On Attempting

Tomasz Placek*

1 Problem

Alicja (3 yrs) attempted to break her mug at our breakfast yesterday. How did she go about this? She slowly moved the mug towards the edge of our table until it tilted and began to fall down. It did not break, however, as I caught it before it hit the floor. I scolded Alicja for *attempting* to break her mug, but she did not agree. So we had a row. To convince her, I gave her a short talk on agency and in particular, on the *stit* theory, but I cannot call it a success. I nevertheless report on our exchange as (I think) it casts light on two questions: (1) to what extent is attempting an intentional concept? And (2) how far can one get in analyzing it in the stit framework?

In what follows I make three assumptions. First, "to attempt" is a so-called accomplishment verb, which means that its meaning changes with tenses. A sentence in the past tense, like "I attempted to prove this lemma yesterday", implicates that I failed to prove it. My assertion, "I am attempting to prove this lemma", however, is consistent with my success as well as with my failure. In this text I leave aside this complexity and arbitrarily assume the past-tense meaning of "to attempt", i.e., I take it that attempting is always unsuccessful. Second, I assume an idealization that attempting occurs at point-like moments. Third, although English grammar prescribes the form "α attempts to break her mug", I suggest a little linguistic reform here: I will write "α attempts that her mug will be (is) broken". In other words, I assume that "α attempts that ..." is a propositional operator, my aim being to find adequate truth-conditions for it.

2 *Stit*, branching time (BT), Agents, and Choices

Stit is an acronym for the phrase "see to it that" which, as Belnap, Perloff and Xu (2001) have suggested, might serve to draw a distinction between agentive and non-agentive sentences. To give an example, the sentence "I am going to Prague" is ambiguous, as in some contexts it is agentive and in others it is not. But, if its assertion in a given context can be paraphrased by using the stit form—that is as

*This work is supported by MNiSW research grant 3165/32. It extends my earlier paper, Placek (2009, in Polish), and corrects its errors. For many insightful comments on attempting and the stit theory, I am very grateful to Nuel Belnap.

"I see to it that I am going to Prague"—the sentence is agentive; otherwise it is not.

The stit theory presents the task of understanding agency by explaining the phrase "see to it that". The phrase is assumed to be represented by a propositional operator, written as *stit*; the task amounts to providing a semantics for this operator. The explanation does not appeal to mental concepts, like the agents being conscious, or driven by some intentions, or constrained by moral norms and values. This does not mean that an agent's actions do not have such aspects. The idea is that basic features of actions can be understood without taking recourse to mental concepts, and if there is a need for such a concept, the stit theory will lend itself to an appropriate supplementation.

The semantics for stit is based on the theory of *branching time* (BT), which was put forward by Prior (1967). A language $\mathcal{L}$ considered has, besides the *stit* operator, the tense operators P (it was the case that) and F (it will be the case that), the operators of historical possibility and historical necessity, respectively, $Poss$ and $Sett$, the tense indexicals like 'now' or 'tomorrow', and the truth-functional connectives $\vee, \wedge, \neg, \rightarrow$ of disjunction, conjunction, negation, and (material) implication, respectively.

It is assumed that the atomic sentences of our language $\mathcal{L}$ are in the present tense and significantly tensed. We read "F : Kasia reads a book" as "Kasia will read a book" and analogously for the past tense operator P. The sentence "$Sett: P$: Kasia goes to Warsaw" is read as "It is (already) settled that Kasia has gone to Warsaw". In a similar vein, "$Poss: F$: Kasia goes to Warsaw" is rendered as "It is (still) possible that Kasia will go to Warsaw".

