Peter Baumann, "Involvement and Detachment: A Paradox of Practical Reason", in: Peter Baumann and Monika Betzler (eds.), *Practical Conflicts*, Cambridge: Cambridge University Press 2004, 244-261.

http://www.cambridge.org/us/knowledge/isbn/item1114574/?site_locale=en_US

10

Involvement and Detachment: A Paradox of Practical Reason

Peter Baumann

If the world were perfect, it wouldn't be.
Yogi Berra

I am going to present what I think is an interesting paradox. I will first give an exposition of the paradox (which could be called a "preface paradox for goals" – for reasons that will become obvious soon). I will then deal with some objections and finally discuss a proposed solution to the paradox.

THE PARADOX

A good starting point is, as always, a triviality: Agents want their goals to be realized. More precisely:

(1) If an agent A has a goal G, then A wants that G will be realized (by A or somebody or something else).

This has to be taken in the *de dicto*-sense – hence not in the *de re*-sense of "something would constitute the goal's realization, and A wants it."[1] I use "A wants that p" in the strong sense of "All things considered, A prefers the truth of 'p' to the truth of 'not-p.'"[2]

Having a particular *desire* is, of course, compatible with not wanting its realization. Goals, however, are different: They imply the desiredness of their realization. It is simply incoherent to say something like, "I have this goal but I do not want to realize it."[3]

The following thing seems trivial also:

(2) Agents have many goals (at any particular time as well as throughout their lives).

As does this:

(3) If an agent A has a finite number of goals $G_1, G_2, \ldots, G_n$, then A wants that G_1 will be realized, A wants that G_2 will be realized, . . . , A wants that G_n will be realized.[4]

Can we infer the following (from (1) and (2) or from (2) and (3))?

(4) Agents want that all their goals will be realized.[5]

No. If A has n goals $G_1, G_2, \ldots, G_n$, then it does not follow that A also has the conjunction of all these particular goals as a goal. A might just not "construct" the conjunction of all his particular goals. However, it would be irrational to have all those particular goals and still not accept their conjunction as a goal if one were asked, "Do you want all your goals realized?" Hence, the following seems also true:

(5) If an agent has n different goals, then he has an indefeasible reason to accept the conjunction of his n different goals as a goal.

An indefeasible reason is a reason that is not "trumped" by another reason. More precisely, an indefeasible reason to x is a reason such that there is neither another, better reason not to x nor another equally strong reason not to x.[6] Two reasons are equally strong iff neither reason is better than the other reason and both reasons are commensurable with each other. Two reasons are commensurable iff either one of them is better than the other or this is not true but an additional amount of what makes one of these reasons a good reason (i.e., an additional amount of evidence, etc.) would (ceteris paribus) make this reason better as well as better than the other reason.[7] One could also call the case of equally strong reasons "indifference" of reasons. For example, I might have both a good reason to go to the movies tonight and a good reason not to go to the movies tonight (but rather to stay home). Nothing might speak against accepting both reasons as good reasons; none of the reasons might be better than the other one and both might be equally strong. In such cases, there is nothing wrong with throwing a coin or just picking one option.[8] And if additional evidence would make my reason to go to the movies a better reason, it would (ceteris paribus) also make it better than my reason not to go to the movies.

To be sure: When I say that a person has an indefeasible reason to accept something as a goal, I do not mean to imply that the person is aware of this reason. The person is not necessarily irrational if she does not see that there is an indefeasible reason to accept something as a goal. Other things might be on her mind. A fortiori, the person is not necessarily irrational if she does not make this goal her own. However, the person is irrational if she would not accept the reason or the goal in

case she came to think about it and could, in principle, notice that it is a good reason. Hence, "having an indefeasible reason to accept something as a goal" has a counterfactual aspect.[9]

Now, (5) does not seem quite right as it stands. A's goals might be mutually incompatible. For logical, analytical, nomological, or particular contingent reasons it might be impossible to realize all of the goals from a given set of goals.[10] This seems to suggest the following thesis: If an agent has n different and mutually incompatible goals, then he has no indefeasible reason to accept the conjunction of his n different goals as a goal. Hence, (5) seems false. Is this true? Let us distinguish between two types of incompatibility between goals: incompatibilities that a rational agent can detect (by logical thinking, noticing of empirical data, etc.) and incompatibilities that a rational agent cannot detect. Since we are talking about human agents here, and since humans are even in the best case only nonideally rational agents,[11] it is really non-ideal, "human" rationality that matters here. Now consider an incompatibility of goals that a human agent could detect (let us call such an incompatibility "detectable").[12] A rational (human) agent would detect the incompatibility and change his system of goals in such a way that the incompatibility would disappear. Hence, since the problem I am talking about here is one about rational agency, the existence of detectable incompatibilities does not threaten (5).

