An Evidence-Based Critique of Intention Cognitivism

John McGuire

School of International Studies, Hanyang University,
Seoul, South Korea
mcguire@hanyang.ac.kr

Abstract

"Intention Cognitivism" (IC) refers to a family of theories concerning the relation between the concepts of intention and belief. While there are important differences between the various theories that belong to this family, each is committed to the idea that an agent intends to X only if they believe that they (probably) will X. In this article I argue against this core commitment of IC on the basis of recent experimental evidence concerning the ways in which ordinary speakers of English use the concepts of intention and belief. After summarizing some of this evidence, I consider a possible defense of IC, one based on the distinction between full versus partial intentions and beliefs. I report on a new study that was conducted specifically to evaluate this defense of IC, the results of which indicate that this distinction fails to protect IC against the compelling empirical evidence against it. Additionally, I consider, and argue against, two recent arguments for IC.

Key words: *intention, belief, intention cognitivism, partial intentions, experimental philosophy*

Received 15 February 2022; Revised 24 July 2022; Accepted 21 August 2022
Journal of Cognitive Science 23(3): 249-282 September 2022

1. Introduction

What exactly is the relation between the concepts of intention and belief? This simple question, which has sustained a lively debate among philosophers of action for more than half a century, has proven to be surprisingly difficult to answer. The following is a brief sample of some of the theories that have been defended in response to it:

I. *Identity theories:*
Intentions are identical to certain sorts of beliefs (Velleman 1989; Setiya 2007; Marušić & Schwenkler 2018).

II. *Compositionality theories:*
Intentions are composed of, and/or can be defined in terms of, certain combinations of beliefs and other mental states (Grice 1971; Audi 1973; Beardsley 1978; Davis 1984).

III. *Positive semantic theories:*
Intentions and beliefs are distinct mental states, but the concepts of intention and belief are semantically related such that an agent can intend to do something only if they believe they (probably) will do it (Harman 1986; Clark 2020).

IV. *Negative semantic theories:*
Intentions and beliefs are distinct mental states, but the concepts of intention and belief are semantically related such that an agent can intend to do something only if they do *not* believe that they (probably) will *not* do it (Mele 1992).

V. *Rational norm theories:*
Intentions and beliefs are metaphysically and semantically distinct, but both concepts are subject to certain defeasible norms of rationality such that it is normally irrational for an agent to hold intentions that are inconsistent with their beliefs

(Bratman 1984, 1987; McCann 1991; Holton 2009).

The foregoing list is not exhaustive: other notable views that do not fall neatly into any of the four categories on this list include those of Anscombe (1957), Davidson (1980), and Thompson (2008). Nevertheless, the list does give some sense of the diversity of views on offer and illustrates why it is unlikely that the debate over the relation between intentions and beliefs will yield a consensus among philosophers of action anytime soon.

However, if there is such a thing as a correct understanding of the relation between the concepts of intention and belief, which is an assumption shared by virtually all who write on this topic, then it really should be possible to make progress toward identifying it. This essay attempts to make such progress, not exactly by defending a particular theory, but rather by arguing against an entire family of theories. More specifically, in what follows I will argue against the first three types of theories on the foregoing list—identity, compositionality, and positive semantic theories. What unites these three types of theories is a commitment, which will be characterized more precisely below, to the idea that the concept of belief is somehow implicit in the concept of intention. The theories that are united by this commitment are often called "cognitive theories of intention." However, the term "cognitivism" as it is applied to the concept of intention, is in need of some clarification.

Paul (2009) introduced the terms "strong cognitivism" and "weak cognitivism" to distinguish identity from compositionality theories of intention. According to Paul (2009: 3), strong cognitivist theories are those that hold that an intention to X is a special kind of belief that one will X, whereas weak cognitivist theories hold that an intention to X is a "composite attitude" that involves the belief that one will X plus "some further, separable *conative* component." If every cognitivist theory is either "strong" or "weak," as Paul defines those terms, then every cognitive theory of intention is either an identity or a compositionality theory. However,

other researchers have used the term "weak cognitivism" in a broader sense to include, not only compositionality theories, but also positive semantic theories. For example, Clark (2020) explicitly rejects the idea that intentions are even partly beliefs and instead defends a version of "weak cognitivism" according to which an intention to X entails the belief that one will X. Similarly, Levy (2017: 223) embraces this broader understanding of cognitivism since he defines "Intention Cognitivism" (hereafter IC) as the view that "intending to [X] entails, or even consists in, believing that one will [X]." For the purposes of this essay, I will accept this broader definition of IC, which includes identity, compositionality, and positive semantic theories of intention. While it is an open question whether negative semantic theories of intention also belong to the IC family, I will not pursue this question in the present context.

The remainder of this essay is structured as follows. In Section 2, after identifying the core commitment of IC, I explain why empirical research concerning the ways in which ordinary speakers of a language use the concepts of belief and intention can shed light on the relation between those concepts. In Section 3, I go on to summarize the experimental studies that have recently been conducted on this issue and show how the results of these studies challenge the core commitment of IC. In Section 4, I consider a possible defense of IC against these experimental results, one that is based on a distinction between *full* versus *partial* intentions and beliefs. I report on a new experiment that was conducted to test whether this distinction does in fact protect IC from the experimental results summarized in Section 3 and show that it does not. Given the experimental evidence against IC, in Section 5 I raise the question of why many philosophers continue to be attracted to this theoretical perspective. I consider two recent arguments for IC—those of Marušić & Schwenkler (2018) and Clark (2020)—and identify significant flaws in both arguments. In the conclusion, I offer my own explanation of IC's ongoing appeal, much of which stems from the widely

accepted fact that intentions are subject to certain rational norms, including norms of consistency, which ensure that an agent's intentions are generally consistent with their beliefs about what they will do. However, IC goes astray, I argue, in misunderstanding the nature and significance of those rational norms. A further and related mistake, one that is made by virtually all those who have defended IC to date, is to approach the question of the relation between concepts of intention and belief from a purely theoretical perspective, as opposed to one that is informed by empirical research. The overall objective of this essay is to present and defend an evidence-based case for the view that IC misrepresents the relation between the concepts of intention and belief.

