Lucky Joint Action

Abstract

In this paper, I argue that joint action permits a certain degree of luck. The cases I have in mind exhibit the following structure: each participant believes that the intended ends of each robustly support the joint action. This belief turns out to be false. Due to lucky circumstances, the discordance in intention never becomes common knowledge. However, common knowledge of the relevant intentions would have undermined the joint action altogether. The analysis of such cases shows the extent to which common knowledge of the participants' intentions can be harmful to joint action. This extends a recent line of research that has questioned the necessity of common knowledge in joint action.

1 Introduction

Traditional accounts of joint action (e.g. Bratman 2013; Miller 2001; Tuomela 2005) comprise at least the following two necessary conditions for joint action to occur. First, an intention condition according to which joint action requires that each agent intend the same interdependent end; i.e. an end whose satisfaction requires that each party enact their respective part. Next, it is argued that these intentions be common knowledge.

Under the banner of "minimal joint action", several authors have recently challenged the common knowledge requirement. Most notably, Olle Blomberg (2016) has argued that, in certain cases, false beliefs about one's co-participants' intentions are compatible with joint action. Stephen Butterfill (2012) made room for the possibility of joint action in young children by simply not including a common knowledge condition in his definition of joint action. Similarly, Cordula Vesper et al. (2010) do not include common knowledge among the "building blocks" of minimal joint action. Lastly, Christopher Kutz (2000) has argued that participants merely need to be open to the disclosure of the relevant attitudes; a requirement which is, on Kutz's estimation, weaker than common knowledge.

Now, although common knowledge of the intentions of each is, according to these authors, not necessary for joint action, such knowledge is, in each case, argued to be compatible with joint action. Common knowledge is always good, yet not always required. In this paper, I will further explore common knowledge failure in joint action by pointing to cases in which common knowledge of the relevant intentions would act as an underminer for the joint action.

The principal idea is that joint actions permit a certain degree of luck. In the relevant cases, the participants believe that the relevant intentions of each robustly favor the joint activity which is undertaken successfully; this belief turns out to be false; the intentions of each do not, in fact, robustly favor the joint activity. However, had the actual intentions of each been common knowledge, the joint action would not have occurred in the first place. More specifically, I have two types of cases in mind; I will, first, analyze a set of cases in which common knowledge of some of one's co-participants' *subplans* would undermine joint action. In discussing these cases, I will rely on the following vignette:

> **Lucky Jog.** Sarah and Bob both intend that they go jogging. Sarah believes that Bob would continue the jog even if it rained. This is important for her! Her intention that they go jogging is conditional on her belief that he wouldn't bail if it rained. Her belief about Bob, however, who would bail if it rained, is false. Fortunately, sunny weather prevails and they complete a happy jog. As it happens, they got lucky.

Second, I will point to a set of cases in which common knowledge of one's co-participants' joint ends would undermine joint action. Consider the following vignette an illustration of this:

> **Forking Trip.** You and I are in Baltimore. I intend that we go to NYC. As a means, I also intend that we go to Philadelphia. You intend that we go to Ocean City. As a means, you intend that we go to Philadelphia. However, I only care about going to Philadelphia with

2

you insofar as it is a means to going to NYC. If I knew that you intended that we go to Ocean City, I would simply fly to NYC. You only care about going to Philadelphia with me insofar as it is a means to going to Ocean City. If you knew that I intended that we go to NYC, you would simply fly to Ocean City. I, however, (falsely) believe that you intend that we go to NYC; you falsely believe that I intend that we go to Ocean City. Upon arriving in Philadelphia, we discover our mismatched intentions. Nevertheless, we jointly went to Philadelphia.

This paper is a contribution to specifying minimally necessary conditions for joint action. Let me briefly explain why such a project is worthwhile. Most importantly, joint action is a pervasive feature of human sociality; that is, "joint action" refers to a basic way of acting together that is distinguished from mere parallel action (e.g. Kutz 2000, 5). To see the difference between both ways of acting, contrast, for instance, the case of two strangers walking down Fifth Avenue next to one another each intending not to run into the other, and the case of two friends walking together down Fifth Avenue. Joint action theory sets out to explicate the distinguishing features of such examples. As the walking-together example indicates, joint actions are not confined to long-term projects such as building a bridge together (Miller, 2001, 75), or going on a trip to New York City together (M. E. Bratman, 2013). A range of short-term activities such as lifting a table together (Sebanz et al., 2006), bouncing a block on a trampoline together (Warneken, et al. 2006), or jointly building a block tower (Bloomberg, 2013) serve as paradigmatic examples. These examples suggest that joint action is a ubiquitous and fundamental feature of human sociality. Concomitantly, philosophical interest derives (at least in part) from its pervasiveness. The literature chimes with this assessment. Margaret Gilbert describes joint action as the "social atom" that lies at the very "foundation of human social behavior" (Gilbert, 2003, 39). Michael Bratman thinks of joint action as the structure that grounds "social coordination and

planning" (M. E. Bratman, 1993, 99). Yet others argue that a cognitive adaption for joint action is the central and most basic evolutionary ingredient separating "hypersocial" (Tomasello et al., 2005) human sociality from the sociality of great apes.

In section two, I will discuss all relevant pieces of terminology and presuppositions. In section three, I will detail two types of assurance in joint action. In section four, I will show why the joint actions exemplified in **Forking Trip** and **Lucky Jog**, require that the beliefs of each about the respective other's intentions be false.

2 Terminology and presuppositions

In the introduction, I used the somewhat technical terms "ends", "intentions", "belief", and "common knowledge". Let me clarify and explicate these notions.

