The Southern Journal of Philosophy (1999) Vol. XXXVII

Normative Expectations, Intentions, and Beliefs

Katarzyna Paprzycka
University of Southern Mississippi

G. E. M. Anscombe noticed that statements naturally interpreted as expressions of intentions (such as "I am going for a walk") could also be interpreted as expressions of beliefs.[1] Indeed, philosophers of action have been struck by the close conceptual ties between intention and belief. So H. P. Grice observes that it would be awkward to express an intention of doing something while denying that one believes that one will do it.[2] He suggests that an agent who intends to do something must believe that he will do it. This position has been subject to much scrutiny. In fact, a whole spectrum of views on the relation between intention and belief has been occupied—from the strong position on which the relation between them is that of logical implication, through weaker positions (such as the view that the relation is that of conversational implicature, or that what is implied is only the belief that the agent will try to do what he intends to do), to the position that one can intend to do something while believing that it is impossible for one to do it.

The aim of this paper is to look at the relation between intention and belief from a different angle. I will examine a different pair of concepts, namely, normative and predictive expectations, which bear an important resemblance to the pair which gave so much heat to the debate. It is telling, for example, that Anscombe's observation also holds for normative expectations. Statements that can be interpreted as expressions of normative expectations (such as "You will walk the dog") can also be interpreted as expressions of predictive expectations (beliefs). After recalling further reasons to think that the resemblance is of a close rather than a distant family sort

---

Katarzyna Paprzycka received her Ph.D. from the University of Pittsburgh. She is currently a full-time mother teaching part-time at the University of Southern Mississippi. Her articles span the areas of philosophy of action, philosophy of science, metaphysics, and social philosophy. Her main project is a book developing a nonindividualist responsibility-based theory of action centered around the concept of normative expectation rather than intention.

(section 2), I propose a certain understanding of the reasonableness of normative expectations (section 3). This notion of reasonableness will give us preliminary tools to conceptualize the relation between reasonable normative expectations and rational predictive expectations (section 4). In section 5, I conclude by considering the light the discussion can shed on the original debate.

I. INTENTIONS AND BELIEFS

It is awkward to announce an intention to φ and at the same time to deny that one will φ. This might be thought to indicate that there is a conversational implicature in play, that the announcing of the intention to φ implicates that one believes that one will φ.[3] As has been persuasively argued by David Pears, this view must be supplemented by a more substantive conception of the relationship between intention and belief.[4] Such an account has been classically advanced by Grice who held that it is a necessary condition of the agent intending to φ that she believe that she will indeed φ.[5]

The strong view on the relation between intention and belief has several advantages. First, it grounds and explains the conversational implicature. Second, it underlies one of the prominent roles of intention, namely, its coordinative function in planning. One of the upshots of forming an intention to φ is that the agent can then count on her φing and develop further plans accordingly. This role of intentions would be impaired if intentions to φ were accompanied by beliefs that one will not φ.[6] Finally, the view firmly establishes the difference between intentions and mere wishes. It has been argued that if intentions were to be disassociated from future beliefs, there would be a danger of them becoming more like idle wishes or hopes.

At the same time, however, the strong view seems to many to be simply too strong.[7] Jane may be intending to attend a yoga class but not be convinced that she will actually be able to as she has heard rumors that the class may be cancelled. Don may be intending to make ten carbon copies as he is writing on a paper but not know if he is succeeding. It has even been suggested that there are cases where one can intend to perform an action that one believes to have little or even no chance of performing, as when a policewoman intends to shoot a runaway despite knowing that he is simply too far off for her to have a good chance of shooting and not missing.

It is sometimes replied that one must pay attention to the content of the agent's intention in those cases. The policewoman does not intend to actually shoot the runaway (in the sense of shooting and hitting him), but rather she may only intend to *try* to shoot him. Jane may be intending not so

much to attend the yoga class, but only to attend it *if* the class is not cancelled. Don may be intending to *try* to make ten carbon copies as he is writing on a paper. In such cases, the agents will also hold a weaker belief that they will try to succeed. In this way, the strong thesis that intention implies belief is saved.[8]

The debate is not closed however. There are those who believe, as does Pears, for example, that the content of the intention to try to φ is really the same as the content of the intention to φ, in which case the strong view falls.[9] There are those like Hugh J. McCann and Kirk Ludwig who take different routes to reach the same conclusion as Pears. They both appeal to cases where agents can perform some actions intentionally while believing that it is impossible for them to succeed. They use different routes to argue that in those cases the agent intended to perform the action.[10]

The debate is clearly very complex and it is getting more and more involved with the resolution of particular issues in the philosophy of action, in particular the relation of the concept of intention not only to that of belief but also to that of intentional action. Rather than getting involved in resolving the tangles, I propose to look at a pair of related concepts: the concepts of normative and predictive expectations. I will discuss the conditions under which it would be natural to think that it is rational for someone who holds another to a normative expectation to expect (in the predictive sense) that he will fulfill the normative expectation. I will then say how that might clarify the relationship between intentions and beliefs.

II. NORMATIVE EXPECTATIONS AND INTENTIONS VS. PREDICTIVE EXPECTATIONS AND BELIEFS

The concept of a normative expectation, while an old inhabitant of sociologists' and social psychologists' conceptual frameworks, has only recently established itself in the philosophical literature.[11] It is best introduced in contrast to the notion of a predictive expectation.

I may expect of a person that he be polite toward people around him and yet, taught by prior experience, expect that he will not. That no contradiction is involved is clear. Two different concepts of expectation are involved: The former expectation is normative, the latter predictive or descriptive. Greenspan characterizes the distinction in terms of the notion of prediction, on the one hand, and the notion of holding the agent to a demand, on the other.[12] Wallace ties the notion of normative expectation with various reactive emotions we are

inclined to feel when the expectation is frustrated (guilt, resentment).[13] The distinction can be sharpened by appealing to the metaphor of direction-of-fit introduced by Anscombe.[14] Predictive expectations (that *p*), like beliefs, have a mind-to-world fit: If it is the case that not-*p*, the fault lies with the expectation. Normative expectations (that *p*), like intentions, desires, etc., have a world-to-mind fit: If it is the case that not-*p*, the fault is with the world which ought to be changed accordingly. More precisely, we can say that a person predictively expects that *p* when (among other things) he is disposed to dismiss the expectation as having been wrong if not-*p*. A person β expects (in the normative sense) of another person α that *p* when β is disposed to sanction α's failure to bring about *p*. The notion of sanctions is broad enough to include the reactive emotions[15] as well as sanctions of lesser moral magnitude like feeling dissatisfied or disappointed by oneself or by another, criticizing oneself or others, and, on the side of positive sanctions, feelings of satisfaction and accomplishment.