What about incompatibilities of goals that a rational human agent cannot – given his limitations – detect ("undetectable" incompatibilities)? For instance, it might be beyond the grasp of even the best mathematicians that one cannot prove a particular theorem and at the same time stick with a certain mathematical theory.[13] Given the plausible principle that "ought" implies "can" and "cannot" implies "not ought," there is no basis for saying that an agent ought to do what he cannot do. Similarly, what we have reason to do is constrained by what we can do. It is not irrational not to do what we cannot do. Some incompatibilities are undetectable, and thus the agent is not irrational if he does not detect and remove such incompatibilities. But does he also have an indefeasible reason to accept the conjunction of his (undetectably) incompatible goals as a goal? I think he does, and here is why. The following general principle seems very plausible: If in cases of type C the agent has an (indefeasible) reason to X and if the agent cannot see a relevant difference between cases of type C and cases of type I, then he also has an (indefeasible) reason to X in cases of type I.[14] Hence, if (5) is true in the case of compatible goals, then it is also true in the case of undetectably incompatible goals.

The upshot of all this is that neither the existence of detectable nor the existence of undetectable incompatibilities between goals threatens (5). The reason mentioned in (5) is indefeasible because there is no better or equally good reason contrary to it. The topic of incompatibility can easily become a red herring; hence, it is a good idea to continue with the main line of the argument.

Now, (2) and (5) lead to

(6) An agent has an indefeasible reason to accept the conjunction of his n different goals as a goal.

(6) together with the rather trivial

(7) If an agent has an indefeasible reason to accept something as a goal, then he has an indefeasible reason to want that that goal will be realized

leads to

(8) An agent has an indefeasible reason to want that all his goals will be realized.[15]

What I said above about indefeasible reasons suggests

(9) If an agent has an indefeasible reason to x, then he does not have an indefeasible reason not to x.

(9) presupposes that reasons are commensurable in the sense that for every pair of reasons R_1 and R_2 it is true that either one of them is better than the other or they are equally good.[16] Hence, if R_1 and R_2 are commensurable and if R_1 is indefeasible, then there is no other reason (different from R_1) that is indefeasible. If R_1 is A's indefeasible reason to x, then there is – a fortiori – no indefeasible reason for A not to x. I come back to this point below and discuss an objection related to it.

(8) and (9) imply

(10) An agent does not have an indefeasible reason not to want that all his goals will be realized.

A fortiori, it is also true that

(10*) An agent does not have an indefeasible reason to want that not all his goals will be realized.

Now, my main problem here – which generates the paradox – is that the following seems also true:

(11) An agent has an indefeasible reason not to want that all his goals will be realized.

(11) is true because something stronger is true:

(11*) An agent has an indefeasible reason to want that not all his goals will be realized.[17]

Why (11) or (11*)? Well, if a person could (and would) realize all of her goals,[18] her life would not only be extremely successful but also extremely boring. This seems true even if we take into account that many bad things that happen are completely beyond our control and have nothing whatsoever to do with the goals we have. Even given that, most human beings would not consider a completely successful life worth living (for them). It is not that we just want to always have goals to pursue (which is also true). Rather, the point is that, in addition, we would not want to always succeed in realizing our goals. In other words, failure, too, is important to us (not only success). Not that anybody would like to have a life full of failures.[19] We do not value failure as such and we have no reason to do so. But if we had the choice (suppose this scenario makes sense) between (a) a life without any failure and (b) a life with some (but not too much) failure, we would rather choose (b). We would even be suspicious of (a). And rationally so. Even though human persons are goal pursuers, they want more: They *want* to be goal-pursuers[20] *and* they want to be goal-pursuers such that not all of their goals are realized. Happiness and the satisfaction of (first-order) preferences is important for us, but it is not everything that matters to us; we also have second-order goals such as the one just described. If all that is true, a person has a very good reason to want that not all of her goals will be realized.[21]

The upshot of all this is that (11*) and (11) are true, too. The reason mentioned in (11) or in (11*) is indefeasible because there is no better or equally good reason contrary to it.

Games offer a good analogy here. For each individual game of, say, Tetris we play, it is true that we want to win it. On the other hand, we do not want to win all of the individual Tetris-games we play. What is the point of playing a game you always win? Such games are boring. Sure, we also do not want to always or often lose. But some failure seems to be part of the attraction of a game. Life is like a game in this respect.[22] Goethe might have been up to something like this when he remarked that we can put up with everything but a couple of nice days. And Yogi Berra later added that if the world were perfect, it wouldn't be.

Now, given all that, the whole problem is this: (11) clearly contradicts (10), and (11*) contradicts (10*). And we seem to have very good arguments for both (10) (or (10*)) and (11) (or (11*)). Both arguments seem valid and use plausible premises. In other words: We have a paradox here – a paradox of practical rationality.