Ends and intentions. Joint actions involve some kind of motivating attitude on the part of each participant. This attitude is sometimes called a "conative" attitude. The exact name and nature of the relevant attitude varies across philosophical accounts. According to Seumas Miller, jointly acting agents are said to share the same "end" (Miller, 2001, 57). Raimo Tuomela speaks of "aim intentions" (Tuomela, 2005). Michael Bratman speaks of regular intentions (with the special content that "we" do something (M. E. Bratman, 2013, 60)). Finally, John Searle famously defends the idea of *sui generis* "we-intentions" (Searle, 1990). At times, the specific differences between those usages are identified to be merely terminological. According to Raimo Tuomela's assessment (2005, 353), for instance, his "aim-intentions" and Miller's "ends" really amount to the same thing. Other times, the differences between these attitudes are taken to be more substantive. Bratman, for instance, argues that only intentions, but not goals, can be "agglomerated" (i.e. can be combined) (M. E. Bratman, 2013, 22). Such

differences won't matter for our purposes. For the purposes of this paper, I will mostly use the term "intentions" and "ends" to refer to the relevant conative attitudes. I will say that several agents "intend that ___" and, when necessary, use the more specific formulation "have as an end that ___."

Contents. I will use Bratman-style formulations to specify the contents of the relevant conative attitude; these contents have the form "that we ψ." Hence, I will write "you and I have as an end that we go to NYC"[1] or "you and I intend that we go to NYC." Using Bratman-style "that we ψ" formulations in part just fixes a convention. However, one not entirely conventional merit is that such formulations succinctly capture the idea that there is interdependence between the participants' roles. For instance, my intention that we go to NYC cannot be satisfied unless you also go. This is because, on Bratman's account, you enacting your role is part of the satisfaction condition of my intention. The idea that joint actions involve some such interdependence is widely shared (e.g. Tuomela, 2005, 340; Miller, 2001, 56; Bratman 2013, 65). Bratman-style formulations capture this idea in a succinct and intuitive way. In what follows, I will call ends/intentions with a "that we" content *interdependent* end/intentions.

Means and Ends. Some intentions are mere means, some intentions are mere ends, and some are both. The relevant distinction between means and ends is that, other things equal, we only care about the means insofar as we care about the end. Hence, giving up the end would, likewise, rationalize giving up the means. The specification "mere" is important, because it is, of course, possible to intend something as a means *and* as an end. To see all this more clearly, reconsider first **Forking Trip**. In this example, our intention that we go to Philadelphia depends on our intention to go to NYC/Ocean City. We each only care about going to Philadelphia insofar as we care about going to NYC/Ocean City with the respective other. This dependence marks our intention that we go to Philadelphia as a

---

[1] Such formulations can be found in Miller (2006, 75).

mere means. This is different from the case in which going to Philadelphia with you has independent appeal; in which case this intention would also be an end.

Famously, Michael Bratman distinguishes between interdependent ends and the means – "sub-plans" (e.g. Bratman 2014, 55) as he calls them – that are realized in order to satisfy these ends. Ends and subplans are not merely distinct insofar as they are structured by the means-end relation. Subplans and ends are also said to be governed by distinct theoretical requirements. Joint action, according to Bratman, permits, for instance, that one leave the relevant sub-plans somewhat unspecified. All that is required for joint action is that one *intend* that the relevant subplans overlap (or "mesh" in Bratman's terminology) (see Bratman 2014, 53). Such openness, however, is not permitted for the intended ends. Bratman requires that all parties *have* the same interdependent end, not merely that they *intend* to have the same end.

Note that Bratman merely requires that the *interdependent* ends of each agent coincide. He does not require that *individual* ends of each participant coincide. Shared intention, according to Bratman, does not "require that the agents participate in the pursuit of the same goals. Perhaps you participate in our shared intention to paint the house because you do not like the present color, whereas I participate because I want to get rid of the mildew" (Bratman 2014, 29). It makes sense that joint action should not require our individual ends to coincide. To illustrate, suppose you and I intend that we go to NYC together. I, however, plan to go on to travel to Boston; Boston is my final destination. The fact that my ultimate end is Boston and not NYC does not undermine the possibility for joint action.

These distinctions are important, because, as announced above, one claim of this paper is that, in some cases of joint action (exemplified by **Forking Trip**), it is required that agents misrepresent their co-participant's *interdependent* ends; not their individual ends.

I should note that, throughout this paper, I will be concerned with two-party joint actions only.

Joint actions involving larger groups are somewhat special in that, in those cases, it may suffice that a proper subset of all participants entertain a certain end. In this paper, I will not be concerned with these intricacies.

Belief, knowledge, common knowledge, and belief of common belief. First, I will leave the notion of belief unanalyzed. Second, in the context of this paper, I only care about the difference between belief and knowledge insofar as knowledge is factive; knowledge entails *true* belief. Third, several agents commonly know a proposition *P* only if each knows that *P*, each knows that each knows that *P*, and so on *ad inf.* Furthermore, if an agent knows that a certain proposition *P* is common knowledge, then it is in fact common knowledge (Bonanno, 1996). All this is straightforward. Common belief, however, is a bit more entangled. A proposition is commonly believed, if each agent believes that *P*, believes that each believes that *P*, and so on *ad inf.*; but unlike the case of common *knowledge*, if an agent believes that a proposition is commonly believed, then this does *not* entail that this proposition is commonly believed (Bonanno, 1996). But apart from Bonanno's formal rendition, the idea is also independently intuitive. After all, an agent's belief that a proposition is commonly believed might just be false.

Note also that common knowledge (belief) merely *entails* such nested knowledge (belief); this is *not* to say that this nested structure *constitutes* common knowledge (belief). Famously, David Lewis (see Lewis, 1969; Cubitt & Sugden, 2003, 185) argued that common knowledge is *defined* in terms of symmetrically positioned reasoning on the part of various agents. Such symmetry is then said to *entail* nested knowledge. Whether Lewis was right need not concern us, because, by contraposition, a failure of such nested knowledge entails a failure of common knowledge.