The distinction between normative and predictive expectations is a close kin to that between intentions and beliefs. For one, as we saw, the distinction between normative and predictive expectations can be sharpened by appealing to the same metaphor of direction-of-fit introduced by Anscombe to draw a distinction between intentions and (some) beliefs. Predictive expectations like beliefs have a mind-to-world direction of fit, normative expectations like intentions have a world-to-mind direction of fit. In addition, what distinguishes normative and predictive expectations centers on a normative element present in the former. When a mother expects her son to be home by 5 p.m., she places a demand on him. It is widely recognized that intentions also involve a normative element, that of commitment.[16] (Since predictive expectations are generally regarded as simply beliefs about the future, I will use the term "belief" rather than "expectation in a predictive sense." Henceforth, unless otherwise indicated the terms "expectation" and "expect" will be understood in the normative sense.)

One of the differences between intentions and normative expectations is that the former are usually thought of as being first-personal (I cannot intend you to go to the movie)[17] while the latter are paradigmatically, though not essentially, interpersonal (you can expect of me that I finish the job on time and I can expect of myself that I do so). This might lead one to be inclined to think that intentions are normative expectations of oneself. That may be ultimately a fruitful stance to take in capturing at least some functions of intentions. For our present purposes, it is sufficient for us to think that intentions involve normative expectations of oneself. When I intend to finish my novel by January, I expect of myself that I do so. Other things being equal, I will blame myself, feel disappointed or just feel

bad if I do not, and I will feel a sense of pride and accomplishment if I do.

III. REASONABLENESS OF NORMATIVE EXPECTATIONS

Before proceeding, it is important to clarify the notion of reasonableness of normative expectations. I will later use the concept as a springboard for some conjectures about what beliefs it would be rational for the expector (the person who expects something of another) to hold.

The notion of reasonableness of normative expectations is very hard to capture precisely. One reason for this is that two component concepts are involved. This can be seen by reflecting on two different ways in which a normative expectation may be unreasonable. One reason an expectation of a person may be unreasonable is that it is *not within her power* to do what she is expected to do. For instance, it would be unreasonable to expect of an athlete who broke a leg that she take part in a race, of a blind person that he drive a car, or of a newly arrived foreigner that he speak like a native. Another reason why an expectation may be unreasonable is of a different nature. It may be that the person has the power to do what we expect of her but it may be nonetheless *inappropriate* for us to expect it of her. Let us suppose that you have a relatively ordinary relationship with your neighbors. You are polite to one another and occasionally help one another out in neighborly matters. There are (many) expectations that it is simply inappropriate for you to hold them to and not because it is not within their power to fulfill them. For instance, it would be inappropriate for you to expect them to regularly mow your lawn, to do your shopping, etc.

These two kinds of cases exemplify two different though equally fundamental concerns with the reasonableness of normative expectations. For want of better terminology, I shall speak of *reasonableness*$_A$ (agent-centered reasonableness) to capture the first sort of case, and of *reasonableness*$_N$ (specifically normative reasonableness) in the second kind of case. It is possible for an expectation to be reasonable$_A$ but unreasonable$_N$. Your expectation of your neighbors that they do your shopping would be reasonable$_A$ (because it is within their power to do so) but it would be highly unreasonable$_N$ (inappropriate) for you to expect it of them. It is also possible for an expectation to be reasonable$_N$ but unreasonable$_A$. A teacher may reasonably$_N$ expect of his student that she turn in the assigned paper on time but the expectation may be unreasonable$_A$ in view of the fact that the student has been taken to the hospital. Henceforth, I will focus exclusively on reasonableness$_A$, as this is the

concept that will allow us to establish a tie between normative and predictive expectations.

No substantive claim will depend on exactly what it means for a performance to be within the agent's power, I shall therefore refrain here from an elucidation of the concept.[18] It will be important, however, to make a distinction between an expectation being prima facie unreasonable$_A$ and it being made unreasonable$_A$ by the occurrence of a defeating condition. There is a large class of normative expectations that are prima facie unreasonable$_A$. It is prima facie unreasonable$_A$ to expect of a person that she make sure that 2 + 2 = 4. It is prima facie unreasonable$_A$ to expect of a person that she prevent the law of free fall from applying. It is prima facie unreasonable$_A$ to expect of a person that he speak all the known languages. All these are instances of expectations regarding performance types that we generally think are not within a normal person's power to accomplish.

Aside from such cases, there are instances where otherwise reasonable$_A$ expectations may become unreasonable$_A$ in the presence of special circumstances, *defeating conditions*. Thus, the prima facie reasonableness$_A$ of the expectation of a student that he complete a class project may be defeated if the student becomes seriously ill. His becoming ill is a defeating condition; it renders it outside of his power to fulfill the expectation. The prima facie reasonableness$_A$ of the doctor's expectation of a patient that he raise his arm may be defeated if the patient suffers from paralysis. Paralysis is a condition that makes the fulfillment of the expectation in question lie outside of the agent's power. In general, in most cases, it will be true that if it is reasonable$_A$ to expect of α that α φ, it is within α's power to φ.[19]

I should point out that the notion of reasonableness$_A$ thus construed does not depend on the agent's beliefs.[20] Consider the following case: Suppose Jim has good reasons to believe that Tamara suffered a paralysis in her left arm. In fact, however, her arm is in quite a normal state. The question is whether Jim's expectation of Tamara that she raise her arm is reasonable$_A$. On the understanding of reasonableness$_A$ I intend here, we should say of such a case that it would be *reasonable*$_A$ for Jim to *expect* of Tamara that she raise her arm, even though it is quite *reasonable*[21] (rational) for him to *believe* that it would be unreasonable$_A$ to expect it of her. The notion of reasonableness$_A$ of normative expectations thus construed does not depend on the expector's epistemic position. In other words, if it is reasonable$_A$ for β to expect of α that α φ (at t) then it is reasonable$_A$ for γ to expect of α that α φ (at t).[22] This is not to deny that there is not a notion of reasonableness (as applied to beliefs, including beliefs about what is reasonable$_A$) that is epistemically responsive. Even if it were reasonable (rational)

for us to believe that an expectation of Tamara that she raise her arm would be reasonable$_A$ (at t), it is unreasonable (irrational) for Jim to believe that the expectation would be reasonable$_A$ (at t). (To avoid confusion, I will use the term "rational" rather than "reasonable" to apply to beliefs.)