To put it differently: What we have here is both a reason to x and a reason not to x (cf. (8) and (11)). Neither of these reasons is better than the other. How could they be? Are they "equally strong"? In that case, none of these reasons would be indefeasible and the contradiction would disappear. One can have a reason to x and a reason not to x and both reasons might be equally strong reasons and good reasons (the case of "indifference"). One reason might tell me to go to the movies and the other might tell me not to go. There is nothing paradoxical about this. Does our case here differ significantly from cases of indifference? Yes! In the case of indifference, there is no problem at all with accepting both reasons at the same time. One can imagine how additional evidence could make one of the reasons a bit better and thus better than the other reason. And one can, for example, throw a coin to solve the practical problem of what to do. In our case, however, there is no such "peaceful coexistence." We must decide between both reasons. It would be inappropriate to throw a coin in order to settle the problem. In our case, we cannot imagine how additional evidence could make one of the reasons better. And even if there were such evidence, it would not make one reason better than the other. We cannot accept both reasons as equally strong reasons. Since neither one is better than the other, both reasons are indefeasible. Hence, we have a paradox here. What can we do about it? Since it is a paradox, something must be done about it.

Even if we have a hard time solving the paradox, we can already learn an interesting lesson from it. The contradiction between (10) and (11) or between (10*) and (11*) is based on a conflict between two perspectives that we take on our actions: the "involved" perspective of the agent who finds himself in a particular situation and tries to reach his particular goals and the "detached" perspective of the reflective person who takes a step back and looks beyond the limits of particular situations and goals. Both perspectives are important and irreducible; there does not seem to be any good reason to give up one of the perspectives for the sake of the other. In addition, it seems that we just cannot do that. That there is not only a tension between perspectives but a paradox about practical reason makes this duality of perspectives even more interesting.[23]

All this could also have interesting implications for our thinking about moral responsibility. I can give only a very rough hint here. Consider the example of a doctor who performs a lot of operations throughout her life. Since "nobody is perfect," she will make terrible mistakes from time to time ("terrible" because of the consequences). Let us assume that our doctor is excellent and makes very few terrible mistakes. If we think of these individual mistakes, we would tend to blame her and hold her morally responsible. However, if we think of all the operations she was

doing and of the fact that she is such an excellent doctor, and if we add that nobody is perfect and that nobody can be blamed for not being perfect, then we would rather tend not to blame her and not to hold her morally responsible. In other words, it seems that we have both an indefeasible reason to hold her responsible for her mistakes as well as an indefeasible reason not to do so. If that is true, then it is not clear what consequences we should draw with respect to our thinking about moral responsibility.

SOME OBJECTIONS

I discuss three objections now that seem quite strong and suggest that nothing needs to be done here because there really is no paradox.

Working Hard for Success

One possible objection attacks the reasoning behind (11) and (11*). One could argue that a person might realize all her goals but not be bored at all because she often has to put a lot of effort into it. The realization of her goals might very often be hard work. Can't we like games we always win but only with much effort? Can't we enjoy the fact that the realization of our goals is not a trivial task? There is a lot of truth in this objection. However, it is doubtful whether this is a good objection against the ideas behind (11) and (11*). A lot depends on what exactly is meant by "realizing a goal with much effort." Consider the example of somebody who tries to write a good paper. If this is hard for the person, then she will typically need to write several different versions until she can come up with an acceptable (for her) final version. Consider the nonfinal versions. For at least some of these versions it is true that the person had the goal to make it a good version (and thus the final version). Sure, one can write a version of a paper (especially an early one) knowing that it probably will not be the final version and thus without having the goal to make it the final, acceptable version. But if it is really hard for the person to write a good paper, then there will be versions she wrote with the goal in mind to make them good versions but that she could not turn into good versions. In other words: If a person succeeds in realizing a goal but only with much effort, then at least in many cases she will need several attempts to realize her goal. But this, of course, implies that she was not always successful and did not manage to realize all of her goals (e.g., the goal to make the version that later turned out to be the second-last version the last

version). Hence, the above objection does not show what it is supposed to show: namely, that a completely successful life could be interesting. It just is not about a completely successful life.

In addition to this "trial and error" aspect of goal pursuing, there is also very often a "goal/subgoal" structure.[24] To write a good paper, our person has to make a good argument. Suppose her conclusion follows from a set of three premises. To come up with a good argument, she has to make each of these premises plausible. Her overall goal to make a good argument thus consists in a series of (three) subgoals (to make each of the three premises plausible). If it is hard for the author to make the argument, then it is hard for her to reach the subgoals. That is, she will – in some cases at least – have to undergo the pain of trial and error with respect to her subgoals. This gives additional support to the above conclusion that nontrivial success at least often implies some failure.

But are there not cases in which the person does not have to try again and still has to put in a lot of effort? Are there such cases of "hard success without failure"? Consider Mike, who is very strong. He set himself the goal to lift his piano and hold it in the air for thirty seconds. He tries, puts a lot of effort into it, sweats, and screams – and succeeds. It seems that there is no failure involved here. He did not aim to lift the piano without any effort. On the contrary, the whole point was that it would not be easy. It seems that there really are cases such as this one. However, many cases are not like this. In such cases – which are usually also the more important ones – we do not have hard success without any failure. So, the upshot of all this is that we have good reason to stick with (11) and (11*). The paradox does not go away.