Turning to the branching time theory, its model $\mathcal{W} = \langle W, \leqslant \rangle$ is a non-empty and partially ordered set, subject to the condition of no backward branching—

$$\forall x, y, z \in W (x \leqslant z \wedge y \leqslant z \rightarrow x \leqslant y \vee y \leqslant x),$$

and the condition of historical connection, i.e.,

$$\forall x, y \in W \; \exists z \in W : \; (z \leqslant x \wedge z \leqslant y).$$

Since histories are defined as maximal chains in W, the last condition implies that every two histories intersect, which justifies the condition's name. Elements of W are called "events", which is rather unintuitive since a BT event is an instantaneous slice of the universe, i.e., a class of point events simultaneous with a given point event. (As the theory assumes absolute simultaneity, it is pre-relativistic.) Finally, $x \leqslant y$ means that y belongs to a possible future of x, or (equivalently) that x is in the past of y. We write $Hist$ for the set of histories in $\mathcal{W}$.

There is a family of undividedness relations $\equiv_e$ on $Hist$, parametrized by events $e \in W$ and defined as $h_1 \equiv_e h_2$ iff $\exists e' \; e' \in h_1 \cap h_2 \wedge e < e'$. It is straightforward to see that $\equiv_e$ is an equivalence relation on $H_{(e)} := \{h \in Hist \mid e \in h\}$. Thus, $\equiv_e$ induces the partition Π_e of $H_{(e)}$.

Some BT models allow for the introduction of a set Ins of instants, thought of as instantaneous moments at which events occur. It is required that (1) every element of Ins and every history intersect at a single event, and that (2) Ins respects ordering $\leqslant$ on W, which amounts to the following conditions. For every $e_1, e_2, e_1', e_2' \in W$, every $i_1, i_2 \in Ins$, every $h_1, h_2 \in Ins$:

$$h_1 \cap i_1 = h_1 \cap i_2 \Rightarrow h_2 \cap i_1 = h_2 \cap i_2 \tag{1}$$

$$(e_1 \in h_1 \cap i_1 \wedge e_2 \in h_1 \cap i_2 \wedge e_1 < e_2 \wedge e_1' \in h_2 \cap i_1 \wedge e_2' \in h_2 \cap i_2) \Rightarrow e_1' < e_2'. \tag{2}$$

We write $i(e)$ for the instant determined by event e.

Turning to the BT semantics, Prior's main semantical idea is to take for evaluation points the event-history pairs in which the event belongs to the history. Such a pair is written as e/h. A BT semantical model $\mathfrak{M} = \langle \mathcal{W}, I \rangle$ for our language $\mathcal{L}$ consists of a BT model $\mathcal{W}$ and an interpretation function $I : Atoms \to \mathcal{P}(W)$, where $Atoms$ is the set of atomic formulas of $\mathcal{L}$. The evaluation is defined as follows:

$$\text{For } A \text{ an atomic formula: } \mathfrak{M}, e/h \models A \text{ iff } e \in I(A); \tag{3}$$

$$\mathfrak{M}, e/h \models \neg\varphi \text{ iff it is not the case that } \mathfrak{M}, e/h \models \varphi; \tag{4}$$

$$\text{and similarly for the other classical connectives;} \tag{5}$$

$$\mathfrak{M}, e/h \models F\colon \varphi \text{ iff there is } e' > e \text{ such that } \mathfrak{M}, e'/h \models \varphi; \tag{6}$$

$$\text{and analogously for a formula with } P \text{ as the main operator;} \tag{7}$$

$$\mathfrak{M}, e/h \models At_i\, \varphi \text{ iff } \mathfrak{M}, e'/h \models \varphi, \text{ where } e' \in i \cap h; \tag{8}$$

$$\mathfrak{M}, e/h \models Sett\colon \varphi \text{ iff for every } h', \text{ if } e \in h', \text{ then } \mathfrak{M}, e/h' \models \varphi. \tag{9}$$

$Poss\colon \varphi$ is defined as equivalent to $\neg\colon Sett\colon \neg\varphi$.

To represent agents' actions, one first postulates a non-empty set $Agents$. Then choices available to $Agents$ at events from W are coded by a "choice function" $Choice : Agents \times W \longrightarrow \mathcal{P}(Hist)$ subject to the conditions: (1) $Choice(\alpha, e)$ is a partition of $H_{(e)}$ and (2) if $h \in Choice(\alpha, e)$ and $h \equiv_e h'$, then $h' \in Choice(\alpha, e)$. We will write $Choice_e^\alpha$ for $Choice(\alpha, e)$ and $Choice_e^\alpha\langle h \rangle$ for the element of $Choice_e^\alpha$ determined by h (this is defined provided that $h \in H_{(e)}$).