IV. WHEN IS IT RATIONAL TO EXPECT THAT HE WILL DO AS EXPECTED?

Let us now consider the connection between a person's expecting something of another (normative expectation) and his belief (predictive expectation) about whether the agent will fulfill the normative expectation. We first need to impose a normative condition on the expectations in question. After all, we are interested in a normative relation between normative expectations and beliefs. Indeed, the debated relation between intentions and beliefs holds only for *rational* intentions.

I will assume that β is a rational expector in the sense that she holds others only to those expectations that she believes to be reasonable$_A$. (If a person holds another to expectations she takes to be unreasonable$_A$, i.e., to lie outside of the agent's power, it would be clearly irrational for her to believe that they will be fulfilled.) Assuming that β knows that she is a rational person and knows what this implies, she will believe that her expectation of α that α φ is reasonable$_A$, and that in turn will imply that she believes φing to be within α's power. So:

> If (rational) β expects of α that α φ then β believes that it is within α's power to φ.

Although the belief that it is not within α's power to φ makes it rational to believe that α will not φ, the belief that it is within α's power to φ is not yet tantamount to the belief that α will φ. For one, the agent may not be reliable in fulfilling the expectation. Mary may expect of Joe that he tidy his workspace but know that Joe is a messy person and so believe that he will not do so (not to her satisfaction at any rate). For another, the agent may have competing plans or be subject to competing expectations. A teacher may expect of his student that she turn in her homework on time. The student may also be expected by her coach to take part in a ping-pong tournament, which will make it impossible for her to fulfill the former expectation. She may prefer to go to the film festival instead. In either case, a teacher who knows about the competing plans cannot unqualifiedly believe that the student will do as expected. Yet another factor that will affect the judgment whether the agent will act as expected is the possibility and likelihood that some defeating condition will occur. Tricia may expect of Sam that he

meet her in the library at 5 p.m. Ceteris paribus she will believe that he will be there on time (and plan on it accordingly). When she finds out that there is a bus strike and that Sam will most probably have needed to take a bus, her belief that he will be there on time should be weakened.

There are a variety of considerations that will affect the expector's belief concerning whether or not the agent will act as expected. Let us try to systematize them.

Motivational Dominance

I will assume that the expector believes that the normative expectation to which she holds the agent is motivationally dominant. There are various ways of understanding this claim. For the purposes of this paper, the following should suffice. Let us suppose that β expects of α that α φ. Let us assume that β believes that α is disposed to respond to β's expectation by φing (provided that α knows about the expectation) except when α is subject to certain competing expectations E_1, ..., E_n (among them some of α's own expectations of himself). We can then say that β believes that the normative expectation to which she holds α is motivationally dominant iff β believes that α is not subject to the competing expectations E_1, ..., E_n. The assumption of motivational dominance is intended to bring the considerations closer to the case involving intentions. After all, intentions are considered to be "all out" judgments that do not compete with other motivating factors. Rather, they are the products of the agent weighing competing reasons. The difference is that while intentions are thought to be intrinsically motivationally dominant, normative expectations are not.

Implicit in this characterization of motivational dominance is the conviction that normative expectations of oneself and normative expectations of another agent are on a par insofar as both can motivate the agent to act. But this thought is controversial. Most philosophers of action subscribe to explanatory individualism, a position according to which an agent can only act on his own pro-attitudes. The thought that an agent acts because another person wants[23] him to so act must be reduced to the thought that the agent acts because of some desire of his own, which might be suitably directed toward the other person's desire. According to an alternative view, explanatory nonindividualism, an agent may sometimes act because another person wants him to so act without thereby acting on any of his own pro-attitudes.[24] Explanatory individualism focuses on the individual in abstraction from his social environment. The individual is pictured as a rational master of his circumstances consulting only the contents of his mind, choosing to act as he sees fit. By contrast, explanatory nonindividualism aims to include the social (in particular interpersonal) context in which the individual acts from the very

outset. The individual is pictured as sometimes acting on his own preferences and expectations, and other times finding himself in situations where he is overwhelmed by the social frames of interaction, to use Goffman's term,[25] and acts on other people's expectations of him.

My characterization of motivational dominance presupposes the position of explanatory nonindividualism. More precisely, we need to assume that the expector, who is assumed to believe that her expectation is motivationally dominant, is an explanatory nonindividualist. I have defended explanatory nonindividualism elsewhere.[26] At present, I will confine myself to a couple of thoughts that should help alleviate the impression that the view is absurd.[27]

One reason why explanatory nonindividualism may appear to be completely mysterious is that the expector's normative expectation of an agent is a state of the expector not a state of the agent. The expectation appears to be completely idle unless it is somehow mediated by the agent's states. A nonindividualist can agree, however, that the normative expectation must be mediated in order to be effective. His position leaves open the possibility that the agent must *believe* that she is subject to the expectation. The only thing he denies is that the expectation *must* be mediated by the agent's *pro-attitudes*.

Note, moreover, that the nonindividualist does not need to construe pro-attitudes of others as having an intrinsic motivating power. Most of our desires directed toward others are completely idle, e.g., my wanting you to read Gombrowicz. This is not to say that others' desires may not motivate us to action in the right kind of interpersonal setting. My wanting my students to read Gombrowicz is no longer as idle and even less so is a lieutenant's command directed at his soldiers. Furthermore, there are situations where it seems most natural for us to construe another person's pro-attitudes as actually overriding our own pro-attitudes. These are typically situations where one person has power over another person. They have been brought out clearly in an experimental setting by Milgram's obedience experiments.[28] The most natural way to think about them is that the agent acts on the experimenter's expectations of him but contrary to his own expectations and desires.[29]

Even so, there may appear to be reasons to construe all cases of action as involving some desire of the agent. Indeed, the remarkable flexibility of our intentional vocabulary can easily secure the feasibility of attributing some desire to the agent.[30] However, the individualist needs to argue not only that it is possible to attribute a desire to the agent in every case of an action, but also that the agent acted *because of* such desire. In other words, it is not enough to show that the agent had a

reason for acting, the individualist must in addition show that she acted *for* that reason.[31] One of the most powerful arguments intended to show that every action must be done on some desire of the agent is the argument from breakdown cases.[32] Consider a case where the agent allegedly acts on another person's want (a child goes shopping because his mother told him to). On reflection, we might be inclined to think that had the child not wanted to go shopping ceteris paribus he would not have gone, thus demonstrating that he must have wanted to go shopping after all. But the argument equivocates between *wanting not* to go shopping (having a con-attitude) and *not wanting* to go shopping (lacking a pro-attitude). The reason why we find the central thought (had the child not wanted to go shopping he would not have gone) intuitive is that we focus on cases where the child had a con-attitude toward going (dug in his heals against going). Indeed, intentional psychology tells us only what the agent would do given that he *has* the relevant intentional attitudes.[33] However, to demonstrate that the child must have had a desire to go shopping, the individualist would have to show that had the child lacked a pro-attitude (rather than possessed a con-attitude) toward going, he would not have gone. What is missing is a consideration of cases where the child has neither a pro-attitude nor a con-attitude toward going. The argument is thus inconclusive. It leaves conceptual space for the nonindividualist to think that at least in cases where the agent lacks his own pro-attitudes, he can be moved to action by others' pro-attitudes.