Throughout this paper, I will be conveniently using examples in which the participants in a joint action have *actual* false beliefs about their co-participant's (and sometimes their own) intentions. This may seem like a departure from the philosophical literature. After all, philosophers in a broadly

Lewisian tradition couch things in terms of mere *dispositional*, or *potential* beliefs (for a review see Paternotte (2011)). In this paper, I'm concerned with the question of whether joint action requires that the contents of these beliefs (actual, dispositional, or potential) be *true*. Hence, I'm not concerned with the question of whether joint action requires common knowledge to be specified in terms of a particular epistemic modality. Therefore, we could phrase the examples presented here in terms of (say) dispositions to believe. To indicate what this would look like, consider the following rephrasing of **Lucky Jog**: Sarah and Bob both intend that they go jogging. Sarah has a dispositional belief that Bob would continue the jog even if it rained. Her dispositional belief about Bob, however, who would bail if it rained, is false. Fortunately, sunny weather prevails and they complete a happy jog. In this sense, I will argue that joint action is compatible with certain false dispositional or potential beliefs.

Let me address one related preliminary worry. Paternotte (2011) observes that common *knowledge*, taken at face value, requires an excessively high epistemic standard and is, therefore, not of much use for anything. In most cases, Paternotte argues, we don't have knowledge of other people's mental states, but, rather some type of justified belief short of being knowledge. To see this, Paternotte has us imagine that a sentence is publicly uttered in the presence of person *A* and person *B*. This sentence, many would hold, is now common knowledge between us. But this is false, Paternotte opines: "There is no way for A to be sure that B correctly heard the statement E" (see Paternotte 2011, 255). Absent such certainty, we could not have common *knowledge* of this event which is why Paternotte crafts a definition of common knowledge that employs lower standards of justification. Now, with this in mind, one might worry that these weakened notions of common knowledge already account for the possibility of error in our mutual belief attributions which is why the present paper might seem superfluous. I think this criticism is misguided. Paternotte is concerned with the fact that, in ordinary cases, a person's beliefs about what yet others believe *might have been false* (but are, in fact, true) which

is why they do not amount to knowledge. For Paternotte this is reason enough to lower the justificatory standards for common knowledge. In the present paper, I will defend the claim that, in many cases, joint action requires that some of the beliefs about others' intentions are *in fact false*. This, however, is not Paternotte's concern.

Rationality. Let me also add an important disclaimer: I will discuss joint action under the assumption of common knowledge of rationality. This is a strong assumption and needs a bit of justification. First, it is simply worth investigating whether fully rational agents could act jointly without common knowledge of some of the relevant intentions. Second, in the philosophical literature, common knowledge of all pertinent intentions is often added for the reason that rational agents would need it to engage in joint action and related cooperative activities (e.g. Lewis, 1969; Bratman, 1987; Rubinstein, 1989; Blomberg 2016). Denying that common knowledge is, in fact, required for jointly acting rational agents is a natural extension of the extant literature. The motivation behind this paper is to introduce the idea that common knowledge of the participants' intentions and subplans is sometimes required to fail. This claim has not yet been defended, neither for fully rational, nor for less than fully rational agents. Third, if fully rational agents do not need common knowledge of their intentions to successfully engage in joint action, then the idea that such common knowledge is required for joint action is *thoroughly* undermined, because a common knowledge requirement is often added precisely for the reason that rational agents would need such knowledge. That said, exploring necessary condition for joint action without assuming common knowledge of rationality is a worthwhile endeavor in its own right. Unfortunately, it'll need to be left for another time.

3   Two Types of Assurance in Joint Action

In this section, I will argue that rationally intending a joint action requires that each participant enjoys

some appropriate degree of assurance concerning the identity and robustness of the intentions of each. In the next section, I will then argue that, in certain cases of lucky joint action, the success of joint action depends on these beliefs being false (thereby not amounting to knowledge).

The first type of assurance (i.e. a belief that rationalizes action) concerns beliefs about the intended *ends* of each participant. The second type of assurance concerns beliefs about the intended *subplans* of each participant. I will discuss each type of assurance in turn.

Let's start with the following, rather uncontroversial, constraint on rational intending:

> **Rational Intention.** One ought not to intend what one believes to be impossible.[2]

In many cases, the absence of common knowledge of the relevant intentions will undermine **Rational Intention**.

> **Cards.** We each intend that we build a house of cards. We each know that we intend that we do so. I, however, falsely believe that you falsely believe that I intend that we play Blackjack.

Given my belief that you believe that I intend that we play Blackjack, I will reason that you will not act on your intention that we build a house of cards; at least, that is, if you are rational. After all, the satisfaction of your intention depends on me enacting my part, which you believe I won't do. Analogously, the satisfaction of my intention depends on you enacting your part, which I now don't think you will do. This puts me in a position to believe that, at least if we are rational, I won't be able to satisfy my intention, which, in turn, makes the pursuit of this intention irrational. The moral, then, is this: common knowledge of the relevant intentions assures each participant that their intention in favor of the joint action is not, in fact, unsatisfiable. Common knowledge failure is then taken to

---

[2] e.g. Davidson, 1978; Tuomela, 1974, 133; M. E. Bratman, 2013, 71; Grice 1957, 1989

undermine such assurance.