2. IC and experimental philosophy

Let us begin by stating more precisely the core claim to which all cognitive theories of intention are committed. While identity and compositionality theories of intention are united by a certain metaphysical thesis—that intentions *consist* (wholly or partially) of certain sorts of beliefs—semantic theories need not accept that thesis. Conversely, the core claim that positive semantic theories make—that intending to X *entails* believing that one will X—is a purely semantic claim and one to which identity and compositionality theories are also committed. Mele (1997: 17) describes this semantic claim as a "confidence condition" on the concept of intention. To distinguish it from other possible constraints on the concept of intention, some of which are discussed below, I number and characterize the key claim of positive semantic theories of intention as follows:

Confidence Condition 1 (CC1):

A intends to X only if (at the relevant time) A believes that they (probably) will X.

CC1 is a core commitment, not only of positive semantic theories, but also of identity and compositionality theories. If, as these latter theories maintain, an intention to X consists (wholly or partially) of the belief that one will X, it follows that any agent who has an intention to X necessarily satisfies the belief condition stated in CC1. Since CC1 is something to which all cognitive theories of intention are committed, albeit for different reasons, any examination of IC should begin with the question of whether CC1 is true.

CC1 specifies a belief that an agent allegedly must have if—and while—they intend to do something. For example, according to CC1, if Henry intends to vacation in France this summer then Henry must believe—as long as he has that intention—that he (probably) will vacation in France this summer. While CCI is sometimes stated without the qualifying term "probably," including that hedge protects it from obvious counterexamples. For example, Henry might intend to vacation in France while at the same time acknowledging that unexpected circumstances *could* arise that would prevent him from taking that vacation. He could become ill, lose his passport, or be unable to travel for any number of reasons. Even though he intends to vacation in France, Henry might believe, not that he *certainly* will do so, but rather that he *probably* will. Including the qualification "probably" in CC1 weakens the key claim of that thesis and thereby makes it more plausible or easier to defend.

Nevertheless, even with this qualification, possible counterexamples to CC1 are easy to find. Bratman (1987) distinguished between two distinct types of counterexamples: (a) cases in which an agent intends to X while being agnostic about whether they will *try* to X; and (b) cases in which an agent intends to X while being agnostic about whether they will *succeed* in X-ing. Bratman's now famous example of the first type of case concerns an agent who intends to stop at the library on his way home from work, while acknowledging that he might forget to do so, especially after getting on his

bicycle and going into auto-pilot mode. Bratman illustrates the second type of case with an agent who intends to carry out a dangerous rescue operation while having some reasonable doubts about whether he will succeed in completing the operation. Similarly, Mele (1992) challenges CC1 with an example that involves agnosticism toward both trying *and* succeeding. He describes a situation in which he gets up to answer a sudden knock on his office door. Although he surely has the intention to open the door as he moves toward it, he may have no conscious beliefs about opening it—neither a belief that he will succeed in opening it nor a belief that he will even try. In Mele's view at least, it is possible for a person in this situation to have an intention to X without having a corresponding belief that they (probably) will X.

In light of such examples, Bratman and Mele both reject CC1, the core commitment of IC. However, not everyone has been persuaded by these arguments. As with so many other alleged counterexamples to philosophical theses, one's response to them seems to be determined, at least in part, by one's antecedent views on the thesis in question. Those who are inclined to reject IC will regard examples of the sort Bratman and Mele describe as evidence in support of their views while those who are inclined to accept IC will search for reasons to dismiss such examples. For example, consider again the case of an agent who intends to answer a knock on their door while apparently being agnostic about whether they will try to open it or succeed in opening it. A proponent of IC could respond to this example by claiming that if the agent intends to open the door the agent must believe that they (probably) will open the door, but the beliefs may be *tacit* or held *unconsciously*. If so, then IC could still be true even if the hypothetical example Mele describes were possible, for IC does not state that agents must be consciously aware of the beliefs they have when they intend to do something.

In order to break out of the motivated reasoning that so easily affects

one's judgments concerning the alleged counterexamples to CC1, it is necessary to appeal to something other than one's own intuitions. This is precisely what experimental philosophy attempts to do. In examining systematically the ways in which ordinary people think about and use certain concepts, experimental philosophers believe that they can gain important insights into the meaning of those concepts. The justification for this methodological approach is that the meaning of many concepts, especially the non-technical concepts of folk psychology (e.g. belief, desire, and intention), is determined by, or at least reflected in, the ways in which ordinary people use those concepts. Experimental philosophy has the potential to shed light on the relation between the concepts of intention and belief by examining, under carefully designed experimental conditions, the ways in which ordinary speakers of a language think about and use those concepts.

To take an experimental approach to the question of whether CC1 is true one can recruit groups of subjects (i.e. ordinary language users), present them with scenarios like the ones described by Bratman (1987) or Mele (1992), and then determine the extent to which they agree or disagree that the agent in the given scenario (a) intends to X and (b) believes they (probably) will X. If CC1 is true, then regardless of what "X" is, the level of agreement with (b) should be roughly equal to (or greater than) the level of agreement with (a). If the results of properly conducted experiments indicate that the level of agreement with (b) is significantly *lower* than that for (a), this would suggest that ordinary speakers of the language do not use the concepts of intention and belief in accordance with CC1, which in turn would constitute evidence against CC1 and, hence, IC. However, as I show in the next section, it is precisely this sort of evidence that has been uncovered in the experimental research that has been conducted thus far on the relation between the concepts of intention and belief.

3. Experimental evidence against IC

In order to determine whether ordinary speakers of English use the concepts of intention and belief in accordance with CC1, McGuire (2020) carried out three experiments, each involving a separate group of subjects. In each experiment, subjects were asked to read a given scenario and then answer two questions—one concerning whether the agent in the scenario has a certain intention and the other concerning whether the agent has the corresponding belief. The scenario that was used in these three experiments reads as follows:

> Tom has been battling lung cancer for the past three years. Unfortunately, he is losing the battle. Yesterday Tom's doctors informed him that the disease is now untreatable and that he has only months to live. When pressed for clarification, the doctors explained that only 30 percent of patients in his condition survive more than six months while no patients in his condition survive more than a year.
>
> This devastating news could not have come at a worse time for Tom, as he and his wife are expecting their first child, who is due to be born in about six months. There is nothing more important for Tom right now than to witness the birth of his first and only child.
>
> Based on the information his doctors have given him, Tom estimates that he has only a 30 percent chance of still being alive when his child is born. With odds such as these, Tom understands that it is unlikely he will get to meet his child. Nevertheless, this unpleasant thought does not weaken Tom's desire or determination to be present for the birth of his child.