Indeed, this seems to be a very natural picture. Suppose a friend comes by and asks Stefan to go to the movies with her. He is neither particularly inclined to go, nor does he have anything against going. He is simply indifferent. It is not at all implausible to suppose that when he does go, it is because his friend wanted him to.[34] Suppose Anna walks briskly, fully involved in her thoughts. Someone asks her what time it is. She answers automatically, as it were, continuing to walk and think. Did she tell the time because she wanted to answer the person's question? It seems plausible to think that she was completely indifferent to it. She answered the question because the person asked it.

I do not claim to have thereby removed all the doubts concerning the feasibility of the idea that normative expectations that others hold us to can motivate us to act. A lot of questions remain unanswered. I only hope to have demonstrated that our expector should not be written off as irrational because of her adherence to explanatory nonindividualism.

Agent's Reliability

Consider the question of the agent's reliability in fulfilling the expectation. If the expector believes the agent to be

unreliable in accomplishing a certain task (of φing), it would be ceteris paribus irrational for her to believe that the agent will fulfill the expectation to φ. If a new assistant has never seen a computer before, it would be irrational to believe that he will write even a simple program (without further training). At the same time, it would be irrational for the expector to hold the agent to the expectation. To believe that the agent is unreliable in accomplishing a certain task is *eo ipso* to believe that it would not be within the agent's power to accomplish the task. As befits a rational expector, she should not hold the agent to the expectation.

(nr) If (rational) β holds α to an expectation that α φ and β takes the expectation to be motivationally dominant and β believes that α is not reliable in φing, then it would not be rational for β to believe that α will φ.

What if the agent is believed to be reliable in fulfilling the expectation? One might think that we should conclude that it would be rational to believe that a reliable agent will fulfill the normative expectation. That will be true but only ceteris paribus. Even reliable agents encounter obstacles in the path of the fulfillment of expectations. Even a reliable assistant with a lot of computer experience, in the best motivational situation, might not complete a program as expected because of power failure, because he has been kidnapped, or because of another kind of accident. These kinds of conditions are defeating conditions. They render the otherwise reasonable$_A$ expectation unreasonable$_A$.

For the sake of clarity, let us first abstract from the very existence of defeating conditions.[35] To say that α is reliable in responding to the expectation that α φ is to say at least that as long as the expectation that α φ is motivationally dominant, α will fulfill it unless some defeating condition intervenes. Since we are abstracting from the existence of defeating conditions, the fact that α is reliable in responding to expectations that α φ will imply that as long as the expectation that α φ is motivationally dominant, α *will* fulfill it. In other words:

(r) On the assumption that there are no defeating conditions, if (rational) β holds α to an expectation that α φ and β takes the expectation to be motivationally dominant and β believes that α is reliable in φing, then it would be rational for β to believe that α will φ.

The assumption abstracting from the existence of defeating condition reveals a pure relation between normative expectations and beliefs. We might also think of it as the default relation since in holding others to normative expectations we

frequently ignore the possibility of adverse circumstances (defeating conditions) occurring.

Defeating Conditions

How does the existence of defeating conditions affect our judgment as to whether the agent will fulfill the expectation or not? Let us distinguish three kinds of situations. In all of them, the rational expector holds the agent to the expectation that he φ, she believes him to be reliable in fulfilling expectations to φ, and she believes her expectation that he φ to be motivationally dominant. In the first situation, she believes that no defeating condition will prevent the reliable agent from completing the task. In the second situation, she believes that it is certain that one of the defeating conditions will occur and that it will prevent the reliable agent from completing the task. The third situation is in between the first two: The expector believes that it is probable but not certain that one of the defeating conditions will occur.

The first kind of situation resembles the pure situation (r). When the expector believes that no defeating condition will occur, it is rational for the expector to believe that the agent will φ.

(r_{ND}) If (rational) β holds α to an expectation that α φ and β takes the expectation to be motivationally dominant and β believes that α is reliable in φing and β believes that no defeating condition will occur, then it would be rational for β to believe that α will φ.

This belief cannot be held with absolute certainty but then no belief could. One may object that it would not be rational for the expector to have this belief since the agent can always not fulfill the expectation simply because he does not want to. We should remember two points, however. First, the question we are asking is only about what it is rational for the expector to believe. We are not asking whether the agent will actually act as expected. Second, we are assuming that the expector believes that her expectation motivationally dominates all other (if any) expectations and plans of the agent. Given this assumption, it is rational for her to believe that the agent, who, she believes, will not encounter any obstacles in his way, *will* act as expected.

In the second kind of situation, when the expector believes that one of the defeating conditions will occur, it is rational for her to believe that the agent will not φ. After all, she believes that the defeating condition will occur and so that it will prevent the agent from fulfilling the expectation.

While the first situation would support the analogue of the strong view, the second situation might appear to undermine it. After all, this is a situation where an expector does not believe that the agent will fulfill the expectation. One might argue,

however, that when the (rational) expector believes that the defeating condition will occur, she is committed to abandoning the expectation. Because she believes that one of the defeating conditions will occur, she also believes that her up-to-now (t) reasonable$_A$ expectation will become unreasonable$_A$ at a later time t', when the defeating condition occurs. As a rational expector (in the above sense of holding others only to expectations she considers to be reasonable$_A$), she believes that she will have to abandon the expectation at t'. But since she believes that the defeating condition *will* actually occur at t', and so that she will have to abandon the expectation (at t'), there seems to be little reason for her to hold the agent to the expectation in the first place. She should abandon it now at t.[36] (Note that the conclusion that she should abandon the expectation at t depends crucially on her believing beyond a shadow of a doubt that the defeating condition will occur at t'. If she were to believe that there is some probability that the defeating condition will not occur at t', ceteris paribus it would be rational for her not to abandon her expectation at t.) Of course, she might still hold the agent to a different expectation, e.g., that the agent try to fulfill the expectation or show himself willing to do so, just not the expectation whose reasonableness$_A$ she believes will be defeated. If so, then the second situation does not falsify the analogue of the strong view. Although it is not rational for the expector to believe that the agent will φ, it is also not rational for the expector to hold the agent to the expectation.[37]

(r_{CD}) If (rational) β holds α to an expectation that α φ and β takes the expectation to be motivationally dominant and β believes that α is reliable in φing and β believes that some defeating condition will occur, then it would not be rational for β to believe that α will φ (and it would not be rational for β to hold α to the expectation).