Above, I said that a failure of common knowledge will "in many" (but not in all) cases violate **Rational Intention**. Olle Blomberg (2016) has argued – I think convincingly – that, in some cases, a failure of common knowledge of these intentions need not violate **Rational Intention**. He gives the following example:

> **Hector and Celia.** Hector and Celia are about to build a block tower. Each intends that they build a block tower, and each intends to do their bit of this joint performance. [...] Hector falsely believes that Celia falsely believes that he intends to cover the top face of each of her blocks rather than to do his bit of their joint performance. (Blomberg, 2016, 318)

Blomberg argues that **Hector and Celia** is compatible with joint action, in part, because it is compatible with rationally intending the joint action (Blomberg, 2016, 319). After all, Hector will think that it doesn't really matter whether Celia thinks that he merely intends to stack wooden cubes on top of her cubes. Stacking blocks on top of her blocks is exactly what he would need to do to act in favor of her intention to build a block tower. Hence, he shouldn't expect her to abandon her intention to build a block tower. Analogously, Hector should likewise believe that his own intention that they build a block tower can be satisfied.

Importantly, even in Blomberg's example the participants in a joint action need some kind of assuring beliefs concerning the intentions of each. His argument is simply that these beliefs need not be a correct representation of the intentions of each. To see this, consider an amended scenario in which Hector falsely believes that Celia intends to play 'race car' with the blocks. Surely, in this case, it would be quite irrational for him to intend to build a block tower with Celia (for the familiar reasons provided in the discussion of **Cards**). Blomberg's point is that the true intentions of each don't need

to become common knowledge; the point is not that no assuring belief structure needs to be in place. Hence, even in this example, each participant needs to be assured that the other's intentions are, in some sense, appropriately related[3] to their own intention. Note also that some assuring belief concerning one's co-participant's ends is necessary to rationally carry out any subplan that is a mere means to one's interdependent end. To see this, consider the case in which Hector beliefs that Celia intends to play 'race car' with the blocks. Surely, in this case, he would also believe that his intention that they build a block tower couldn't be realized. And since mere means are dependent on the believed attainability of one's end, Hector would be irrational in retaining any of the intended subplans in favor of building a block tower.

Furthermore, in Blomberg's example, joint action is compatible with each knowing the respective other's interdependent end. Both intend that they build a block tower. Common knowledge of this intention would, surely, not undermine joint action.

I will now turn to a second type of assurance in joint action that is closely related to Bratman's idea that, in joint action, the agents' subplans need to exhibit some (yet to be specified) degree of "mesh" (M. E. Bratman, 2013, 54). Bratman gives an illustrative example: Suppose we each intend that we go to NYC.

> Your and my subplans can mesh even if they do not match. Perhaps your subplan specifies that we not go during rush hour, whereas mine leaves that issue open yet our sub-plans are co-realizable. Further, what is central to shared intention is that we intend that we proceed by way of sub-plans that mesh. This can be true even if, as we know, our sub-plans do not now

---

[3] Due to space constraints, I have to let the notion of "appropriately related intentions" unanalyzed.

mesh, so long as we each intend that in the end our activity proceed by way of a so-lution to this problem. Nor need we each be willing to accept just any specifi-cation of activities of each that would suffice for the intended end. (M. E. Bratman, 2013, 54).

One of the ideas expressed in this example is that while subplans don't need to be fully specified, we nevertheless need not be willing "to accept just any specification". I'd now like to elaborate on this idea and formulate a necessary condition concerning the required level of specification for joint action.

The rough idea is this: In joint actions, agents need to be assured not only that the intended action is possible (as argued above), but also that the intentions of each will persist through a range of counterfactual situations; i.e. that these intentions are sufficiently robust.[4] To see this, consider the following vignette:

> **Canceled Jog.** Jane and Trevor both intend that they go jogging. It is common knowledge between them that they so intend. Jane, however, believes that Trevor will bail at the slightest sign of rain. Jane, on the other hand, doesn't care about the rain. In fact, she greatly enjoys the rain while jogging. There are clear signs of rain today; so Jane, thinking of Trevor's disposition to bail, abandons her intention that they go jogging.

Jane's decision seems perfectly rational. Given that Trevor will likely bail, it's simply too risky for her to set out to go jogging with him. Hence, for Jane it is not enough to know that Trevor intends as she does. She needs further assurance that his intentions are somewhat robust, or, as Bratman would have it, that there is sufficient degree of mesh between their subplans. But how robust exactly?

---

[4] To be sure, Bratman points to "three forms of persistence interdependence" of intention (Bratman, 2013, 70ff) in joint action. Roughly these are: First, in many cases, my intention to do something with you often depends on you persistently wanting to do this thing with me. Second, we must both continue to believe that our success is realistically possible. Third, sometimes persistence of our intentions is grounded in moral obligation (e.g. due to a promise). Bratman's thoughts on intention persistence, as will become clear, are orthogonal to my discussion of believed mesh of intentions.

When formulating a constraint that captures Jane's reasoning in **Canceled Jog** we need to proceed with care. We certainly don't want to stipulate a fixed credential threshold to indicate a participant's certainty that the intentions of each will be satisfied. To see this, consider the following example:

> **Risky Jog.** I promise you one million dollars if you go on a five-mile jog with my friend Bob. You know that Bob will bail at the slightest sign of rain. It's looking rainy right now and you believe that, most likely, Bob will bail before the five miles are completed. The payoff, however, is alluring; so you intend that you and Bob go on a five-mile jog.

In **Risky Jog**, you have a low degree of belief that your intention will be satisfied. However, given the particular settings of the case, it seems perfectly reasonable for you to intend as you do. Hence, we need a constraint weak enough to capture the reasoning behind **Risky Jog**. I propose the following provisional definition: If two agents, A and B, intend that they J, then:

> **Moderate Robustness (provisional).** For each player, there exists a minimal set of circumstances, C_A (C_B), such that, if A (B) continues to intend that they J, then A (B) believes that it is commonly believed by all participants that they each intend that they J in all circumstances specified in C_A(C_B).