In one of the experiments that McGuire (2020) reports on, after reading the above scenario, 74 subjects were asked to indicate, on a 7-point scale from 1 ("Disagree") to 7 ("Agree"), the extent of their agreement or disagreement

with each of the following sentences:

(a) Tom intends to witness the birth of his child.

(b) Tom believes that he probably will witness the birth of his child.

The key finding of this experiment was that the level of subjects' agreement with (b) (M = 3.2) is significantly lower than the level of subjects' agreement with (a) (M = 6.5). Furthermore, the modes for (a) and (b) were 7 and 3, respectively, indicating that subjects tended to agree strongly with the ascription of intention but were neutral or mildly disagreed with the ascription of belief.

In a second experiment, the relevant intention and belief probes were worded differently to distinguish clearly between subjects' judgments about sentences that ascribe mental states to Tom and their judgments about Tom's mental states. After reading the scenario described above, 99 subjects were asked to answer following three questions:

(a) In your opinion, to what extent does Tom *hope* to witness the birth of his child?

(b) In your opinion, to what extent does Tom *intend* to witness the birth of his child?

(c) In your opinion, to what extent does Tom *believe* that he probably will witness the birth of his child?

For each of these questions, subjects had seven options from which to select (1. "Not at all"; 2. "Very weakly"; 3. "Weakly"; 4. "Moderately"; 5. "Strongly"; 6. "Very strongly"; 7. "Absolutely"). Question (a) was included in this experiment to ensure that subjects were not conflating the concepts of intention and hope. The main result of this second experiment is that the mean score for (c), the question about Tom's beliefs, (M = 4.0) was again significantly lower than the mean score for (b), the question about Tom's intention, (M = 5.6), and the modes for the responses to (b) and (c) were 7 and 4, respectively.

The key finding of the first experiment was replicated in the second and also consistent with the result of the third experiment (not summarized here). Overall, the three experiments show, each with slightly different questions, high levels of agreement among subjects with the idea that Tom *intends* to witness the birth of his child but moderate to low levels of agreement with the idea that Tom believes he (probably) will witness the birth of his child. These results conflict with CC1 and, hence, with IC. If CC1 were true, one would expect this fact to be reflected in the ways in which ordinary language speakers use the concepts of intention and belief. In particular, one would expect that speakers who judge that an agent intends to X would also judge that the agent believes they (probably) will X. However, this turns out not to be the case. Experimental evidence concerning how ordinary language users judge agents in scenarios such as the one described above indicate that they do not use the concepts of intention and belief in accordance with CC1.

While many philosophers have thought that there are logical or semantic relations between the concepts of intention and belief, not all of them have accepted CC1. Some have thought that that the confidence condition stated in CC1 is too strong and needs to be weakened. The following is a list of alternative constraints on the concept of intention that could be, and have been, proposed.

Confidence Condition 2 (CC2):

> A intends to X only if (at the relevant time) A does not believe that they (probably) will not X.

Confidence Condition 3 (CC3):

> A intends to X only if (at the relevant time) A does not believe that it is highly unlikely that they will X.

Confidence Condition 4 (CC4):

> A intends to X only if (at the relevant time) A does not believe that it is impossible to X.

These foregoing conditions are ordered in terms of decreasing levels of confidence in the agent. For example, CC2 requires more confidence (or less doubt) in the agent regarding whether they will perform the intended action than does CC3. Similarly, CC3 requires more confidence in the agent than does CC4. Since each of these three confidence conditions (CC2 – CC4) state weaker belief-constraints on the concept of intention than CC1 does, they too are relevant to the truth of IC. Consider, for example, the weakest constraint of all, CC4, which basically states that an agent who intends to X cannot (at the same time) believe that it is impossible to X. Suppose, for the sake of the discussion, that CC4 were false. If so, this would mean that an agent *can* intend to X while simultaneously believing that it is impossible to X and, hence, believing that they certainly will not X. If an agent can intend to X while being certain that they will not X, then surely an agent can, contrary to what CC1 states, intend to X while not believing that they (probably) will X. Indeed, since the belief that one certainly will not X is inconsistent with the belief that one (probably) will X, one would expect any rational agent who holds the first belief not to hold the second belief simultaneously. Similar remarks apply to CC2 and CC3. Evidence against any of these confidence conditions is in effect evidence against CC1 and, hence, IC. Experimental philosophers have carried out studies on each of these confidence conditions (CC2 – CC4) and have found that none of them accords with ordinary usage of the concepts of intention and belief. Here I will briefly summarize the experiments that Buckwalter et al. (2021) carried out to test CC4.

To determine whether ordinary speakers of English use the concepts of intention and belief in accordance with CC4, Buckwalter et al. (2021) ran three different experiments, the first of which employed five different scenarios, each presented in two contrasting modes ("possible" / "impossible"), for a total of 10 conditions in the experiment. For the purposes of this discussion, it is primarily the "impossible" mode of each

scenario that is relevant. The following is an example of the "impossible" mode for one of the scenarios that was used in this experiment.

> Arnold is a highly trained military operative. He has been captured by enemy forces and is about to be tortured. Arnold believes that as a matter of brain chemistry, it is completely impossible to withstand this kind of torture. After considering his oath to his country, Arnold says, "It's impossible but I will withstand the torture." (Buckwalter et al. 2021: 322)

After reading the scenario, the 50 subjects in this condition (as in all other conditions in the experiment) were asked to indicate on a 7-point scale—from 1. "Strongly disagree" to 7. "Strongly agree"—the extent to which they agree or disagree with several statements, including the following two:

(a) Arnold intends to withstand the torture.

(b) Arnold believes that it is impossible to withstand the torture.