Future Goals

So far I have not said anything about the temporal aspect of having goals. One might suspect that this could lead to another objection. Let us start with some relatively unproblematic remarks about goals and time. Some goals are "time-specific" in the sense that their content indicates a time within which or at which the relevant event is supposed to happen. I might have the goal to reach you on the phone between now and midnight or at midnight. Some goals are very vague with respect to the relevant time, such as the goal of somebody who has never been to Mexico to go there some day. In this case, the relevant time might be any time during the person's life. Apparently, there are also goals that are not time-specific at all. A poet might have the goal that some day somebody will appreciate

his poems. Goals can be realized a long time (even an indefinitely long time) after the death of the person who has this goal. Whether goals are time-specific or not, they are always held by a person at a certain time or during a certain time. And our goals change over time, perhaps not all of them (such as the goal to stay out of trouble) but many of them. Some goals disappear because they have been realized and others disappear because we could not realize them. We often give up goals because we decide they are not worth it or because we just forget them (which is an unconscious way of giving up a goal). In addition to that, we continue to develop new goals as we go along. We usually do so until the end of our life. This also explains why there are in almost everybody's life some goals that remain unfulfilled: We die before they could be realized. This fact, however, is much less interesting than an objection lurking behind it in the background.

Let us assume that a person has a set of goals S_1 between t_1 and t_2 and a set of goals S_2 at some later time, between t_3 and t_4.[25] Let us further assume – for the sake of simplicity – that no goal is a member of both S_1 and S_2. Let "P_1" refer to the person between t_1 and t_2 and "P_2" to the person between t_3 and t_4. Couldn't P_1 say "I want all the goals I now have realized but not all of the goals that I will have in the future [P_2's goals]"? P_1 has the set of goals S_1 but not the set of goals S_2. Hence, it is only true that

(12) P_1 has an indefeasible reason to want that S_1 will be realized (cf. (8))

and

(13) P_1 does not have an indefeasible reason not to want that S_1 will be realized (cf. (10)).

It is, however, also true that

(14) P_1 does not have an indefeasible reason to want that S_2 will be realized.

The reason is simply that P_1 does not have these goals (S_2). P_1 can coherently allow for the nonrealization of S_2. We can generalize this:

(15) P_1 does not have an indefeasible reason to want that all the goals she will ever have in life ($S_1, S_2, \ldots, S_n$) will be realized.

The argument behind (11) and (11*) also shows that

(16) P_1 has an indefeasible reason not to want that all the goals she will ever have in her life ($S_1, S_2, \ldots, S_n$) will be realized.

The whole point of the objection is that (13) (as well as (12)) does not contradict (16). Hence, there is no paradox. A person has an indefeasible reason to want the realization of her present goals. She can – without contradiction – at the same time have an indefeasible reason not to want the realization of all her present as well as future goals. The paradox is an apparent one and exploits an ambiguity of the phrase "all the goals": "all the present goals" versus "all the present as well as future goals."

I can think of two replies to this objection. First, one can argue that even with respect to the present goals of a person, the paradox arises because the person has an indefeasible reason not to want all of her present goals realized:

(17) P_1 has an indefeasible reason not to want that S_1 will be realized.

And (17) contradicts (13). Hence, our paradox does not go away.

There is another reply. It denies (14) and (15). P_1 does indeed have an indefeasible reason to want S_2 realized even though P_1 does not (yet) have the set of goals S_2.[26] S_2 is not just any set of goals. It is not like another person's goals. It is a set of some of P_1's future goals. P_1 is the same person as P_2. Hence, P_1 has a good reason to be interested in P_2's well-being, that is, in his own future well-being.[27] He has good reason to "identify" himself with his future self. We can say that

(18) P_1 has an indefeasible reason to want that S_2 will be realized

and

(19) P_1 does not have an indefeasible reason not to want that S_2 will be realized.

We can generalize this:

(20) P_1 has an indefeasible reason to want that all the goals she will ever have in life ($S_1, S_2, \ldots, S_n$) will be realized

and

(21) P_1 does not have an indefeasible reason not to want that all the goals she will ever have in life ($S_1, S_2, \ldots, S_n$) will be realized.

(21), of course, contradicts (16). Hence, the paradox is still with us.