3 First idea: attempting is risking

Our analysis of attempting will proceed in stages. The first idea is to link attempting to risking. Namely, α attempted that the mug be broken (R) at instant i means that she risked that R at instant i. We suggest analyzing this in the following manner: there was an option available to α in which certainly $\neg R$ (the mug is not broken) at i, but α chose another option which allowed that the mug be broken at i (see Figure 1). In symbols our proposal reads:

$\mathfrak{M}, e/h \models \alpha$ risks that $At_i R$ iff

1. $\exists h' \in Hist\colon h' \in Choice_e^\alpha\langle h \rangle \wedge \mathfrak{M}, e/h' \models At_i R$,

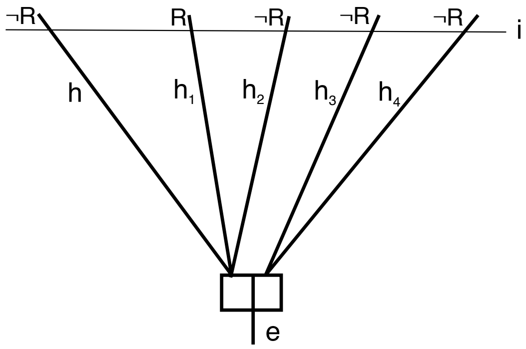

Figure 1: At e/h agent α risks that $At_i R$.

2. $\exists A \in Choice_e^\alpha\ \forall h'' \in A\colon \mathfrak{M}, e/h'' \models At_i \neg R$.

Risking so described relies on forsaking a sure thing in favor of a not-so-sure thing. The concept is too strong, however. By the above explication, I should not say that Alicja risked breaking her mug (even though she put it on the very edge of our table), if I believe that, if left alone, the mug could still break, either for some physical reasons, or because of the workings of other agents — our cat, for instance. A question thus arises with respect to what do we risk? In particular, should we require of a state that a risky agent forsake that it has a sure-thing characteristic? We will return to this question in the next section.

Note also that in this analysis, attempting is ubiquitous. By deciding to spend this day in bed, I would ensure that quite a few things would (not) happen. For instance, I would ensure that my milk pot would now be intact. Yet, I got up, I decided to boil milk for Alicja, put the pot on our stove, and ... forgot about it. The pot would have burned if my wife had not intervened. So, by deciding to get up, I attempted to burn the pot, and most likely I attempted an indefinite number of other things.

It seems we may improve on our first analysis by combining it with some concept of guaranteeing: I attempted that R because first I risked that R and second, although later I could still guarantee that $\neg R$, I did not do it. So we turn next to guaranteeing.

4 Guaranteeing I

Consider the four diagrams of Figure 2. In each diagram at e/h it is true that agent α risks that R at instant i. In the top two diagrams, after taking this risk (by picking the "left" option), α can ensure at a later event e' that $\neg R$ at i. One might got the impression, however, that guaranteeing is achieved by an accidental fact: α's choices at e' (as coded by $Choice_{e'}^\alpha$) combine nicely with possibilities on which α has no influence. In the top-right diagram, α has no way of choosing the third history from the left rather than the first or the second, but fortunately, in this history we have $\neg R$ at i. And in the top left diagram, there is simply no history which α cannot choose.

The bottom two diagrams, however, contain histories that destroy guaranteeing, as they have R at i, and α cannot exclude them. Before addressing this problem, we need to decide what should be on the right-hand side of those diagrams that illustrate guaranteeing. This amounts to deciding with respect to what we risk or attempt. In other words, if Alicja attempted to break her mug by taking such and such an action, what should we say would happen to the mug if she did not take this action? What should be demanded of choice $A \in Choice_e^\alpha$, which α does not take at e/h, that is, $A \neq Choice_e^\alpha\langle h\rangle$?