Buckwalter et al. (2021) report that the mean scores for responses to (a) and (b)—averaged across the "impossible" mode of all five scenarios—were 5.3 and 5.5. The main finding of this experiment, at least for the purposes of this discussion, was that the mean intent attribution and mean belief attribution were both significantly above the midpoint. In other words, subjects assigned to the "impossible" mode of each of the five scenarios in this experiment interpreted the relevant agents such that they simultaneously intended to X and believed that it is impossible to X. This evidence suggests that ordinary language users do not use the concepts of intention and belief in accordance with CC4, which, for reasons discussed above, constitutes further evidence against CC1 and, hence, IC.

4. IC and partial intentions

The various experiments that have been conducted to test the idea that there is a confidence condition in the concept of intention all have one thing

in common: they all involve scenarios in which an agent seems determined (to one extent or another) to achieve some goal that they recognize they are unlikely (to one extent or another) to achieve. The agents in these various scenarios can be described as having determination but lacking confidence. The experiments based on these scenarios indicate that ordinary speakers of English interpret such agents—those with determination but little or no confidence—as having an intention to do something they do not believe they will do. As such, these scenarios provide compelling counterexamples to CC1 and, hence, IC.

In defending their version of strong cognitivism, Marušić & Schwenkler (2018) consider and respond to alleged counterexamples to CC1 that are similar to the ones used in the experiments described above. For example, Marušić & Schwenkler (2018) consider an example that Mele (1992) introduced of a basketball player who prepares to throw a ball into a basket from the foul line, knowing that he has only a 45 percent success rate when throwing from that spot. In this case too, one might interpret the agent as having an intention to sink the shot while lacking the belief that they (probably) will sink it. Marušić & Schwenkler (2018) reply that examples such as this one are not at all counterexamples to IC; rather, in their view, these case merely illustrate that intending, like believing, can be *partial*. The following passage is illustrative:

> Is it possible to intend to do something that one does not believe one will do? Our reply is that all the examples that are supposed to illustrate this possibility are cases where a person *partially* intends to do something that she *partially* believes she will do. (Marušić & Schwenkler 2018: 325)

Marušić & Schwenkler (2018) go on to describe the different senses in which, in their view at least, intentions and beliefs can be partial. First, intentions and beliefs can be weak (or held weakly) as opposed to strong (or held strongly). According to Marušić & Schwenkler, agents have weak

(and hence partial) intentions or beliefs when they have not completely made up their minds on the relevant issues. Second, intentions and beliefs can be partial in the sense of being conditional or incomplete. For example, Mary might believe that Henry will vacation in France this summer come hell or high water, but she more likely believes that Henry will vacation in France this summer as long as certain circumstances (e.g. Henry becoming seriously ill, war breaking out in France, etc.) do not arise that would prevent him from travelling or cause him to change his mind. In the latter case, according to Marušić & Schwenkler, Mary's belief is partial in the sense of being conditional.

Marušić & Schwenkler (2018) go on to apply this distinction between full versus partial intentions and beliefs to Mele's example of the basketball player who believes that there is a less-than-even chance of sinking a shot from the foul line. Marušić & Schwenkler suggest that not only is the agent's belief (that he will sink the shot) partial in this case, but so too is the agent's intention since it is "pre-conditional," as they describe it, on the "*internal* possibility that he might accidentally shoot the wrong way" (2018: 327). They conclude that since the agent's intentions and beliefs are both partial, this example fails to present a counterexample to their version of IC. Since the agent's intention in this case is partial, all that is required is a partial belief, and that is one that the agent presumably has if he thinks he has a 45 percent chance of sinking the shot.

Marušić & Schwenkler (2018) suggest that their "pre-conditional analysis" of intentions applies, not only to Mele's example of the basketball player, but to *all* cases in which "an agent's lack of confidence in what she will do arises from a concern that she may simply fail to execute her intention" (2018: 327). One dubious consequence of this view is that nearly all intentions a rational agent might have turn out to be partial, since any intention can be frustrated in any number of ways, and rational agents are typically aware of this. For example, I intend to cycle to work today, but

my intention could easily be frustrated by a flat tire, a stolen bicycle, or an unfortunate fall off my bike. If I am rational, I will acknowledge and accept these facts, but if I do, then according to Marušić & Schwenkler, my intention must be partial, even though it might seem to me to be as full or unconditional as any intention I ever have. Marušić & Schwenkler reply that this consequence of their view "may be surprising but is not implausible" (2018: 327). However, whether it is plausible or implausible is something that should be decided, not by Marušić & Schwenkler, who are motivated to embrace this pre-conditional analysis to defend their version of IC, but rather on the basis of evidence concerning how the concepts of intention and belief are used by ordinary language users.

At least one of the experiments that McGuire (2020) reports on is relevant to this question. Recall that in the second experiment summarized above in Section 3, subjects were presented with the scenario of Tom, the dying patient who wants to witness the birth of his child, and asked to indicate, on a 7-point scale, the extent to which he intends to witness the birth of his child. The answer options, once again, ranged from 1. "Not at all" to 7. "Absolutely." According to Marušić & Schwenkler's pre-conditional analysis of intention, Tom's intention in this case must be partial since his lack of confidence about witnessing the birth of his child arises from a concern that he may fail to execute his intention due to his deteriorating health. If this were the correct way of understanding Tom's intention, then one would expect that fact to be reflected in the responses experimental subjects give to the question about Tom's intention. In particular, one would expect that the most commonly selected option would be one of the five options on the Likert scale that represent *partial* intention (i.e. options 2 - 6). However, as was noted above, the most commonly selected option in that experiment—the mode—was 7, suggesting that subjects do not regard Tom's intention as a partial intention. Nevertheless, while this experiment is suggestive, it does not necessarily disconfirm Marušić & Schwenkler's pre-

conditional analysis of intention, for the answer options in that experimental set-up ("Weakly," "Moderately," "Strongly" etc.) reflect only one of the ways that intentions can be partial, according to Marušić & Schwenkler's analysis. In order to overcome this limitation, I conducted a new experiment to test their pre-conditional analysis of intention, the results of which I will now summarize.