Conditional Goals

Here is still another major objection. Don't we have conditional goals? More precisely: Couldn't it be true that some of our goals have the following logical form (with parantheses indicating scope):

A has the goal that (p, but if and only if not all other goals are going to be realized).[28]

We need to assume not that all our goals are conditional but only that some of them are (we can leave open how many would be enough). Conditionality of this sort implies some kind of holism. One cannot understand and pursue one (conditional) goal without having other goals in mind. To be sure: Having the goal that (p, but if and only if not all

other goals are going to be realized) does not imply that one has a certain other particular goal. But it requires that one keeps one's other goals in mind, whatever they are. Conditionality also implies a certain kind of indeterminacy. If A has just two different goals, namely,

the goal that (p, but if and only if not all other goals are going to be realized)

and

the goal that (q, but if and only if not all other goals are going to be realized),

then we can neither infer that A will try to bring it about that p nor that A will try to bring it about that q. What we can say is that A will try to bring it about that either p or q. This again shows the nonatomistic nature of conditional goals: The person who pursues them has complex sets of goals in mind and not just lots of isolated individual goals.

How does conditionality bear on our thesis that there is a paradox about goals? Well, it seems possible that all the goals (conditional and unconditional) of a person are being realized without any threat of boredom or life losing its point. Let us (for the sake of simplicity) assume that a person has just two goals, both of them conditional:

the goal that (p, but if and only if not all other goals are going to be realized)

and

the goal that (q, but if and only if not all other goals are going to be realized).

Let us further assume that she brings it about that p but not q. This means that all her goals are realized even though she did not bring it about that both p and q. The latter seems to guarantee that life will not be boring or lose its point. Hence, there is no way to get to (11) or (11*) and the paradox is gone. By the way, one might reply that this objection also cuts the other way: Isn't it really *p and q* that matters here? Shouldn't we say that there still is an important sense in which the person has not realized all her goals? I do not think so: These really are conditional goals.

The problem with this conditionality objection is that it is hard to see why one should think that we have such conditional goals. The main motivation to assume this seems to be that it would avoid the paradox. This, however, is an ad hoc argument and thus not a good one. Moreover, it seems more psychologically realistic to assume that people do not have these kinds of goals; at least, it seems that many of us do not and for those of us who do, it seems to be an exception rather than the rule. Even

if some of us sometimes have such conditional goals, this would not make a difference big enough to threaten the ideas behind (11) or (11*).

To avoid a possible misunderstanding, I am not saying that we do not have goals such as this one:

Ann has the goal that (she makes it to the movies tonight if and only if she does not reach her goal to convince Jack to prepare dinner for her).

But is this not a perfect example of a conditional goal? No, not in the above sense. For conditional goals in that sense, the nonattainment of other goals as such is an essential part of the content of the goal, no matter what these other goals are. Not so in the example just given. Here the individuality of the other goals matters but not their realization as such and independently from their content. We can reformulate the description of Ann's goal without loss of something essential in the following way:

Ann has the goal that (she makes it to the movies tonight if and only if she does not convince Jack to prepare dinner for her).

Here, mention of "goal attainment" is not an essential part of a description of the content of the goal. However, it is essential in the case of conditional goals. Hence, we cannot reformulate the description of conditional goals in the same way in which we can reformulate a description of Ann's goal. This shows that Ann's goal is not a conditional goal in the sense that is relevant here. Furthermore, Ann's goal presupposes that she has two individual goals (make it to the movies; convince Jack to cook) such that success with respect to one of them is linked with failure with respect to the other one. There will thus be some failure. However, the idea of conditional goals is introduced here as a way to circumvent the possibility of failure.

I want to end this section with three shorter objections. The first has to do with (9) above. I said that (9) presupposes the commensurability of reasons. But couldn't one deny the commensurability of reasons, let (9) collapse, and say that the reasons behind (10) and (11) (and (10*) and (11*)) are incommensurable? If yes, then the paradox disappears and an incommensurability between the reasons mentioned in (8) and (11) would take its place:

(8) An agent has an indefeasible reason to want that all his goals will be realized

and

(11) An agent has an indefeasible reason not to want that all his goals will be realized

would both be true.

This would certainly be an interesting alternative and perhaps even still another argument for the possible incommensurability of reasons for action.[29]

However, it would, I think, not be very convincing. Why should one assume that there is an incommensurability here – just because this would avoid the contradiction? There might even be incommensurabilities somewhere else (we can leave that open), but why here? To assume incommensurability between the reasons mentioned in (8) and (11) would seem ad hoc and perhaps even question begging (against the paradox). I can see no independent argument that would show that we face incommensurable reasons here. One certainly cannot avoid a contradiction just by assuming that it really is a case of incommensurability.

It also does not help at all to take the contradiction as evidence against the idea that our two reasons are both indefeasible reasons. This would, again, beg the question against the argument that there is a paradox here. Similar problems arise if one considers the above contradiction as a reductio of the assumptions behind (11) or (11*) or, alternatively, of the assumptions behind (10) or (10*). This would be ad hoc, motivated only by the fact that rejecting those assumptions avoids the contradiction. One would need independent reasons showing that the arguments leading to (11) and (11*) – or to (10) and (10*) – are not acceptable. It is hard to see any reasons such as that. It seems that there really is a paradox. So what can we do?

WHAT IS TO BE DONE?