This is a mouth full. But an illustration will illuminate the mechanics of this principle. In **Canceled Jog**, Jane's set C_J contains rainy circumstances. She will continue to intend that they go jogging, only if she believes that she herself and Trevor would continue the jog in rainy circumstances. Hence, in **Canceled Jog**, she abandons her intention that they go jogging, because she does not believe that Trevor's intention would persist under rainy circumstances. In this case, it is false that each believes that all participants' intentions persist under rainy circumstances, because Jane herself doesn't believe that Trevor's intention would so persist.

14

C is a "minimal" set. This means that a participant would abandon her intention given any proper subset of *C*. Furthermore, an agent's set *C* is held rationally if it maximizes the player's utility. The thought is simple: Suppose Jane thinks that it will likely rain and the disutility she would experience from Trevor's bailing under rainy circumstances would be great. In this case, the expected utility associated with going on a jog with Trevor might just be too low to justify maintaining and acting on her intention.

Let's, next, be precise about the exact content of the belief of common belief specified in **Moderate Robustness (provisional)**. The added precision will lead to a slight amendment of this principle. Let's again use **Canceled Jog** as an example. Jane's intention that they go jogging is predicated on her belief that Trevor will jog even if it rains. Rainy circumstances are part of her set C_J. Does she have to believe that rainy circumstances are also part of Trevor's set C_T? No, because although Trevor might be happy continuing the jog under conditions of rain, he might be equally happy to seek shelter should Jane wish to do so. Rainy circumstances might not, as it were, be part of his *minimal* set C_T; i.e. his intention that they go jogging might not depend on him thinking that she will continue the jog in rainy circumstances. Now, intuitively, it only matters to Jane *that* Trevor won't bail if it rains; it does not matter to her whether running in the rain is part of his minimal set C_T. Instead, Jane needs to believe that Trevor believes that she shares all the elements specified in the set C_T indexed to *him*.

Furthermore, Jane needs to believe that Trevor believes that Jane believes that he shares the elements of C_J indexed to *herself*. To see this, consider the following line of reasoning: if Trevor thought that Jane believed that he didn't share the elements in her set C_J then he would be in a position to reason that she will abandon her intention that they go jogging. This, in turn, would make it rational for him to abandon his own intention that they go jogging. The upshot, then, is this: Trevor and Jane are required to entertain the nested type of beliefs typical of common belief; however, the contents of

these nested beliefs alternate from one level to the next in the way indicated above. Based on these considerations we should reformulate the above principle as follows: If two agents, *A* and *B*, intend that they *J*, then:

> **Moderate Robustness.** For each player, there exists a minimal set of circumstances, C_A (C_B), such that, if *A* (*B*) continues to intend that they *J*, then:
>
> *(i)* *A* (*B*) believes that they each intend that they *J* in all circumstances specified in C_A (C_B).
>
> *(ii)* *A* (*B*) believes that *B* (*A*) believes that they each intend that they *J* in all circumstances specified in C_B (C_A). (The crucial point is that the indexes "$_B$" and "$_A$" are reversed as compared to *(i)*. More generally, these indexes reverse at each level relative to the previous one.)
>
> *(iii)* *A* (*B*) believes that *B* (*A*) believes that *A* (*B*) believes that they each intend that they *J* in all circumstances specified in C_A (C_B) and so on *ad inf.*

This concludes my rendition of assurance in joint action. First, the participants need to be assured that the satisfaction of their intentions is possible. Second, the participants need to be assured that the relevant intentions favor the joint action in a robust fashion.

4 Intentions in the context of assurance

In this section, I will put to work what we have learned so far and argue that, given each participant's enjoyment of the kind of assurance set out above, acting jointly sometimes depends on misrepresenting the intentions of one's co-participant. To start seeing this, first re-consider **Lucky**

Jog:

> **Lucky Jog.** Sarah and Bob both intend that they go jogging. Sarah believes that Bob would continue the jog even if it rained. This is important for her! Her intention that they go jogging is conditional on her belief that he wouldn't bail if it rained. Her belief about Bob, however, who would bail if it rained, is false. Fortunately, sunny weather prevails and they complete a happy jog. As it happens, they got lucky.

Consider another example with larger scope:

> **Lucky Marriage.** Ian and Mia stand on the altar each vowing: "I promise to be true to you in good times and in bad, in sickness and in health." Both Ian and Mia intend this to be an honest expression of their conviction; they are, as it were, sure that they will be able to keep this promise. So they get married. By all measures, their marriage turns out to be truly wonderful. It lasts for 50 happy years until they both pass away. The "bad" times, in which, as they promised, they would stay together never came. There was no sickness, no temptation and the like. Had such bad times come around, Mia would not have stood by Ian. The same is true of Ian. However, had it been known to them that their marriage could not endure such bad times, they would have never gotten married in the first place. In fact, had it been known to them that they themselves could not keep this promise, they would not have gotten married either. Their belief that the respective other had robust intentions to stay together was crucial to them, despite the fact that, under pressure, it would not have held up. They both got lucky![5]

---

[5] It might be objected that a marriage somehow isn't a joint *action* (but rather a *project*). The example is included mainly to show how ubiquitous the falsity of such robustness related beliefs really is.