Much more common than the second kind of situation is the third type of situation, where the expector believes that it is probable (rather than certain) that one of the defeating conditions will occur. In such a case, the expector believes that it is probable that the agent will be prevented from completing the assigned task by the occurrence of a defeating condition. This means that it would be irrational for her to believe that the agent *will* φ. This belief stands in conflict with the belief that because of the defeating condition the agent *will* probably *not* φ. But there is another belief that it is rational for the expector to hold, namely, that it is probable that the agent will not φ *because of* the defeating condition, i.e., that the agent will *try to* φ but will probably be prevented from φing successfully

by the occurrence of the defeating condition.[38] Note, however, that unlike in the second case, the expector can still hold the agent to the expectation that he φ. Since she only believes that the expectation will probably be defeated, it is not as clearly rational as before for her to abandon the expectation now.

(r_{PD}) If (rational) β holds α to an expectation that α φ and β takes the expectation to be motivationally dominant and β believes that α is reliable in φing and β believes that it is probable that some defeating condition will occur, then it would not be rational for β to believe that α will φ (though it would be rational for β to hold α to the expectation).

The account presented embodies the central insight of the strong view in taking the relation between normative expectations and beliefs in the pure situation to be that of entailment. But it also finds space for the intuitions underlying the opposition to the strong view. I will close the section by considering three cases I mentioned earlier, which purport to illustrate the failure of the strong position, to see how the account applies to them.

Jane expects herself to attend a yoga class but is not convinced that she will actually be there as she has heard rumors that the class will be cancelled. In this scenario, Jane must give some credibility to the rumors for her not to be convinced that she will attend the class. In other words, she believes that it is probable that the class will be cancelled (which is a defeating condition to her attending the class), and hence believes that she might not actually attend the class despite the fact that she expects it of herself. In other words, the example falls into category (r_{PD}). We should not expect that the entailment relation will be manifested under such circumstances and, indeed, it is not.

Don expects himself to make ten carbon copies as he is writing on a paper but has doubts whether he is succeeding. We should distinguish two cases here. In one, Don believes himself to be quite reliable in making ten carbon copies when he sets out to do so. He has done it many times, he knows what paper to use, how much pressure to apply, how many times the carbon paper can be used, etc. In such a case, if Don has doubts whether he is making ten carbon copies, it is very likely due to the fact that he has doubts whether he applies enough pressure on the carbon paper (perhaps he is weak due to illness) or whether he has mixed in the wrong kind of paper (he has been absent-minded recently). In other words, when Don believes himself to be reliable in the activity, he will normally believe that he is succeeding (case [r_{ND}]) unless he has reasons to think that some defeating condition occurred (case [r_{CD}] or [r_{PD}]). But

it is also possible (this is the second case), that Don neither is nor believes himself to be reliable in fulfilling the expectation to make ten carbon copies. He has no knowledge of how to do it, has never done it, and the task seems formidable to him. In such a case, not only is it rational for him to believe that he will not succeed, it will be rational for him not to hold himself to the expectation in the first place (case [nr]). On either reading, the example does not approximate the pure situation. It is no wonder then that the entailment relation is not revealed.

The same is true in the third case of the policewoman who expects of herself that she shoot the runaway despite knowing that he has run simply too far for her to have a good chance of not missing him. As the case is described by McCann,[39] the policewoman believes that she will actually miss her target because the runaway is too far off. This is a case when the expector believes that the defeating condition *will* occur—in fact she believes that it has already occurred (case [r_{CD}]). It seems plausible to think that it would be irrational for the expector to hold the agent (in this case, herself) to the expectation that she believes to be unreasonable$_A$. In other words, in this case, it would be irrational for the policewoman to expect of herself that she shoot the runaway. She may still expect herself to shoot in the runaway's direction, to frighten the runaway, to make a point by not giving in, and so on. Once again, the case can be accommodated on the above account.

There is thus no straightforward answer to the question whether rational β who holds α to an expectation that α φ and who takes the expectation to be motivationally dominant, should also believe that α will φ as expected. The account sketched resembles the strong view on the relation between intention and belief in the pure situation (r) and the case where the expector believes that no defeating condition will occur (r_{ND}). It also remains compatible with the strong view when the expector believes the agent to be unreliable (nr) and when the expector believes that some defeating condition will occur (r_{CD}). In those situations, although it is not rational for the expector to believe that the agent will do as expected, it is not rational for the expector to hold the agent to the expectation. But the account also makes room for some intuitions underlying the disagreements with the strong view in particular in cases where the expector believes that it is probable that some defeating condition will occur (r_{PD}).

V. NORMATIVE EXPECTATIONS, INTENTIONS, AND BELIEFS

The account presented is a view on the relation between normative expectations and beliefs; it is not a view on the

relation between intentions and beliefs. But there is no reason to think that it will fail to apply to the normative expectations that the agent holds herself to. Since intentions involve normative expectations of oneself, one might expect that it will shed some light on the relation between intentions and beliefs. I want to close by considering just that.

The account is similar to the strong view in that it takes the entailment relation between normative expectations and beliefs about the future to be the pure or the default relation. It inherits some advantages of that position. First, normative expectations (just like intentions) play a prominent role in planning, though paradigmatically in the planning of interpersonal activities. If I am working as part of a team and am (reasonably$_A$) expected to finish my part of the team's project, it will be rational for all members of the team to count on my actually finishing it, which among other things means that it is rational for them to believe that I will finish it. The account allows one to preserve the thought that the lack of the connection would be a detriment to the possibility of coordinating such activities while finding room for a challenge presented by Alfred Mele.[40]

Mele argues that there need not be any connection between intentions and beliefs for the former to play a coordinating role in human activities. He takes as an example a particular basketball play that is to allow the team to score five points. All of the players have rehearsed the kind of play so they can coordinate well, but at the same time they all believe that their chances of succeeding are very slim indeed. Mele takes this to show that the coordination of activities does not require any connection between intentions and beliefs (or indeed between normative expectations and beliefs). In the above terms, the problem with Mele's generalization from the example is that the case falls into category (r_{PD}). The expectors believe that it is very probable that some defeating conditions (namely, interference by the other team, their own exhaustion, etc.) will occur and prevent them from carrying out the planned play. But this does not tell us anything about what happens in the pure situation (r) were the defeating conditions not considered a factor. Suppose that one of the players (X) expects another (Isaiah, in Mele's example) to throw a ball in such a way as to allow him to catch it. Suppose further that X believes that Isaiah is reliable in making such shots, that the expectation is motivationally dominant, and that no defeating condition will occur.[41] If in such a scenario, X believes that Isaiah will not throw the ball as expected, this would surely constitute a good reason for X not to expect himself to do his part in the play. In short, the problem with Mele's example is the same as the problem with the other cases that purport to illustrate that the strong view on the relation between intention and belief is false.