In this experiment, which I call "Experiment 1", a total of 88 American adults were recruited through the online survey hosting company Survey Monkey. The scenario that was used in this experiment, which is a slightly modified version of one that McGuire (2020) introduced, reads as follows:

> A presidential motorcade is heading toward a convention center, where the president is scheduled to give a speech. Across the street from the convention center, on the top floor of a hotel, sits a sniper, holding a rifle, waiting for the motorcade to arrive. The sniper's plan is to shoot the president as soon as he steps out of the limousine. The sniper has gone over the details of this plan many times, and he knows how difficult it will be to pull it off. Visibility at this time of the evening is poor, and due to security concerns, the president will be moving quickly as exits the car. There will be very little time to line up a clear shot on the president. Given these unfavorable circumstances, the sniper estimates that he has only a 20 percent chance of hitting his target, but that thought does not deter him at all. He knows that he may never have a better opportunity to shoot the president than this one. As the motorcade comes to a stop in front of the convention center, the sniper peers through the scope on his rifle and adjusts his aim. A few seconds later, the president emerges from the limousine and the sniper immediately pulls the trigger.

After reading the scenario, subjects were asked to answer four questions, which were presented together on a single screen. The answer options to each question were ordered randomly, and subjects could modify their

answers to any of the questions at any time before clicking on a final submission tab. The first two questions, which were included for the purpose of screening out subjects who did not read or understand the scenario, were as follows:

1. When does the president arrive at the convention center?
   (a) Morning
   (b) Afternoon
   (c) Evening
2. Does the passage indicate how many people are in the limousine?
   (a) Yes
   (b) No

The final two questions, which were designed to collect data on how subjects interpret the intentions and beliefs of the sniper, were as follows:

3. In your opinion, which of the following best describes the sniper's intentions?
   (a) The sniper fully intends to shoot the president.
   (b) The sniper partially intends to shoot the president.
   (c) The sniper does not intend to shoot the president.
4. In your opinion, which of the following best describes the sniper's beliefs?
   (a) The sniper fully believes that he will shoot the president.
   (b) The sniper partially believes that he will shoot the president.
   (c) The sniper does not believe that he will shoot the president.

A total of 22 subjects answered either or both of the comprehension questions incorrectly; the responses of these subjects were excluded from the results presented here as it is doubtful whether or to what extent these subjects read or understood the scenario and questions. The responses to Questions 3 and 4 that were collected from the remaining 66 subjects are summarized in the 3x3 contingency table below (Table 1).

Table 1. Subjects' attributions of intention and belief to the agent in Experiment 1 (N = 66)

		Belief			
		Full	Partial	No	Total
Intention	Full	19	37	1	57
	Partial	1	2	1	4
	No	3	0	2	5
	Total	23	39	4	66

Note. "Full" = "The sniper fully [intends to / believes he will] shoot the president."
"Partial" = "The sniper partially [intends to / believes he will] shoot the president."
"No" = "The sniper does not [intend to / believe he will] shoot the president."

In response to Question 3, concerning the sniper intentions, 57 subjects (86.4%) attributed a full intention to shoot the president, 4 subjects (6.1%) attributed a partial intention, and 5 subjects (7.6%) judged that the sniper did not intend to shoot the president at all. In response to Question 4, concerning the sniper's beliefs, 23 subjects (34.8%) attributed a full belief that he would shoot the president, 39 (59.1%) attributed a partial belief, and 4 (6.1%) judged that the sniper did not believe that he would shoot the president. A McNemar-Bowker test was performed to determine whether the row and column marginal frequencies in the above contingency table are equal. The test results indicate that the frequencies are significantly different, X^2 (3, N = 66) = 36.1, $p < 0.001$, and a post-hoc analysis confirmed that this result is due to the different attributions for full versus partial intentions and beliefs. That is, whereas 37 (56.1% of) subjects interpreted the sniper as having a full intention and partial belief, only one subject interpreted the sniper as having a partial intention and full belief; this asymmetry was found to be statistically significant ($p < 0.001$). A significant asymmetry was also found when the data from all 88 subjects was analyzed (i.e. even when the responses of the subjects who failed the comprehension questions were included in the analysis), X^2 (3, N = 88) = 33.8, $p < 0.001$.

If Marušić & Schwenkler (2018)'s pre-conditional analysis of intention reflects some truth about the concept of intention—that agents only partially intend to X when they believe that they may fail to execute their intention to X—then the majority of subjects in this experiment should have interpreted the sniper as having only a partial intention to shoot the president. This is because the sniper's lack of confidence that he will shoot the president—just like the basketball player's lack of confidence that he will sink the shot—arises from a concern that he may fail to execute his intention. However, the majority of subjects in this experiment did not interpret the sniper's intention as partial. On the contrary, even though a majority of subjects (59.1%) judged that the sniper *partially believes* that he will shoot the president, the majority of subjects (86.4%) judged that the sniper *fully intends* to shoot the president. These results suggest that subjects do not attribute intentions in accordance with Marušić & Schwenkler's pre-conditional analysis of intention. Nor does the distinction between full versus partial intentions and beliefs protect their brand of IC—strong cognitivism—from counterexamples such as the one described in the sniper scenario. Although Experiment 1 enables subjects to distinguish clearly between full versus partial intentions and beliefs, this opportunity does nothing to align their attributions of intentions with their attributions of beliefs; subjects continue to attribute intentions and beliefs very differently and not in accordance with the core commitment of IC.

The foregoing should not be interpreted as suggesting that the distinction between full versus partial intentions is fictitious or meaningless. One can still distinguish between an agent who is fully committed to achieving some goal X and an agent who is inclined to X but not quite fully committed; one might say that the former agent fully intends to X whereas as the latter agent intends to X only in some partial or weakened sense. However, no matter how committed an agent is to achieving a goal, it does not follow that the agent *will* achieve it or that the agent must *believe* they will achieve

it. Two equally good tennis players may enter a match against each other, both fully intending to win, even though each realizes that one of them is going to lose and that it could be either one of them. It may be strategically beneficial for each player to believe they themselves will win, but neither the metaphysical structure of intentions nor the semantics of the concept of intention requires either one of them to hold such a belief. The mistake with Marušić & Schwenkler (2018)'s pre-conditional analysis of intention is its assumption that agents who lack full confidence that they will achieve some goal must also lack a full intention to achieve it. Empirical evidence concerning how people actually use the concepts of intention and belief indicates that this assumption is false.