(1) Surely the mug would not break. Formally,

$$\forall h' \in A\, \forall e'\, (e' \in i \cap h' \rightarrow e'/h' \models \neg R.$$

Clearly, this condition is too strong. We may still believe that α risks breaking the mug, even though we think that without α's action the mug's integrity is nonetheless endangered by our cat, a draught, or Mary's action. A possible obstructive action of some other agent does not get α off the hook, so to speak.

(2) Other agents, working separately or jointly, could ensure that the mug is not broken. The idea is that option A contains a history in which the mug is broken at i, but other agents have choices available that would lead to $\neg R$ at i in the appropriate history. This makes α's attempting dependent on choices available to other agents, and this is not a good idea.

(3) α herself might guarantee that $\neg R$. Despite having been able at e/h to guarantee that $\neg R$ at i, she chose at e another option such that she could not later guarantee that $\neg R$ at i. However, in the present reading we allow for the obstructive actions of other agents and obstructive indeterministic workings on the part of nature even in the "guaranteeing option". The idea here is to understand guaranteeing in a rather relaxed, or counterfactual way: if there were no obstructive workings by other agents or nature, α could guarantee $\neg R$ at i. Thus the task is to put in brackets the obstructive workings of other agents and nature. To analyze this idea, we will employ the concept of simple strategies of Belnap, Xu, and Perloff (2001), which we even further simplify for the present purposes.

5 Guaranteeing II and simple strategies

Our aim now is to analyze, in the framework of simple strategies, the phrase "α can guarantee that Q at instant i", where we assume the relaxed interpretation of "guaranteeing" as indicated just above.

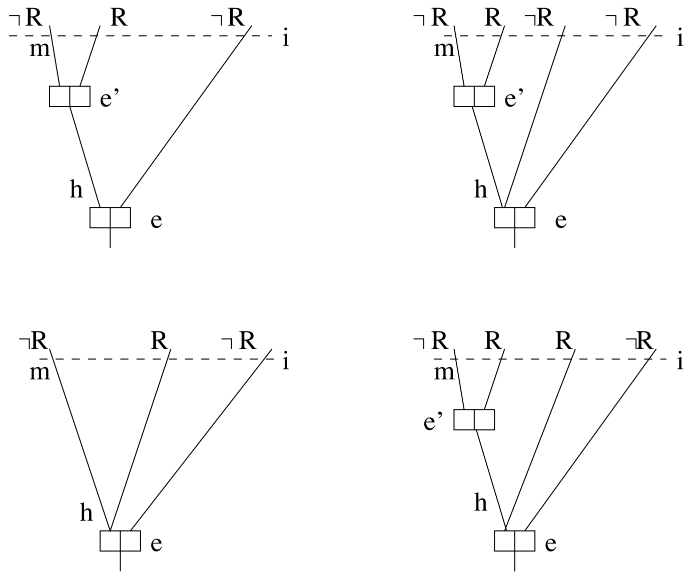

Figure 2: Guaranteeing I.

A strategy s available for $\alpha \in Agents$ is a partial function on W such that:[1]

$$\text{if } e \in Dom(s), \text{ then } s(e) \in Choice_e^\alpha.$$

We say that s is a strategy for α in field M if $Dom(s) \subseteq M \subset W$. To define simple strategies, one needs the following notions:

1. Admissibility: s admits e iff

$$\forall e_0\ (e_0 \in Dom(s) \wedge e_0 < e \rightarrow e \in \bigcup s(e_0));$$

2. Primariness: strategy s is primary iff $e \in Dom(s) \rightarrow s$ admits e;

3. Backward closure: for strategy s for α in field M, we say that s is backward closed in M iff

$$\forall e_1, e_2\ (e_2 \in Dom(s) \wedge e_1 < e_2 \wedge e_1 \in M \rightarrow e_1 \in Dom(s));$$

4. Simplicity: a strategy s for α in M is simple iff it is primary and backward closed in M.

Consider now the set of histories such that at the pair "event e, history", the sentence "At_iQ" is true, i.e.,:

$$H_e^{At_iQ} = \{h \in H_{(e)} \mid e/h \models At_iQ\}.$$

[1]To simplify, we assume here that a strategy is strict in the sense that $s(e) \in Choice_e^\alpha$, rather than $s(e) \in \mathcal{P}(Choice_e^\alpha)$.