They pertain to cases where the pure relation between intention and belief is clouded by the need to take defeating conditions into account. The proposal outlined finds central room for the thought that coordination requires such a strong relation between normative expectations (in cases [r] and [r_{ND}]) while accommodating Mele's counterexample (in case [r_{PD}]), just as it finds central place for the thought that there exists an entailment relation between normative expectations and beliefs (in cases [r] and [r_{ND}]) while accommodating counterexamples (cases [nr], [r_{PD}] and [r_{CD}]).

The second advantage inherited by the view proposed is the power to explain the conversational implicature between the announcement of an intention (which involves a normative expectation of oneself) and a belief. It is plausible to suppose that in the absence of any explicit mention of some defeating condition, it is assumed that the speaker either believes that none will occur or does not take defeating conditions into consideration. If so then it is plausible to think that the pure relation will hold between the announced intention and belief. Note, however, that there is not a similar conversational implicature in play between the announcement of normative expectations and beliefs. We are not as befuddled when someone says, "I expect of him that he come on time but I don't think he will" as we would be were she to say "I intend to come on time but I don't think I will." The reason for this difference is that unlike intentions, normative expectations (whether of oneself or of others) are not considered to be intrinsically motivationally dominant.

The account differs from the strong view on the relation between intention and belief in that it takes the entailment relation between normative expectations and beliefs to hold in the pure situation where defeating conditions do not play a role. It therefore allows conceptual space for the dissenters from the strong view *as long as* the expector's reason for withholding the belief that the agent will φ is her belief that it is probable that some defeating condition will occur. In other words, we should not deny that there is an entailment relation between normative expectations and beliefs across the board, as it were. Rather, the denial is limited to some rather exceptional circumstances, namely, those where there are reasons to believe that some defeating condition might occur.

At the same time, the view sketched is subject neither to the objection that normative expectations lose their power—that they become more like idle wishes, nor to the objection that the content of normative expectations is weakened when the content of the beliefs is weakened. Consider the first point. What is distinctive about the attitude of expecting something whether of oneself or of another is that one is (appropriately) disposed to sanction the fulfillment and frustration of the

expectation. This is not the case when one merely wishes or hopes some state of affairs to obtain. In such cases, one is normally not prepared to impose any sanctions—indeed, such an imposition would be inappropriate. The difference between my hoping to get into medical school and my expecting of myself that I do so is that in the latter but not in the former case it would be ceteris paribus appropriate for me to think well of myself, say, if I do get in and badly if I do not. The propriety of one's disposition to sanction the fulfillment or frustration of an expectation does not in any way depend on one's beliefs about whether the expectation will be fulfilled or frustrated. So there is no danger that the belief that an expectation will not be fulfilled will turn the expectation into an idle wish.

Let us turn to the second point. The defenders of the intention-entails-belief thesis argue that in cases where the agent does not believe that he will φ but believes that he will only try to φ, he cannot be properly said to intend to φ but only to intend to try to φ. In such a case, the entailment relation between intention and belief will still hold. The weakness of the content of the belief will be inherited from the weaker content of the intention.[42] Let us see what happens to the content of normative expectations.

A natural way of understanding the content of normative expectations is in terms of what fulfills and frustrates them, which in turn can be understood in terms of the conditions under which it would be appropriate to impose positive and negative sanctions. Consider first a situation where the expector believes that no defeating condition will occur, and consequently believes that the agent will fulfill the expectation. In such a case, the content of her expectation is that the agent φ. She is (appropriately) disposed to sanction positively the agent's fulfillment of the expectation as she is (appropriately) disposed to sanction negatively his frustration of the expectation. Of course, if one of the defeating conditions were to interfere, it would be appropriate for her to withhold the sanctioning of the frustration or fulfillment of the expectation.

Let us then turn to the situation where the expector believes that it is very probable that one of the defeating conditions will occur and so believes that the agent will probably not fulfill the expectation. Can she expect of the agent that he φ? At first glance, one might suspect that the expectation must be somewhat weaker, after all it is very probable that the expectation will not be fulfilled. But the appearance is deceiving. The expector's dispositions to sanction should not change at all. She will be (appropriately) disposed to positively and negatively sanction the agent in the very same circumstances (of fulfillment and frustration of the expectation, respectively), as she will be prepared to withhold her sanctions in case one of the defeating conditions occurs. The only thing that

changes is that it is more likely than it was before that she will have to withhold her sanction because of the occurrence of the defeating condition. In such a case, although the expector cannot rationally believe that *the agent will* φ, she can rationally expect *of* the agent *that he* φ. Indeed, some of our intuitions support this claim. Suppose that a teacher expects of a student that he finish a project he started by Wednesday but that, at the same time, she believes that it is very likely that he will not be able to do so because it is very likely that a power outage will prevent him from completing it. When the student nonetheless does finish the project, it is natural to say that he fulfilled the teacher's expectation that he finish it not just her expectation that he try to finish it.

While I have proposed a view on the relation between normative expectations (including normative expectations of oneself) and beliefs, I have not advocated a position on the relationship between intentions and beliefs. I want to stress two points. On one hand, it is important to appreciate that the debate between those who hold the strong view and the dissenters is motivated theoretically. The position one holds on the relationship between intention and belief will depend on the theory of intention one advocates. I do not have such a theory. At the same time, I am disinclined to speculate on the relationship in part because the only full-blown theories of intention[43] (as opposed to reasoned conjectures) do not give a univocal answer to the question. On the other hand, however, the account of the relation between normative and predictive expectations sketched allows one to find a place for many of the threads present in the debate on the relationship between intention and belief. The thoughts underlying the strong view are given central place in being taken to hold in the default relation between normative expectations and beliefs. At the same time, we can take seriously the challenges to the strong view as long as they are tied to beliefs about defeating conditions. This might lead one to risk the conjecture that what connection there is between intentions and predictive expectations is inherited by intentions from the motivationally dominant normative expectations they involve. I do not pretend to have proven this conjecture but only to have made it worth entertaining.[44]

NOTES

[1] G. E. M. Anscombe, *Intention,* 2nd ed. (Ithaca: Cornell University Press, 1963), 1.