5. Arguments for IC

Given the empirical evidence against IC, the question to ask at this point is why so many philosophers continue to be attracted to this theoretical perspective. In what follows, I examine the arguments put forward by two of the most recent advocates of IC: Clark (2020) and Marušić & Schwenkler (2018). These arguments are worth considering, not only because they shine light on some of the reasoning that motivates IC, but also because they reveal a common mistake that is made by virtually all IC theorists—that of thinking about the relation between concepts of intention and belief from a purely theoretical perspective as opposed to one that is informed by empirical research concerning the ways in which those concepts are actually used by the speakers of a language.

According to Clark (2020: 310), "the central argument for [intention] cognitivism is that intentional action [and future-directed intention] entails at least some awareness of what one is doing." In other words, according to Clark, an agent can intend to X (or X intentionally) only if the agent has some idea that they will X (or are X-ing). It is the former claim concerning

the state of intending, as opposed to the latter claim about intentional action, that is of concern in the present context. In support of this claim, Clark provides the following example:

> Suppose you are about to open the door and someone asks, 'Why are you going to let the cat out?' If you had no idea you were about to let the cat out – perhaps you did not even know there was a cat – then although it may be true that you were going to do that, you did not intend to do it. (Clark 2020: 310)

Following this observation, Clark (2020) immediately presents the following thesis concerning the concept of intention:

> T1: "If you have no idea that you will X, then although it may be true that you will X, you do not intend to X" (Clark 2020: 310).

In the remainder of the article, Clark goes on to explain and defend his preferred version of IC, which he calls "non-inferential weak cognitivism" (2020: 310). Thus, the central argument for IC, according to Clark, is really just an assertion of T1, presented as an analytic truth, with the foregoing cat scenario offered as an illustration of this truth.

In order to evaluate this argument one must first of all clarify what exactly T1 means. As a first step, T1 can be rephrased as follows:

> T1*: A intends to X only if A has some idea that they will X.

But what does it mean to say that "A has some idea that they will X"? An assertion of this sort evidently attributes to A either certain *expectations* concerning their own future actions or at least some *awareness* of certain possibilities concerning their own future actions. Consider, for instance, the example Clark introduces to defend T1. To say that A has some idea that in opening the door they will let out the cat is to say either that A expects to let out the cat by opening the door or that A thinks they might let out the cat by opening the door. These cognitive states of expectation or awareness are best expressed in terms of the more general concept of belief. Thus, the relatively obscure T1, which is logically equivalent to T1*, can be clarified

further in terms of T1**.

> T1**: A intends to X only if either (a) A believes that they will X or (b) A believes that they might X.

The consequent in T1** is a disjunction of two claims—(a) and (b). If (b) is subtracted from T1**, the remainder is just CC1, the core claim of IC. Conversely, adding (b) to the consequent in CC1 results in a weakened version of CC1, one that is less vulnerable to counterexamples and hence more likely to be true. However, precisely because T1** is a weakened version of CC1 it does not suffice to establish the truth of CC1, the core claim of IC. That is, even if T1** were true, it would not follow that IC is true.

However, a further problem with arguing for IC on the basis of T1** is that there is evidence that T1** is false. As was noted in Section 3, Buckwalter et al. (2021) present evidence indicating that, as far as ordinary speakers of English are concerned, agents can intend to X even when they believe that it is impossible to X. These experimental results suggest that agents can intend to X without believing that they will—or even that they *might*—X. In other words, if CC4 is false, then so too is T1**, and there is indeed evidence that CC4 is false, evidence which Clark (2020) and other IC theorists to date have failed to consider.

To say that there is evidence that T1** is false is not to say or suggest that an agent can intend to X in the absence of any beliefs at all. It is both possible and plausible that an agent who intends to X must have certain contextually specific beliefs. For example, if A intends to let out the cat, then A presumably believes that cats exist, that there is a cat in the vicinity, that cats are ambulatory, and so on. However, precisely the same could be said of desires and other pro-attitudes. The fact that agents must have *some* beliefs in order to have intentions (or other pro-attitudes) provides no justification for IC, which requires, not that an agent who intends to X has some beliefs, but rather that an agent who intends to X has the more

specific belief that they (probably) will X.

On Clark's behalf, one might object that experimental evidence concerning how ordinary speakers use the concepts of intention and belief does not, or perhaps cannot, refute a semantic hypothesis like T1. However, whether or not this sort of experimental evidence *refutes* T1, it is at least evidence against it, evidence that needs to be taken into account. The key question concerns what reasons there are for and against thinking that T1 (or T1**) is true. Notice that in attempting to justify IC on the basis of T1 Clark too appeals to intuitions about how one would use the concept of intention in a given context (e.g. opening a door that might result in a cat being let out). The difference between Clark's argument for IC and the foregoing rebuttal to it is not that one argument appeals to ordinary language use while the other does not, but rather that one of these appeals is supported by evidence concerning how speakers of a language actually use the concepts of intention and belief while the other is based only on one's own intuitions about how those concepts should be used.

Let us turn then to consider Marušić & Schwenkler's (2018) argument for IC, which has been described by one of its critics as "simple and elegant" (Hauthaler 2022: 1). The core claim that Marušić & Schwenkler (2018) seek to defend is that intentions are beliefs based on practical reasoning: "To intend to do something," they write, "is neither more nor less than to believe, on the basis of one's practical reasoning, that one will do it (2018: 309). In the following passage, which is worth quoting in full, they explain the considerations that lead them to embrace this view.

> The aim of strong cognitivism is to explain the relation between practical and theoretical reasoning—that is, to explain how reasoning about *what to do* relates to reasoning about *what will happen* in the future. This relation is difficult to understand if practical and theoretical reasoning are seen as answering different sorts of questions—if, say, the conclusion

> of practical reasoning is supposed to be a judgment about what it would be good to do, or what one ought to do, or what one has most reason to do. By contrast, we hold that these two forms of reasoning are related because practical reasoning is concerned with a factual question—the question of what one is *going to do*. This is why we say that practical reasoning and theoretical reasoning both issue in beliefs. (Marušić & Schwenkler 2018: 309-10)

Before evaluating the argument that Marušić & Schwenkler advance in this passage, I will provide some relevant context to clarify it.

