, there is no direct inference from (D) to (E), but there is an inference if we add the premise that your belief in *P* is consistent with your other beliefs. I can see no reason why we should accept this qualification: if a consistent belief generates a nonrelative reason, why does not an inconsistent belief? But in any case, a different version of the same argument works against this reply too.

Suppose you believe *P* and you do not believe *Q*, and suppose your belief in *P* is consistent with all your other beliefs. Plainly, not believing *Q* rationally requires you not to believe *P*:

$\neg$B(Q) O$\rightarrow$ $\neg$B(P).

(I am not suggesting that the relation of rational requirement always permits contraposition, but it does in this case.) Then exactly the same argument will go through, but this time your belief in *P* is consistent with your other beliefs.

There is one special sort of belief reasoning that is genuinely reason-giving, indeed ought-giving, in a nonrelative way. Some reasoning concludes in a tautology, which rests on no premises; the premises are discharged along the way. You ought nonrelatively to believe a tautology. This is an exception to my general claim that reasoning is not reason-giving or ought-giving. However, there is no parallel exception for intention reasoning, since intention reasoning has to have an intention as a premise.

The strict notion of 'ought' can be weakened in two different ways. A *reason* is a weakened sort of ought; it is a *pro tanto* ought. A *rational requirement* of the sort expressed in (D) is also a weakened sort of ought; it is weakened by being made relative. It is easy to muddle the two types of weakening, and slide between a relative ought and a nonrelative *pro tanto* ought – between a rational requirement and a reason. But they are different.

What difference does it make? Suppose you believe *P* and have no reason not to believe *Q*. Then, if your belief in *P* gave you a reason to believe *Q*, in the sense of a *pro tanto* ought, you actually ought to believe *Q*, because this reason would not be opposed by a contrary reason. But in fact it is not necessarily the case that you ought to believe *Q*. That may well depend on whether you ought to believe *P* in the first case. Your belief in *P* merely requires you, in the sense of a relative ought, to believe *Q*.

To state this rational requirement is merely to repeat the assumption I made at the start, that *Q* follows from *P*. It restates this assumption at the level of mental states – of beliefs. Because *Q* follows from *P*, correct reasoning will take you from a belief in *P* to a belief in *Q*. That is to say, a belief in *P* rationally requires a belief in *Q*. At the beginning of Section 5, I said that reasoning is normatively guided. In recognizing the notion of a rational requirement, I have done nothing but make this more explicit. Some states of mind rationally require others. Reasoning is not ought-giving or reason-giving in anything but a misleading sense. Nor are the states of mind it is premised on.

7. Intentions are not reason-giving

Everything I have said about belief reasoning applies to intention reasoning too. If you intend an end, and believe some act is a necessary means to it, your intention and belief rationally require you to intend the means. Reasoning will bring you to intend it. This is simply to restate the correctness of intention reasoning such at (B). It does not follow that you ought to intend the means, nor that you have a reason to intend it.

For example, if you intend to go to St Andrews, and believe that getting off at Leuchars is a necessary means of getting there, it does not follow that you ought to get off at Leuchars or have a reason to do so. From one point of view, this is obvious. It might be, for instance, that

you ought not to be going to St Andrews in the first place.

In *Intentions, Plans and Practical Reason*, Michael Bratman (1987, pp. 23–7) points out that a theory of practical reason must not imply 'bootstrapping', as he calls it. Just because you intend to go to St Andrews, that intention cannot possibly create a reason for you to go there if you have none already. An intention cannot justify itself, by providing itself with a reason. Still, your intention must play a role in your reasoning and lead you to other intentions, including the intention of getting off at St Andrews. The purpose of Bratman's book is to elucidate the vital role of intentions in the reasoned planning of our lives. It creates a puzzle. If your intention leads to other intentions by reasoning, does it not create a reason for those other intentions? That would still be a sort of bootstrapping; a reason is being pulled into existence out of nothing. Now we have a solution to this puzzle. There is no bootstrapping. One intention gives rise to another by means of reasoning, but no reasons are involved. There is only the relation of rational requirement, which is given us by the correctness of the reasoning. There is something normative about the process, but the normativity attaches to the reasoning and not to the intention that emerges from it.

When we acquire beliefs or intentions by reasoning, we may hope we ought to have them or have reason to have them, but we cannot be sure of it. It is often thought that, to justify our beliefs and intentions, we must give reasons for them. But we cannot do that, and it is the wrong idea of justification. The correct way to justify beliefs and intentions is to cite the reasoning that led to them, with its premises. If you are asked to give reasons for your beliefs and intentions, you should interpret this as a request to cite your reasoning.

8. Summary

Like all reasoning, practical reasoning is a process that takes you from some of your existing mental states to a new mental state. Theoretical reasoning concludes in a belief; practical reasoning in an intention. If a piece of reasoning is correct, it concludes in a state that is rationally required by the states it is derived from. But it does not follow that you ought to be in the concluding state, or have a reason to be in it, even if you are in the states it is derived from.

This paper considered only one sort of practical reasoning: instrumental reasoning. If you intend an end, from this intention together with an appropriate belief, instrumental reasoning leads you to intend some means to the end.

The simplest cases of instrumental reasoning are where you believe a particular means is necessary to the end. In those cases, the correctness of the reasoning is ensured by the logical validity of its content.

References

Åqvist, Lennart: 1984, 'Deontic logic', in D. Gabbay and F. Guenthner (eds), *Handbook of Philosophical Logic, Volume II*, D. Reidel, Dordrecht, pp. 605–714.

Bratman, Michael E.: 1987, *Intentions, Plans and Practical Reason*, Harvard University Press, Cambridge, Massachusetts.

Broome, John: 1998, 'Practical reasoning', typescript.

Hume, David: 1978, *A Treatise of Human Nature*, eds L. A. Selby-Bigge and P. H. Nidditch, Oxford University Press, Oxford.

Kant, Immanuel: 1948, *Groundwork of the Metaphysic of Morals*, trans H. J. Paton, in *The Moral Law*, Hutchinson, London.

Makinson, David: 1998, 'On a fundamental problem in deontic logic', in *Deon '98: Proceedings of the Fourth International Workshop on Deontic Logic in Computer Science*,

pp. 3–42.

University of St Andrews
Department of Moral Philosophy
Fife KY16 9AL
Scotland

Symbols

→ (Right arrow)
¬ (Logical negation sign)
O→ (Letter 'O' followed immediately by right arrow)