Sarah, as well as Ian and Mia have false beliefs about the robustness of the respective other's as well as their own subplans. Each misrepresents the robustness of their co-participant's as well as their own intentions. Nevertheless, the actions described in these vignettes are perfectly joint for the following reasons: First, it just seems intuitive that these are cases of joint action. Should we pay attention to such intuitions when defining joint action, or should we rely on independent theoretical considerations instead? I think we should, in fact, pay attention our considered intuitions for the following reasons. First, there is no agreement about the theoretical role of joint action. Butterfill (2012), for instance, argues that joint action plays a role in *enabling* mindreading in infants. Joint action should, therefore, not presuppose mindreading which is why Butterfill opts for a sparse definition of joint action. However, whether (and the extent to which) mindreading needs *enabling* is, in fact, contentious. Many have argued that mindreading is an innate capacity and is experimentally evident in infants as young as 6 months (for a review consult Carruthers (2013)). It seems to me that the theoretical role of joint action is insufficiently understood and it is, therefore, difficult to base definitions of joint action on such further theoretical embedding. Second, theory-driven definitions of joint action need to respect our considered intuitions on pains of not changing the subject. A case in point is an exchange between Butterfill (2012) and Blomberg (2015) who criticizes Butterfill's (2012) minimal account of joint action precisely because it overextends to cases that we would not intuitively label to be cases of joint action. This is not the point to discuss their exchange; however, we should note that theory-driven definitions of central philosophical concepts often run the risk of changing the subject by abandoning the initial target phenomenon. Explicating the precise role joint action plays in broader social scientific concerns is, of course, an important concern; but I will set this task aside for now.

That said, one uncontroversial theoretical desideratum of joint action theory is that is a basic and ubiquitous form of social interaction (see above). This assessment would have to be revised, if it were

denied that **Lucky Jog** and **Lucky Marriage** are joint actions. After all, there is nothing at all contrived about these cases. Jointly acting agents are regularly mistaken about the degree of robustness of their co-participants' intentions. In some cases, the problematic counterfactuals do materialize, which, in turn, leads to a breakdown of the joint action. However, in many other cases, in which the relevant counterfactuals do not materialize, this leads to the pursuit of perfectly normal joint actions.

A critic might object that joint action marks, in some sense, an ideal way of acting together; a way of acting together that is particularly *safe* and can't, therefore, come about through lucky circumstances. I think this criticism is misguided. Quite frankly, joint action does not mark an ideally safe way of acting together. Rather, – and the philosophical literature concurs – joint action is a basic and ubiquitous type of social action (see above). It is, as it were, the *actual base* of sociality, and not an ideal to live up to. There are in fact countless ways to make joint actions more safe and reliable (e.g. through third-party enforced contracts, explicit promising, mutually known expectations, or habitual action (e.g. Michael & Pacherie, 2014)) that go way beyond what is minimally required for joint action to occur.

Therefore, giving up on the idea that joint action is ubiquitous and socially basic is a high theoretical cost. After all, philosophical interest in joint action is, as I showed at the beginning of this paper, stoked, at least in part, by the conviction that joint action is a fundamental way of acting together.

The lesson from this discussion, then, is this: In the context of joint action, we should require that all participants have some appropriate degree of assurance concerning the range of subplans intended by each participant. We should, however, not require that these beliefs be true. Furthermore, in **Lucky Jog**, and **Lucky Marriage**, both pairs of agents engage in join action, yet knowledge of the range of subplans intended by each would readily undermine the joint action.

Let me now apply this insight to a second set of cases. In these cases, the participants' intended

interdependent ends are incompatible and knowledge of these ends would undermine the joint action. To see this more clearly, re-consider **Forking Trip.** In this case, going to Philadelphia requires that we each have false beliefs about the respective other's end. If we knew that our ends didn't match, we would abandon our trip altogether and our jointly going to Philadelphia would be undermined. Our jointly going to Philadelphia depends on our false beliefs about the intended interdependent ends of each. The intuition, however, is that we jointly go to Philadelphia. Given the above analysis of assurance, we're now in a position to defend this intuition on principled grounds.

To be very clear, the claim is *not* that **Forking Trip** is entirely devoid of matching interdependent intentions. Surely, in this example, both of us have matching interdependent subplans; namely that we go to Philadelphia; and this subplan is itself an interdependent intention with a "that we φ"-type content. Merely our intended interdependent *ends* are incompatible and not jointly realizable.

Now, why should we believe that **Forking Trip** presents a case of joint action? First, our interaction, while on the way to Philadelphia, is, in all important behavioral respects, indistinguishable from the case in which we both intend, as an end, that we go to Philadelphia. In both cases, we appropriately support one another, and our actions are equally interdependent.

Second, a critic might further object that the impossibility to co-realize both ends is problematic for a different reason; namely, that it renders the ends irrational. This objection is, again, misguided. According to the intuitive principle **Rational Intention**, rationally intending does not require that it be possible to satisfy the intention. Rather, mere *belief* that the satisfaction is possible is required. A stronger and, arguably, more controversial principle of rationality is required to render the relevant ends in **Forking Trip** irrational.

Third, a critic might further object that the relevant beliefs are not robust enough, because joint action requires, as Kutz (2000) puts it, that "no one would modify his or her plans in virtue of

disclosure." In **Forking Trip**, however, disclose of our ends would undermine the joint action. This point is no doubt true; but it overextends. In **Lucky Jog** and **Lucky Marriage**, the joint action would not survive the revelation of the actual robustness of each participant's intention. However, at least in those cases, this does not support the judgment that these actions are not truly joint actions. The critic would have to further argue that the actual ends are special in some sense, such that the joint action should survive the revelation of the ends, but not their subplans. This move, however, seems unmotivated.