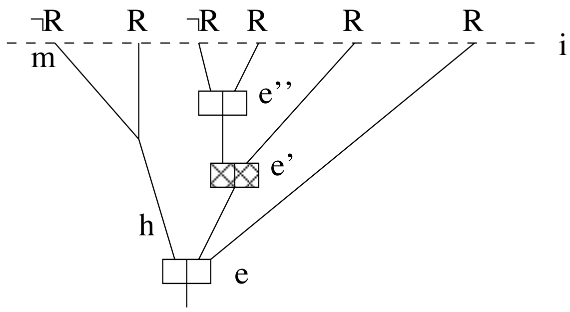

Figure 3: Alicja and our cat at the table. White and shadowed boxes represent Alicja's actions and our cat's actions, respectively.

With this set, we explain the phrase "α can guarantee that Q at instant i" as follows:

$e/h \models \alpha$ can guarantee that Q at instant i iff there is a simple strategy s for α in $M = \bigcup Choice_e^{\alpha}\langle h \rangle$ such that: $\bigcap_{x \in Dom(s)} s(x) \subseteq H_e^{At_i Q}$.

Let us return to Alicja's actions at the table. The situation is depicted by Figure 3, with white boxes representing Alicja's choices and a shadowed box indicating the choices of some other agent, say, our cat. Can Alicja guarantee that the mug is not broken, $\neg R$, at instant i? At e/h it is not true that she can guarantee that $\neg R$ at i: there is no strategy for Alicja with field $M = \bigcup Choice_e^{Alicja}\langle h \rangle$ such that $\bigcap_{x \in Dom(s)} s(x) \subseteq H_e^{At_i \neg R}$. In other words, if Alicja selects the "left" option at e, she cannot guarantee that the mug is not broken. However, she can guarantee this, if she chooses the "right" option at e, because there is a strategy for Alicja with field $M = \bigcup Choice_e^{Alicja}\langle h'' \rangle$ such that $\bigcap_{x \in Dom(s)} s(x) \subseteq H_e^{At_i \neg R}$ for arbitrary $h'' \in Choice_e^{Alicja}\langle h' \rangle$. The strategy relies on taking the "right" option at e and then the "left" option at e''. Note that in the "right" option the verdict is that Alicja can guarantee ..., despite a possible obstruction by our cat at e'.

6 A final analysis

We have come to our final analysis:

$e/h \models \alpha$ attempts that $At_i R$ iff

1. $e/h \not\models At_i R$, and
2. $e/h \not\models \alpha$ can guarantee that $\neg At_i R$, and
3. there is $h' \in H_{(e)}$ such that: $e/h' \models \alpha$ can guarantee that $\neg At_i R$.

My story about Alicja's behavior at the table now goes like this: At e/h Alicja attempted to break her mug in 1 sec. How did she do that? At e/h she placed her mug on an edge of our table. At this event it was not yet settled whether the mug would break: both scenarios were possible. Alicja had other options at e: she might have left her mug in front of her, where it stood. Yet, even if she had chosen to leave the mug, it might still have broken, which might have happened spontaneously (the farmost right history), or not. In the latter scenario, our cat will come to the mug (event e). The cat will face a choice, to break the mug outright or to play with it. In the latter option, Alicja has a choice to save the mug or not (at e''), by chasing the cat away or not.

7 Objections

Our analysis leaves aside the intentional aspects of agents' actions. The two objections presented below try to show that, as a result, the analysis is both too wide and too narrow.