First, this paradox resembles the well-known preface paradox for beliefs.[30] It is quite reasonable for an author of a book to assume (in the preface) that not everything she says in the book is true. We know we are not infallible, and it is only rational to assume that we make mistakes from time to time. On the other hand, it can be perfectly rational for an author to sincerely believe every single thing she says in the book; in this case it is also rational to believe that everything said in the book is true. This, of course, leads to a contradiction; the same contradiction arises for everybody who holds a plurality of beliefs. You cannot escape the preface paradox by writing no books or only books without prefaces.

There is a convincing solution to the preface paradox for beliefs that goes back to Frank Ramsey.[31] Beliefs are no yes/no matter but rather allow for different degrees. This allows us to assign subjective probabilities close to 1 (but below 1) to belief contents such that (according to the probability calculus) the conjunction of many of those contents is assigned a very low probability (especially if they are probabilistically independent from each other). The person can thus hold each and every particular belief (even assigning a very high subjective probability to it) and still not bet much on the conjunction of those contents.

One might expect or hope that there is an analogous solution to our ("preface") paradox for goals. Unfortunately, however, nothing like that will do, even if we assume degrees of desiredness (rational or not) similar to degrees of belief. The attempted analogy breaks down because there is no way to compute the degree of desiredness of a conjunction exclusively on the basis of the degree of desiredness of the conjuncts. To have a coffee now might have a desiredness of .9, and to read a novel now one of .5. It simply does not follow that, for example, the desiredness (rational or not) of reading a novel and at the same time having a coffee is below each single desiredness (say, .45). There are no implications whatsoever as to the desiredness (rational or not) of the conjunction. The combination of coffee and novels might have any possible degree of desiredness without rendering the person irrational. Hence, the strategy to solve our problem by exploiting the analogy to the preface paradox for beliefs and to a Ramseyan solution of it will not work. The question remains an open one: What can we do?

Notes

I would like to thank Monika Betzler, Luc Bovens, Jon Cameron, Ruth Chang, Michael Clarke, Gisela Cramer, Christoph Fehige, Jörg Fehige, Harry Frankfurt, Ulrike Heuer, Luis Eduardo Hoyos Jaramillo, Keya Maitra, Bob Plant, Wlodek Rabinowicz, Abdul Raffert, Neil Roughley, Thomas Schmidt, Jörg Schroth, and an audience at the University of Memphis for discussions and for comments on earlier versions of this paper.

1. Sentences such as (1) could raise the question of whether we are allowed to quantify into intensional contexts and what the exact logical analysis of such sentences is. However, we do not have to deal with this question here since nothing essential hinges on it and the crucial point is pretty clear.
2. Some goals are recurrent: At t_1, I want to eat; after eating, at t_2, I do not want to eat any more; some time later, at t_3, I regain the goal to eat something. Has my goal to eat been realized or not? The situation at t_2 seems to suggest a positive answer, whereas the situation at t_3 seems to suggest a negative answer. If one looks more closely at the situation, one realizes that the answer must be positive: At t_1, my goal was, more precisely, to eat something *soon*. At t_2 this goal has been realized. My goal at t_3 was to eat something soon. This goal is of

the same type as the former one (eating soon, whenever that might turn out to be). However, it is a different particular goal because t_1 is different from t_3 and the identification of the time belongs to the full description of the respective goal. Hence, it is important to distinguish between the realization of a particular token of a goal (eating at t_2, etc.) and the realization of a type of goal (eating soon, whenever that might turn out to be). The former sense of "goal" is the relevant one here. One other point to be mentioned is that goals differ with respect to their importance for the person. For the sake of simplicity, I disregard this complication here; it would only make the argument more complicated than necessary.

3. This reminds one of Moorean sentences, such as, "It's raining but I don't believe it" (see Moore 1952: 542f.). The incoherence above differs from Moore's paradox insofar as it is not tied to the first person or the present tense.
4. The case of infinitely many goals is more tricky but also much less realistic; hence, we do not have to deal with it here. The attributions of wants above are *de dicto*, again, and not *de re*.
5. This is taken in the *de dicto*-sense.
6. The word "indefeasible" could create the impression that indefeasible reasons *cannot* be overriden by other reasons. However, given the explanations above, it only follows that indefeasible reasons *are in fact* not defeated by other reasons. One could alternatively say: *Given* the reasons the person has, an indefeasible reason cannot be defeated. Indefeasible reasons are not overriding reasons (see Raz 1986: 339). Somebody has an overriding reason to x iff he has a better reason to x than not to x and if this reason is a good reason to x (rather than no or no good reason to x). To put it differently: A reason to x is an overriding reason to x iff it is stronger than any reasons not to x. I confine myself here to indefeasible reasons, but one could, perhaps, make a similar argument in terms of overriding reasons.
7. See Raz 1986: 325 (who does not accept this as a definition, though). According to the completeness axiom of classical decision theory, either two options are equally good or one is better than the other (see, e.g., Luce and Raiffa 1957: 223, 25).
8. See Ullmann-Margalit and Morgenbesser 1977 for the idea of "picking" rather than "choosing" an option.
9. One could call this use of "reason" an "objective" use and contrast it with a "subjective" use that implies that the person is in fact aware of the reason. (5) is much less convincing if one takes "reason" in the subjective sense.
10. Two goals are logically incompatible if their contents contradict each other; two goals are analytically incompatible if their contents "analytically contain" a contradiction (let us not worry about analyticity here); two goals are nomologically incompatible if there is a law of nature excluding the realization of both goals; two goals are incompatible for particular contingent reasons if they cannot both be realized because of contingent singular circumstances. By the way: Neither does incompatibility imply incommensurablity nor does the latter imply the former.
11. See, e.g., Simon 1983: 3ff.