The distinction between practical and theoretical reasoning is commonly understood as that between, on the one hand, reasoning about *what is to be done* and, on the other, reasoning about what is the case. However, as Kane (1998: 21) points out, the phrase "what is to be done" is ambiguous: "it can signify what I (or someone) 'should' or 'ought' to do; or it can signify what I 'will' (i.e., 'choose' or 'decide') to do." Kane goes on to conclude that practical reasoning can issue in two kinds of judgments—either normative judgments about what one ought to do or, alternatively, choices or decisions about what one will do. On this view, which is a very common one in the philosophy of action, there are in fact three types of reasoning: (a) theoretical reasoning that issues in beliefs about what is the case, (b) practical reasoning that issues in beliefs about what one ought to do, and (c) practical reasoning that concludes with the making of decisions and/or the formation of intentions. However, there seems to be something missing from this standard triad. In addition to having a rational belief that one *should X* (and perhaps forming an intention to X), it seems possible for an agent to have a rational belief that they *will X*. The traditional understanding of practical reasoning says nothing about this last sort of belief, or the reasoning upon which it might be based. Marušić & Schwenkler (2018) want to fill this lacuna in the traditional way of thinking

about practical reasoning. The key issue they are concerned with is how agents can have rational beliefs about what they *will* do—not just what they *should* do—when matters are up to them.

While the question that Marušić & Schwenkler (2018) address is a good one, the answer that they provide to it is not. In the paragraphs that follow, I explain how it is possible to understand how agents can have rational beliefs about what they will do, when matters are up to them, without supposing, as Marušić & Schwenkler do, that those beliefs must be identical to the intentions agents have formed on the basis of practical reasoning. Marušić & Schwenkler point out that an agent's beliefs about what they will do cannot be grounded simply in evidence about what is independently likely to happen; they correctly note that "if matters are up to the agent, then this sort of evidence will necessarily be inconclusive" (2018: 311). They then infer from this fact that an agent's belief that they will X must be based on reasoning from considerations about what is worthwhile or good (i.e. practical reasoning).

There is, however, a third option, one that Marušić & Schwenkler overlook, which is that an agent's beliefs about what they will do can be based on the *decisions* they have made and the *intentions* they have formed or acquired. An agent who forms an intention to X has a reason to believe that they will X, not because "intending to X" is equivalent to, or entails, "believing that one probably will X," but rather because intentions are *conduct-controlling* mental states. Intentions control conduct in the sense that they initiate, motivationally sustain, and guide actions (Bratman 1987; McCann 1991; Mele 1992). This is not to say that any agent with an intention to X is guaranteed to X but rather that an agent who has an intention to X is at least likely to try to X, as long as they maintain that intention and nothing prevents them from acting on it. For this reason, intentions are correctly described as *dispositions* to behave in certain ways (Bratman 1987; McCann 1991; Gillesen 2017).

In order to understand how the dispositional nature of intentions undermines Marušić and Schwenklers' argument for strong cognitivism, it may be helpful to illustrate with an example. Suppose that Richard is thinking about renovating his house and that it is entirely up to him whether or not renovates; that is, he is under no obligation to do so. Suppose furthermore, that Richard engages in practical reasoning, leading to the conclusion that he *should* renovate the house in order to increase its market value. On that basis, let us assume, he makes a decision and forms the intention to renovate his house. Now if, as Marušić & Schwenkler suggest, Richard's intention to renovate is identical with a belief that he will renovate, then Richard's belief is quite irrational. It is a form of wishful thinking for anyone to believe that they will do something (e.g. renovate a house) simply because practical reasoning leads them to conclude that they *should* do it. Nor is it possible for Richard to form a reasonable belief that he will renovate simply by considering evidence about what is independently likely to happen. That evidence, as Marušić & Schwenkler correctly point out, would necessarily be inconclusive. However, once Richard makes a decision and forms the intention to renovate, from that point on it would be perfectly rational for him to believe that he (probably) will renovate. In this case, Richard's belief about what he will do would be based, not on considerations about what is independently likely to happen, or simply on reflections about what is good or worthwhile, but also—and crucially—on his intention to renovate. Without that intention, there would be no reason at all for Richard to believe that he will renovate, if it really is up to him whether he renovates.

This is not to say that an agent who forms an intention to X will necessarily form the belief that they (probably) will X. Whether an agent does or does not form such a belief in any given case depends on further contextual details. However, once an agent forms an intention to X, the agent does from that point on at least have a prima facie reason to believe

that they (probably) will X, a reason which they lack prior to having that intention. For example, having decided—and formed the intention—to shoot the president, the sniper described above in Experiment 1 has a prima facie reason for believing that he (probably) will shoot the president. At the same time, after assessing the conditions on the ground at the time of the shooting, the sniper may also have some reason to believe that he (probably) will *not* shoot the president. Whether the sniper believes that he (probably) will shoot the president will ultimately depend on his assessment of both the strength of his intention and the extent of the challenges he faces in executing his intention.

To summarize, the key question that leads Marušić & Schwenkler to embrace the idea of "practical beliefs," which they identify with intentions, concerns how an agent can form a rational belief that they will do something when it is up to them whether they do it. But there is a straightforward answer to that question, one which avoids the assumption that intentions are identical with beliefs and the related assumption that there is a type of reasoning that proceeds from normative judgments to factual beliefs. All rational beliefs about what one will do should be based upon, and responsive to, evidence, but the evidence need not be restricted to considerations of what is independently likely to happen. When matters are up to the agent, the evidence can and should also include things over which the agent has some control, namely, their own decisions and intentions. One's rational beliefs about what they will do are not identical with one's intentions; rather, they are preceded by, and based upon, one's intentions (among other considerations). The key flaw in Marušić and Schwenkler's argument for strong cognitivism is their failure to acknowledge this alternative basis for the rational beliefs that agents can have about what they will do when matters are up to them.

6. Conclusion

While there are many different versions of IC, what unites them all is a commitment to CC1, which states a fairly strong confidence condition on the concept of intention. According to IC theorists, an agent intends to X only if they believe that they (probably) will X. Much of the reason for thinking that intentions are subject to such a confidence condition stems from the fact that intentions are, as Bratman (1987) and others have emphasized, subject to certain rational norms, including a norm of consistency. That is, there is a well-recognized rational requirement that an agent's various intentions be consistent with one another, given the agent's beliefs. However, while the fact that intentions are subject to such norms may motivate the thought that intentions and beliefs are somehow related, it does not justify the assumption that CC1 is true. To think it does is to misunderstand the nature of the rational norm of consistency that governs intentions.