Let me present one pressing objection in detail. Ben Laurence (2011) and Hans Bernhard Schmid (2016) have recently employed and expanded the Anscombian idea that acting intentionally requires, quite generally, some form of teleological rationalization (Laurence, 2011; Schmid, 2016). Thus, intentionally acting agents can answer the question "Why?" they act as they do. To illustrate, suppose you fill up a kettle with water in order to make tea. When asked "Why should you want to fill it up?'" (Laurence, 2011, 278) you might answer "Oh, because I want to make tea" (Laurence 2011, 278). Intentional action is said to depend on such rationalizations. Similarly, intentional *joint* agency is then, likewise, said to depend on the availability of such rationalizations. What is more, each participant's contribution to a joint action needs to be rationalized with reference to a *joint* (i.e. collective) action. To illustrate further, suppose a band of robbers are robbing a bank. The various robbers have different tasks. One robber's job is to crack open the safe. If their robbing the bank is indeed a joint activity then the robbers' individual actions (e.g. cracking open the safe) are rationalizable with regard to the joint end of robbing the bank. If asked, "Why is he cracking the safe?" the answer "because the band of robbers is knocking over Mellon bank" (Laurence 2011, 278) is appropriate. Laurence further requires that in joint actions such rationalizations should be *the same for each participant* (see Laurence, 2011, 282). Hence, the explanation "because the band of robbers is knocking over Mellon bank"

marks the teleological endpoint of their joint action that rationalizes each robber's contribution.

This line of reasoning generates an apparent objection against the claim presented in this paper. Presumably, in **Forking Trip**, each of us will rationalize their actions differently. When asked "why are you on this bus?" I will answer "because we are going to NYC", citing *my* interdependent end; you, on the other hand, will answer "because we are going to Ocean City", citing your interdependent end. But each of the rationalization that we provide are false. We're neither jointly going to NYC, nor are we jointly going to Ocean City. Our joint action merely extends as far as one of our subplans. Hence, the teleological rationalization each of us provides is false.

To answer this objection, we should first understand why we should accept this particular rationalization criterion in the first place. Schmid stresses that intentional agency requires that agents have available some rationalization for their actions. He explains that "it has to be apparent *to us* what it is we're doing, or else intentional action breaks down" (Schmid, 2016, 52). To see this more clearly, he provides the following example for such a breakdown in agency:

> **Fridge**. "Imagine that during a short break after some hours of intense work on a paper at your desk, still thinking about your paper, you find yourself in the kitchen, opening the fridge, not knowing what it is you're doing there. Perhaps your cluelessness does not run all the way down to your present bodily movements—you know perfectly well that you're opening the fridge—but you have no idea as to the question of why you're doing it. Were you about to get something from there, or put something back? You still feel utterly lost and rather stupid for a moment, and intentional action has broken down [...]" (Schmid, 2016, 52).

Individual agency, Schmid argues, requires that an agent *can answer* the relevant "Why?" question (not

just that the question is answer*able*) by citing her end. *Joint* agency, as Schmid and Laurence suspect, requires that participants can answer the "Why?" question by citing their joint end.

Now, I take it that, in **Forking Trip,** the mere *availability* requirement of such a rationalization is satisfied**.** After all, in said vignette both of us will rationalize their actions with regard to an intended joint action. If you are asked "Why are you on this bus?" you will answer "because we're going to Ocean City together"; likewise, I will answer "because we're going to NYC together." Hence, although false, we do have a rationalization at hand.

Should we add the further requirement that each of our answers to the "Why?" question be *true*? The answer, I think, is not obviously "yes." The simple reason is that agents can be mistaken about the particular explanations they invoke to rationalize their actions. Yet such confabulation does not obviously seem to render their actions unintentional (unlike the way the actions in **Fridge** seem unintentional and devoid of agency). Carruthers (2011, 342) and Wegner (2002) discuss experimental cases in which subjects carry out an instruction that was given to them under hypnosis and who "will often confabulate an explanation for their action citing some or other particular intention" (Carruthers, 2011, 342). For instance, subjects will follow the instruction "when I see the book on the table, I shall place it on the shelf." When later asked why they placed the book on the shelf, subjects confabulate an intention such as that they intended to tidy the room. Similarly, Schmid contends that our intentions are not always transparent to ourselves. He invokes the example of two friends quarreling all evening and only later realizing that the intention behind the quarrel was to break up the friendship (Schmid, 2016, 57).

Although we should agree that agents should be able to give a rationalization for their intentional actions, it seems at best controversial that intentional agency should depend on the *truth* of those avowed rationalizations. Now, the rationalizations that the agents in **Forking Trip** would provide are

false. When I answer the question "Why are you on this bus?" by saying "because we're going to NYC", I'm saying something false; however, as we've seen we shouldn't require such rationalizations to be true – neither in the context of individual agency, nor in the context of joint agency.

Summing up, we should agree that, in the context of joint action, there should be a description under which we act jointly. This description may well correspond to a subplan of ours. In **Forking Trip**, this description is "jointly going to Philadelphia". We can further concur with Schmid's claim that intentional agency requires that an agent be able to have available an answer to the relevant "Why?" question. After all, in **Fridge** intentional agency really seems to have broken down; hence intentional agency seems to depend on the availability of such rationalizations. We can also concede that, in the context of joint action, the answer to this "Why?" question should refer to a joint action (e.g. "I φ because *we* ψ"). However, we don't need to concede that the answers to such "Why?" questions necessarily be veridical. Bar such a veridicality requirement, the Anscombian approach to joint action does not undermine the jointness of **Forking Trip**. For these reasons, I conclude that there are no reasons to doubt that Forking trip is an any less genuine case of joint action than the ones put forward by the traditional accounts.

Lastly, a critic might object that, in **Forking Trip**, both of our ends could not, in fact, be co-realized; however, the objection continues, we should require that, in the context of joint action, the ends of each must be mutually co-realizable. Note that this objection is close to being a brute denial of the idea that **Forking Trip** presents a joint action; after all, one of the central claims of this paper is that, in joint actions, the intended ends don't need to be co-realizable. As such, the objection is much more a mere assertion of a clash of intuitions than it is a structured objection. A *mere* clash of intuition produces a stand-off. However, I think by now we have gathered extensive reason to suppose that this example does present a case of joint action. We saw that this example is compatible with an

intuitive way to spell out rational intending, it is a natural extension of our assumption that joint actions are socially basic, and the behaviors specified in this example are behaviorally indistinguishable from more mundane cases of joint action.