12. The distinction between what we can and what we cannot detect is, of course, very vague; in many cases it is not clear at all whether the agent could or could not do it.
13. I owe this point and example to Ruth Chang.
14. In this sense of "subjective" (not to be confused with the one mentioned above), rationality is subjective.
15. This is taken in the *de dicto*-sense and collectively, not distributively.
16. See, however, Chang 1997a: 25–7.
17. This is taken in the *de dicto*-sense and collectively, not distributively.
18. To avoid possible misunderstandings: I do not have the case in mind in which a person can realize a goal immediately or without any effort (see the first objection below on the aspect of effort).
19. For an argument on why an agent cannot always or very often fail to reach his goals, see Baumann 1996: 50–7.
20. See, e.g., Frankfurt 1992: 13ff. and Ullmann-Margalit 1992: 73ff.
21. For the sake of simplicity, I am assuming here that the person is never ignorant or in error about the sucesses and failures of the actions she has performed. Nothing essential changes (it only gets more complicated) if we make the more realistic assumption that people are sometimes ignorant (we forget things) or in error about their past failures and successes. A lot would change if we would assume that people are often or always ignorant or in error about the outcome of their past actions. This, however, is not a realistic assumption at all and we can disregard it for this reason. To be sure, very often an agent cannot know in advance whether he will succeed or fail to reach the goal he is just trying to realize. However, past experiences gives him a lot of evidence about what can be expected for the future. So, the element of ignorance about some future outcomes does not make an extremely successful life substantially less boring.
22. This also tells us something about the idea of omnipotence. To be sure, the completely successful agent as such is not necessarily omnipotent: He can realize all the goals he in fact has, but not necessarily all possible goals. However, a completely successful agent would probably develop more and more goals and thus "converge" toward omnipotence. My argument suggests that omnipotence is not a good ideal for human beings.
23. See, of course, Nagel 1986.
24. See Bratman 1987 for the importance and complex structure of plans.
25. One could be tempted to think that in the last analysis persons have goals at points of time rather than during time intervals. This, however, is highly unrealistic. No human being can have a goal just for a millisecond. If an extremely short time interval is too short, then a fortiori an extensionless point of time is much too short.
26. That I want that my future goal F (that p) will be realized does not imply that I already have goal F. Suppose that I do not want that p now (or even want that not-p); suppose further that I know that I will want that p in the future. Hence, we can say that I have a want of the following form:

    (a) I now want that (p, if/ as soon as I have the goal that p).

    We also assumed that

(b) I will have the goal that p.

(a) and (b) do, of course, not imply

(c) I now want that p.

Hence, we can also avoid the conclusion that I both do want that p and do not want that p. Analoguously, I might want that you reach your goals without sharing your goals.

27. Some philosophers – such as Parfit 1984 – deny both this view of personal identity and its implications for practical rationality. I cannot go into these discussions here. It is clear that there is a certain (rational) bias toward the present; this, however, does not effect my reply to the objection above.
28. Goals can be "conditional" in still another sense: "If and only if not all other goals are going to be realized, then A has the goal that p." It is, of course, impossible that all our goals are like that. Everything would "hang in the air." It would be completely indeterminate what the person would be trying to do. So, let us assume that only some of our goals are like that. Does this case show anything interesting about our paradox? I do not think so. It tells us only that there are certain conditions under which persons develop certain goals.
29. See the contributions in Chang 1997b. Chang 1997a prefers the expression "incomparability" to "incommensurability"; nothing substantial depends on it and I use the latter expression because talk about "incomparable" goals or reasons sounds a bit odd. I would, of course, be pleased if friends of incommensurability would find my argument useful for their own purposes.
30. See Makinson 1965: 205ff.
31. See Ramsey 1990: 111.

Bibliography

Baumann, Peter. 1996. Mephistos Problem: Über den Zusammenhang von Absichten und Handlungserfolgen. In Christoph Hubig and Hans Poser (eds.), *Cognitio Humana – Dynamik des Wissens und der Werte. XVII Deutscher Kongress für Philosophie (Leipzig 1996)*. Leipzig: Institut für Philosophie, vol. 1, 50–7.

Bratman, Michael. 1987. *Intention, Plans, and Practical Reason.* Cambridge, Mass.: Harvard University Press.