Bratman (1987: 32) clearly recognized that the rational norms governing intentions are defeasible in the sense that "there may be special circumstances in which it is rational of an agent to violate them." It is a mistake, in his view, to think of these norms as principles that admit of no exceptions. Furthermore, and related to this last point, it is a mistake to think these norms reflect the metaphysical structure of intentions or the semantics of the concept of intention. An agent who does on some occasion violate the consistency norms governing intentions is not a logical impossibility but rather an agent who, on that occasion, may be guilty of some limited form of irrationality, as we all are from time to time. McCann (1991: 34) captured both of these points when he wrote that intentions, which he called "settled objections," are "always subject to consistency demands, but only as regards their rationality, and only with a *ceteris paribus* clause that exempts [certain] agents." The sort of agents McCann had in mind in making this

remark were those like sniper in Experiment 1. Faced with unfavorable shooting conditions, the sniper reasonably estimates the chances that he will hit his target are very low. Given these circumstances, it is perfectly rational for him to believe that he (probably) will not shoot the president and also to renounce the contrary belief that he (probably) will shoot the president. But all of this is still consistent with the idea that there is still *some* chance—however small—that he *will* hit his target. If there is such a chance, and the sniper is aware of it, which he surely is (since he estimates that he has a 20 percent chance of hitting his target), then it may be rational for him to maintain his intention to shoot the president and to act on it. This is because, as the sniper acknowledges in the scenario, he may never have a *better* chance to achieve his goal of shooting the president as well as the fact that the costs of failing to shoot the president may for him be very low. In this case, the reasoning that would justify maintaining the intention to shoot the president would be a form of cost-benefit reasoning rather than instrumental reasoning.

This is not to suggest that all violations of the norm of consistency are rational. Buckwalter et al.'s example of the prisoner who is about to be tortured is perhaps an example of one that is not. If the prisoner sincerely believes that it is completely impossible to withstand the torture, then it is no doubt irrational of him to maintain an intention to do so. However, it could be that even though the prisoner (partially) believes that it is impossible to withstand the torture, he might also (partially) believe that is possible to do so. Interestingly, Buckwalter et al.'s experimental design does not rule out this interpretation of their scenario. For those who do interpret the scenario in this manner, the agent might still be seen as rational in the limited sense of having a reason to maintain an intention to X despite strongly believing that they will not X. On the other hand, if an agent in this situation does maintain an intention that is flatly inconsistent with their beliefs, then the agent would be guilty of criticizable irrationality. However,

even in this case neither metaphysical nor semantic considerations preclude the agent from having that intention.

The key mistake that is made by those who embrace IC consists in understanding the norms of rationality that govern intentions as being rooted in metaphysical or semantic truths that make it impossible for an agent to intend to X without at the same time believing that they (probably) will X. A second and related mistake that runs through the writings of all IC theorists to date is that they have approached the question of the relation between the concepts of intention and belief from a purely theoretical perspective without regard to evidence concerning how those concepts are actually used by ordinary speakers of a language. When one does take an empirical and experimental approach to understanding how these concepts are applied to the interpretation of human behavior one finds that violations of CC1, the core commitment of IC, are not only possible, they are commonplace.

Acknowledgement I would like to thank two anonymous referees at this journal for helpful comments on an earlier version of this article.

Declarations

Conflict of Interest The author has no conflict of interest to declare.

References

Anscombe, E. (1957). *Intention.* Oxford: Basil Blackwell.

Audi, R. (1973). Intending. *Journal of Philosophy*, 70: 387-402.

Beardsley, M. (1978). Intending. In A. Goldman & J. Kim (Eds.), *Values and Morals* (pp. 163-184). Dordrecht: Reidel.

Bratman, M. (1984). Two faces of intention. *The Philosophical Review*, 93: 375-405.

Bratman, M. (1987). *Intention, Plans, and Practical Reason.* Cambridge, MA: Harvard University Press.

Buckwalter, W., Rose, D., & Turri, J. (2021). Impossible intentions. *American Philosophical Quarterly*, 58 (4): 319-332. https://doi.org/10.2307/48619317

Clark, P. (2020). Intentions, intending, and belief: non-inferential weak cognitivism. *Pacific Philosophical Quarterly*, 101: 308-327.

Davidson, D. (1980). Intending, in *Essays on Actions and Events.* Oxford: Oxford University Press, pp. 83-102.

Davis, W. (1984). A causal theory of intending. *American Philosophical Quarterly*, 21 (1): 43-54.

Gillessen, J. (2017). Flat intentions – crazy dispositions? *Philosophical Explorations*, 20 (1): 54-69.

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

Harman, G. (1986). *Change in View: Principles of Reasoning.* Cambridge, MA: MIT Press.

Hauthaler, N. (2022). Strong cognitive weaknesses. *Analytic Philosophy.* Advance online publication. https://doi.org/10.1111/phib.12252

Holton, R. (2009). *Willing, Wanting, Waiting.* Oxford: Clarendon Press.

Kane R. (1998). *The Significance of Free Will.* New York: Oxford University Press.

Levy, Y. (2017). Why cognitivism? *Canadian Journal of Philosophy,* 48 (2): 223-44.

Marušić, B., & Schwenkler, J. (2018). Intending is believing: a defense of strong cognitivism. *Analytic Philosophy*, 59 (3), 309-40.

McCann, H. (1991). Settled objectives and rational constraints. *American Philosophical Quarterly*, 28, 25-36.

McGuire, J. (2020). Is there a confidence condition in the concept of intention? *Philosophical Psychology*, 33 (5), 705-730.

Mele, A. (1992). *Springs of Action: Understanding Intentional Behavior.* New York: Oxford University Press.

Paul, S. (2009). How we know what we are doing. *Philosopher's Imprint* 9 (11), 1-24

Setiya, K. (2007). *Reasons without Rationalism.* Princeton: Princeton University Press

Thompson, Michael (2008). *Life and Action.* Cambridge, MA: Harvard University Press

Velleman, J.D. (1989). *Practical Reflection.* Princeton: Princeton University Press.