So far, we have looked at examples of lucky joint action in which incompatible intentions remain undisclosed to the participants. We may further ask whether joint action *depends* on the non-disclosures of the pertinent incompatible intentions. The answer, I think, is that it does not. But let's first explicate why one might think that it does. Reconsider, **Forking Trip**; suppose that, having arrived in Philadelphia, we find out that we are on our way to different cities. In hindsight we might be inclined to judge that, really, we didn't jointly go to Philadelphia. Concomitantly, we may want to say that our jointly going to Philadelphia depends on our not finding out that our intentions were incompatible all along. I think this assessment is wrong. To see this, we should, once again, see the parallel to **Lucky Jog**. Suppose we come home from a nice jog. As theorists, I have argued that, at this point we are entitled to judge that Sarah and Bob jointly went for a jog. If lucky joint action depended on non-disclosure of the incompatible intentions, this judgement would be inappropriate. Rather, we would only be entitled to judge that Sarah and Bob went on a jog together *provided that they won't find out about the discrepancy in intention later on.* This, I think, is unintuitive. After we come home from a nice jog, the question whether we went on a jog together is settled. Similarly, consider the case in which, having arrived in Philadelphia, we both get urgent calls from home and the both of us have to rush back to Washington D.C. At this point it seems to be settled – provided that the above analysis is right, that is – that we went to Philadelphia together. If lucky joint action were dependent on non-disclosure of the discrepant intentions, we would have to hold off on the judgement whether we went to Philadelphia together.

5   Conclusion

In this paper, I've argued that joint actions can be "lucky". In these cases, jointly acting agents are mistaken about their co-participant's subplans and interdependent ends. Furthermore, in the relevant cases, the joint action *depends* on these mistaken beliefs; common knowledge of the relevant intentions would undermine the possibility for joint action. Hence, the analysis of "lucky joint actions", as I called them, shows the extent to which common knowledge can be harmful to joint action.

References

Binmore, K., & Samuelson, L. (2001). Coordinated action in the electronic mail game. Games and Economic Behavior, 35(1-2), 6–30.

Blomberg, Olle (2015). Shared Goals and Development. *Philosophical Quarterly* 65 (258):94-101.

Blomberg, O. (2016). Common knowledge and reductionism about shared agency. *Australasian Journal of Philosophy*, *94*(2), 315–326.

Bonanno, G. (1996). On the logic of common belief. *Mathematical Logic Quarterly*, *42*(1), 305–311.

Bratman, Michael (1987). *Intention, Plans, and Practical Reason*. Center for the Study of Language and Information.

Bratman, Michael E. (1992). Shared cooperative activity. *Philosophical Review* 101 (2):327-341.

Bratman, M. E. (1993). Shared Intention. *Ethics*, *104*(1), 97–113.

Bratman, M. E. (2013). *Shared agency: A planning theory of acting together*. Oxford University Press.

Butterfill, S. (2012). Joint action and development. *The Philosophical Quarterly*, *62*(246), 23-47.

Carruthers, Peter (2011). *The Opacity of Mind: An Integrative Theory of Self-Knowledge*. Oxford University Press.

Carruthers, Peter (2013). Mindreading in Infancy. *Mind and Language*, 28 (2):141-172.

Cubitt, R. P., & Sugden, R. (2003). Common knowledge, salience and convention: A reconstruction of David Lewis' game theory. *Economics and Philosophy*, *19*(02), 175–210.

Davidson, D. (1978). Intending. In *Philosophy of history and action* (pp. 41–60). Springer.

Gilbert, M. (2003). The structure of the social atom: Joint commitment as the foundation of human social behavior. *Socializing metaphysics*, 39–64.

Grice, H. P. (1957). 'Meaning', Philosophical Review, 66: 377–88.

Grice, H. P. (1989). Studies in the Way of Words, Cambridge, Mass.: Harvard University Press.

Kneeland, T. (2012). Coordination under limited depth of reasoning. University of British Columbia Working Paper.

Kutz, Christopher. (2000). Acting Together. *Philosophy and Phenomenological Research* 61 (1): 1–31.

Laurence, Ben (2011). An Anscombian approach to collective action. In Anton Ford, Jennifer Hornsby & Frederick Stoutland (eds.), *Essays on Anscombe's Intention*. Harvard University Press.

Lederman, Harvey (2017). Uncommon Knowledge. *Mind*:fzw072.

Lewis, David (1969). *Convention: A Philosophical Study*. Harvard University Press.

Michael, J., & Pacherie, E. (2014). On commitments and other uncertainty reduction tools. *Journal of Social Ontology*, 1–34.

Miller, S. (2001). *Social action: A teleological account*. Cambridge University Press.

Schmid, H. B. (2016). On knowing what We're doing together: Groundless group self-knowledge and plural self-blindness. *The epistemic life of groups: essays in the epistemology of collectives. OUP, Oxford*, 51-74.

Searle, J. R. (1990). Collective Intentions and actions. *Intention in communication*, *401*, 401.

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: bodies and minds moving together. Trends in cognitive sciences, 10(2), 70-76.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and brain sciences*, *28*(05), 675–691.

Tuomela, Raimo (1974). *Human Action and its Explanation*. Institute of Philosophy, University of Helsinki.

Tuomela, R. (2005). We-intentions revisited. *Philosophical Studies*, *125*(3), 327–369.

Vesper, C., Butterfill, S., Knoblich, G., & Sebanz, N. (2010). A minimal architecture for joint action. *Neural Networks*, *23*(8-9), 998-1003.

Warneken, F., Chen, F., & Tomasello, M. (2006). Cooperative activities in young children and chimpanzees. *Child development*, *77*(3), 640-663.

Wegner, D. (2002). The Illusion of Conscious Will. MIT Press.