Chang, Ruth. 1997a. Introduction. In Ruth Chang (ed.), *Incommensurability, Incomparability, and Practical Reason.* Cambridge, Mass.: Harvard University Press, 1–34.

Chang, Ruth (ed.). 1997b *Incommensurability, Incomparability, and Practical Reason.* Cambridge, Mass.: Harvard University Press.

Frankfurt, Harry. 1992. On the Usefulness of Final Ends. *Iyyun: The Jerusalem Philosophical Quarterly* 41: 3–19.

Luce, R. Duncan, and Howard Raiffa. 1957. *Games and Decisions: Introduction and Critical Survey.* New York: Wiley.

Makinson, D.C. 1965. The Paradox of the Preface. *Analysis* 25: 205–7.

Moore, George Edward. 1952. A Reply to My Critics. In Paul Arthur Schilpp (ed.), *The Philosophy of G. E. Moore* (The Library of Living Philosophers, vol. IV, 2nd ed.). New York: Tudor, 533–677.

Nagel, Thomas. 1986. *The View from Nowhere.* Oxford: Oxford University Press

Parfit, Derek. 1984. *Reasons and Persons.* Oxford: Clarendon.

Ramsey, Frank P. 1990. Knowledge. In Frank P. Ramsey, *Philosophical Papers*, ed. David Hugh Mellor. Cambridge: Cambridge University Press, 110f.

Raz, Joseph. 1986. *The Morality of Freedom.* Oxford: Clarendon.

Simon, Herbert A. 1983. *Reason in Human Affairs.* Stanford, Calif.: Stanford University Press.

Ullmann-Margalit, Edna. 1992. Final Ends and Meaningful Lives. *Iyyun: The Jerusalem Philosophical Quarterly* 41: 73–82.

Ullmann-Margalit, Edna, and Sidney Morgenbesser. 1977. Picking and Choosing. *Social Research* 44: 757–85.

Practical Conflicts

New Philosophical Essays

Edited by

PETER BAUMANN

University of Aberdeen

MONIKA BETZLER

Universität Göttingen

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

© Cambridge University Press 2004

This book is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of Cambridge University Press.

First published 2004

Printed in the United States of America

Typeface ITC New Baskerville 10/13 pt. *System* LaTeX 2ε [TB]

A catalog record for this book is available from the British Library.

Library of Congress Cataloging in Publication Data

Practical conflicts : new philosophical essays / edited by Peter Baumann
and Monika Betzler.
p. cm.
ISBN 0-521-81271-2 – ISBN 0-521-01210-4 (pbk.)
1. Practical reason. 2. Conflict (Psychology) I. Baumann, Peter. II. Betzler, Monika.
BC177.P72 2004
128′.4 – dc22 2003061213

ISBN 0 521 81271 2 hardback
ISBN 0 521 01210 4 paperback

Contents

Foreword *page* vii

1 Introduction: Varieties of Practical Conflict and the Scope of Practical Reason — 1
Peter Baumann and Monika Betzler

2 Willing the Law — 27
J. David Velleman

3 The Myth of Egoism — 57
Christine M. Korsgaard

4 Thinking about Conflicts of Desire — 92
Henry S. Richardson

5 Putting Together Morality and Well-Being — 118
Ruth Chang

6 The Second Worst in Practical Conflict — 159
Isaac Levi

7 Personal Practical Conflicts — 172
Joseph Raz

8 Sources of Practical Conflicts and Reasons for Regret — 197
Monika Betzler

9 Conflicting Values and Conflicting Virtues — 223
Nicholas White

10 Involvement and Detachment: A Paradox of Practical Reason — 244
Peter Baumann

11 Outcomes of Internal Conflicts in the Sphere of *Akrasia* and Self-Control 262
Alfred R. Mele

12 Are There Insolvable Moral Conflicts? 279
Peter Schaber

13 Moral Dilemmas of Transitional Justice 295
Jon Elster

14 Do Conflicts Make Us Free? 316
Barbara Guckes

List of Contributors 334
Name Index 336
Subject Index 338

Foreword

The idea for this anthology grew out of many discussions and jointly taught seminars on matters of values, autonomy, practical reasons, and rational choice at the Philosophy Department at Göttingen University (Germany).

All of the papers collected in this anthology are original contributions except Christine Korsgaard's "The Myth of Egoism," which was delivered as the 1999 Lindley Lecture at the Department of Philosophy at the University of Kansas. We thank Christine Korsgaard and the Department of Philosophy for allowing us to reprint the lecture here, which, so far, has been accessible only to a relatively small audience.

Many people have encouraged us to pursue this book project. We are particularly grateful to the contributors and special thanks go to Rüdiger Bittner and R. Jay Wallace for their ongoing support and many helpful suggestions. Many thanks go to Jon Cameron for producing the index. We also thank Terence Moore of Cambridge University Press for his valuable guidance during the preparation of this volume.