[2] H. P. Grice, "Intention and Uncertainty," *Proceedings of the British Academy* 57 (1971): 263–279.

[3] This view has been proposed by Davidson in "Intending," in *Essays on Actions and Events* (Oxford: Clarendon Press, 1980), 83–102. There, he also seems to be denying that there is any stronger relation

between intention and belief, but he appears to retract this position in his "Reply to David Pears," in *Essays on Davidson,* ed. B. Vermazen and M. B. Hintikka (Oxford: Clarendon, 1985), 211–215.

[4] David Pears, "Intention and Belief," in *Essays on Davidson*, ed. B. Vermazen and M. B. Hintikka, 53–88.

[5] This position is further defended by Gilbert Harman, *Change in View. Principles of Reasoning* (Cambridge, Mass.: The MIT Press, 1986) and J. David Velleman, *Practical Reflection* (Princeton, N.J.: Princeton University Press, 1989).

[6] This point is challenged by Mele in "Intention, Belief, and Intentional Action," *American Philosophical Quarterly* 26 (1989): 19–30. I consider Mele's challenge in the final section of the paper.

[7] See, e.g., Davidson, "Intending"; Stuart Hampshire, *Thought and Action* (Notre Dame, Ind.: University of Notre Dame Press, 1982); Kirk Ludwig, "Impossible Doings," *Philosophical Studies* 65 (1992): 257–281; Hugh J. McCann, "Rationality and the Range of Intention," in *Studies in the Philosophy of Mind,* ed. P. A. French, T. E. Uehling, Jr. and H. K. Wettstein (Minneapolis: University of Minnesota Press, 1986), 191–211.

[8] See Harman, *Change in View*; J. D. Velleman, *Practical Reflection*, 115 ff. This position should not be confused with the view according to which an intention to φ implies only that one believes that one will try to φ. This view has been advanced by Hampshire (*Thought and Action*) and is criticized by Audi, see his "Intending," in *Action, Intention, and Reason* (Ithaca: Cornell University Press, 1993), 56–73.

[9] This point is challenged by Velleman, *Practical Reflection*, 115–116).

[10] Both McCann and Ludwig hold that whenever one φs intentionally, one intends to φ. This so-called simple view is controversial, however. Its challengers include: Michael E. Bratman, *Intention, Plans, and Practical Reason* (Cambridge, Mass.: Harvard University Press, 1987); Gilbert Harman, "Practical Reasoning," *The Review of Metaphysics* 29 (1976): 431–463 and *Change in View*; Alfred R. Mele, "She Intends to Try," *Philosophical Studies* 55 (1989): 101–106; Myles Brand, "Intention and Intentional Action," in *Contemporary Action Theory,* ed. G. Holmström-Hintikka and R. Tuomela (Dordrecht: Kluwer Academic Publishers, 1997), 197–217.

[11] The notion of normative expectation has been most important in the literature on omissions, though some authors do (e.g., Steven Lee, "Omissions," *The Southern Journal of Philosophy* 16 (1978): 339–354) and some do not (e.g., Patricia G. Smith, "The Concept of Allowing," *The Southern Journal of Philosophy* 22 [1984]): 223–240 and "Contemplating Failure: The Importance of Unconscious Omission," *Philosophical Studies* 59 [1990]: 159–176) invoke the distinction between normative and predictive expectations. Patricia Greenspan has used the notion of reasonable normative expectations to define the concept of freedom; see "Behavior Control and Freedom of Action," *Philosophical Review* 87 (1978): 225–240 and "Unfreedom and Responsibility," in *Responsibility, Character, and the Emotions,* ed. F. Schoeman (Cambridge: Cambridge University Press, 1987), 63–80. R. J. Wallace appeals to the notion of normative expectations in giving a compatibilist theory of moral responsibility; see his *Responsibility and the Moral Sentiments* (Cambridge, Mass.: Harvard University Press, 1994).

[12] Greenspan, "Unfreedom and Responsibility," 72.

[13] This distinction is not crisp because, as Wallace himself recognizes, predictive expectations are also often associated with various kinds of emotions (*Responsibility and the Moral Sentiments*, 20–21.)

[14] Anscombe, *Intention*, 56–57.

[15] Peter Strawson, "Freedom and Resentment," in *Perspectives on Moral Responsibility,* ed. J. M. Fischer and M. Ravizza (Ithaca: Cornell University Press, 1993), 45–66; R. J. Wallace, *Responsibility and the Moral Sentiment.*

[16] See e.g., M. E. Bratman, *Intention, Plans, and Practical Reason*; Robert Brandom, *Making It Explicit* (Cambridge: Harvard University Press, 1994).

[17] One notable exception is Vermazen ("Objects of Intention," *Philosophical Studies* 71 [1993]: 223–265), who argues that the objects of intentions do not need to be actions at all. However, Vermazen still believes that whenever one intends something that is not one's action, one must intend to do something, i.e., one must have an intention whose object is one's action.

[18] Such an account is developed in my *Social Anatomy of Action. Toward a Responsibility-Based Conception of Agency* (Ph.D. Dissertation: University of Pittsburgh, 1997).

[19] There is one exception that I will mention briefly and set aside. Suppose that it is not within the agent's power to φ because of the presence of a certain defeating condition, but that it was within the agent's power to prevent that condition from occurring. For example, the driver's drunken state renders it outside of her power to drive safely but it was within her power to prevent herself from reaching such a state. In such a case, it is arguable that it is reasonable$_A$ to expect of her that she drive safely despite the fact that she is drunk. I shall ignore cases of this sort in the discussion below.

[20] This is a claim about the concept of reasonableness$_A$ as I intend it to be used. I do not deny that there are other, potentially useful, concepts that differ from reasonableness$_A$ in this respect.

[21] Note that this concept of reasonableness applies to beliefs not to normative expectations. I am relying entirely on the reader's intuitions concerning what is and what is not reasonable to believe. In particular, I am not offering an explicit view concerning the relation between reasonableness of beliefs and reasonableness$_A$ of normative expectations.

[22] Note that this is not in general equally plausible for reasonableness$_N$.

[23] I am construing the terms "want" and "desire" very broadly. They are intended to have the same technical meaning as the somewhat awkward term "pro-attitude," which can be captured in terms of the metaphor of direction of fit. See Michael Smith's "The Humean Theory of Motivation," *Mind* 96 (1987): 36–61 and *The Moral Problem* (Oxford: Blackwell, 1994); for the application to the interpersonal case, see my "Collectivism on the Horizon: A Challenge to Pettit's Critique of Collectivism," *Australasian Journal of Philosophy* 76 (1998): 165–181.

[24] Although explanatory individualism is almost universally held, there are important exceptions. See, e.g., Annette Baier, *Postures of the Mind. Essays on Mind and Morals* (Minneapolis: University of

Minnesota Press, 1985); Noel Fleming, "Autonomy of the Will," *Mind* 90 (1981): 201–223; Leszek Nowak, "Man and People," *Social Theory and Practice* 14 (1987): 1–17 and *Power and Civil Society. Toward a Dynamic Theory of Real Socialism* (New York: Greenwood Press, 1991); Georg Henrik von Wright, "Explanation and Understanding of Action," in *Practical Reason* (Ithaca: Cornell University Press, 1983), 53–66.

[25] Erving Goffman, *Frame Analysis. An Essay on the Organization of Experience* (Boston: Northeastern University Press, 1986).

[26] Paprzycka, "Collectivism on the Horizon."

[27] I should note here that both explanatory nonindividualism and individualism are positions about intentional psychology as long as it is understood on a personal rather than a subpersonal level; for the distinction see Daniel C. Dennett, "Three Kinds of Intentional Psychology," in *The Intentional Stance* (Cambridge, Mass.: Bradford Books, 1987), 43–82.

[28] Stanley Milgram, *Obedience to Authority* (New York: Harper & Row, 1969).

[29] Gibbard describes these cases as "social akrasia," see his *Wise Choices, Apt Feelings. A Theory of Normative Judgment* (Cambridge: Harvard University Press, 1990). Milgram thinks that one needs to introduce the notion of an "agentic shift" (where the control over one agent's actions is transferred to another) to account for the phenomena, see *Obedience to Authority*. Indeed, there are evolutionary reasons why our tendency to conform to others' will should have developed. Boyd and Richerson argue that it may be beneficial for an individual to use tried out behavioral patterns rather than to risk that his own solutions will be selected away; see their *Culture and the Evolutionary Process* (Chicago: Chicago University Press, 1985).

[30] I offer a nonindividualist explanation of this rather remarkable feature of intentional psychology in "False Consciousness of Intentional Psychology" (unpublished, available at http://orca.st.usm.edu/paprzycka/falint.html).

[31] The significance of this distinction for contemporary action theory is discussed in Davidson's "Actions, Reasons, and Causes," in *Essays on Actions and Events*, 3–19.

[32] For a more detailed consideration of the argument, see my "Collectivism on the Horizon."

[33] One could respond that intentional psychology as a whole tells us also what would happen (namely, nothing) when the agent lacks the relevant attitudes. However, since the response relies on the assumption that intentional psychology offers a complete picture of human agency it is question-begging against the nonindividualist.

[34] One may object here that even though Stefan does not have a desire to go to the movie, there is surely some desire on which he acts. And our intentional psychology provides us with plenty to choose from: the desire to spend time with the friend, the desire to do something beside homework, the desire to fulfill the friend's desire, and so on and so forth. The abundance of attitudes generated by intentional psychology obscures the rather extreme nature of the claim that is both central to the objection and the target of explanatory nonindividualism. Whence the confidence that Stefan *must* have had some desire that would not merely rationalize but *explain* why he

acted in the way he did? The nonindividualist is not opposed to the existence of a variety of possible situations where the just mentioned desires *do* play an explanatory role. She is merely concerned to argue for the *possibility* that Stefan does not act on any of them—for example, because he may be indifferent not only to going to the cinema but also to spending time with his friend, to doing something other than homework that night, or to fulfilling the friend's desire. Ultimately, one can only engage in the debate between nonindividualism and individualism at the level of arguments not at the level of examples. None of the examples I have given in the text are meant to be conclusive illustrations of the truth of nonindividualism. They can only aspire to spark some nonindividualist intuitions in the readers.

[35] Note that this assumption is tantamount to "bracketing" the very existence of defeating conditions altogether not merely to supposing that the expector believes that no defeating conditions will occur in a particular case. Given this assumption, the expector cannot have any beliefs concerning defeating conditions at all. For a methodological account of abstraction, see in particular Renata Zielińska, "Contribution to the Problem of Abstraction," in *Idealization II: Forms and Applications,* ed. J. Brzeziński, F. Coniglione, T. A. F. Kuipers and L. Nowak (Amsterdam: Rodopi, 1990), 9–22; Leszek Nowak, *The Structure of Idealization* (Dordrecht/Boston: Reidel, 1980) and "Abstracts Are Not Our Constructs. Mental Constructs Are Abstracts," in *Idealization I: General Problems,* ed. J. Brzeziński, F. Coniglione, T. A. F. Kuipers and L. Nowak (Amsterdam: Rodopi, 1990), 193–206.

[36] It is tacitly assumed here that the time when the defeating condition occurs, t', is prior to the time when the agent is expected to φ.

[37] Note that there is an important difference between it being rational for β to hold α to an expectation and it being reasonable$_A$ for β to hold α to the expectation. The latter concept is not epistemically responsive, while the former is. To say that it is rational for β to hold α to an expectation is to say that it is rational for β to believe that it is reasonable$_A$ for β to hold α to the expectation.

[38] It is important to stress that the content of the belief that the agent will try to φ (but will be probably prevented from φing successfully by the occurrence of the defeating condition) is much more specific than the content of the belief that the agent will try to φ. There are two ways of understanding the concept of trying. Suppose that the agent would have φed except that a defeating condition D occurred thus preventing her from φing successfully. In such a case, one might say that she only tried$^{(-D)}$ to φ (i.e., would have φed were it not for the occurrence of D). Such a concept of trying is tied to the occurrence of a particular defeating condition. A more general concept of trying would allow for all possible defeating conditions (with respect to φing) $D_1, \ldots, D_n$ to interfere with the agent's action. To say that the agent tried to φ in this sense is to say that she would have φed were it not for the occurrence of $D_1, \ldots,$ or D_n. Only the former concept is intended here. The latter concept is the focus of Hornsby's work, for example, where the agent can try to do something without even moving his body. See Jennifer Hornsby, *Actions* (London: Routledge & Kegan Paul, 1980).

[39] McCann, "Rationality and the Range of Intention"

[40] Mele, "Intention, Belief, and Intentional Action."

[41] Note that it may be the feature of the example itself that the players may always believe that there will be a defeating condition at work, e.g., the sheer difficulty of the task. In such a case, the above reasoning is only available counterfactually.

[42] This suggestion is made by Harman, *Change in View* and Velleman, *Practical Reflection.*

[43] Bratman, *Intention, Plans, and Practical Reason*; Harman, *Change in View.*

[44] I would like to thank Marcin Paprzycki and an anonymous referee for this journal for helpful comments on an earlier version of the paper.