Consider a careful pedestrian on a sidewalk, choosing whether or not to start crossing the street. As there is no incoming traffic to be seen at this moment, he elects to briskly walk through the crossing. When he is in the middle of the road, however, with a cosmic velocity a mad car appears, heading directly towards him. Fortunately, the mad driver turns his vehicle left at the last moment, saving the pedestrian. Now, did the pedestrian attempt to get injured by deciding to cross the street? Clearly, given his option of not crossing the street, he could arguably guarantee that he be uninjured. He chose to cross the street, however, the consequence being that he could not guarantee that he be uninjured. So, by our analysis, the pedestrian attempted to get injured. But, intuitively speaking, he did not. "It was not his intention to get injured" comes as a natural comment.[2]

In the second story, a platoon of soldiers is surrounded by a dreadful enemy.[3] There is no way for them to escape the ambush. Nevertheless, the commanding officer gives the platoon an order to attack. As a result, all the soldiers die. The officer, later tried for foolhardily sending his soldiers to certain death, maintains a defense that by giving the order to attack, he *attempted* to get the platoon out of the ambush. So his fate depends on the truth at e/h of the sentence: "The officer attempts that the platoon be out of the ambush at some later instant i", where e is a moment of issuing the order and h is an arbitrary history in which the trial occurs. In symbols,

$$e/h \models \text{the officer attempts that } At_i\, Out\,.$$

The prosecution, however, has solid evidence that it was impossible to escape the ambush. Also, on the basis of prior evidence concerning the enemy's cruelty, it is clear that if the soldiers had surrendered, the enemy would have killed them. Thus, well before issuing the order, there was no possible scenario in which the platoon survives. This means that in every history comprising the event of issuing

[2] I owe this example to John Kearns.

[3] This story comes from Tim Childers. I am indebted to Michał Araszkiewicz for explaining to me the definitions of attempting as they occur in the Polish criminal code.

the order the sentence $\neg$ *Out* is true at instant i. By the concept of guaranteeing, this means that at e/h the officer can guarantee that $\neg At_i\, Out$. Accordingly, the second clause of our final analysis is not satisfied, so this analysis yields a negative verdict: the officer did *not* attempt to get his platoon out of the ambush.

Common sense might say the opposite, in particular, if the officer convincingly shows that (1) he thought (mistakenly) that the situation was not hopeless and (2) he had a plan (unrealistic, as it was) of how to rescue the soldiers, and the order in question was a part of this plan. As far as I could learn, the court verdict might go both ways, depending on the nature of the officer's mistaken assessment of the situation.

My own views agree with the final analysis, however. The officer did not attempt to rescue the platoon. He thought (believed) that he was attempting to save the platoon, and, perhaps, one can even make a case that he was justified in believing that he was attempting to save the platoon. Nevertheless, my strong intuition is that, given the hopeless predicament, the officer was mistaken: he was not attempting to save the platoon.[4]

Dismissing the second story as a counter-example to the final analysis, I am left with the first story which shows that the analysis is too wide. The story, however, suggests clearly a condition to be added to our analysis: an agent's intention. The challenge (which we do not know how to address) is to express this condition in the stit framework, however. For the record, I nevertheless put down this idea as an informal, unofficial, and post-final analysis:

$e/h \models \alpha$ attempts that At_iR iff

1. $e/h \not\models At_iR$, and
2. $e/h \not\models \alpha$ can guarantee that $\neg At_iR$, and
3. there is $h' \in H_{(e)}$ such that: $e/h' \models \alpha$ can guarantee that $\neg At_iR$, and
4. $e/h \models \alpha$ intends that At_iR.

Examples similar to our first story as well as criminal codes indicate that intending, planning, or even pre-meditating have some role in attempting. To obtain some clarity about what exactly this role is will be a task for a future project. Yet another task is to see if the recent stit logic for belief, desire, and intention by Semmling and Wansing (2009) can be harnessed to analyze the mental aspects of attempting.

Tomasz Placek
Department of Philosophy, Jagiellonian University
Grodzka 52, 31-044 Kraków, Poland.
tomasz.placek@uj.edu.pl

[4] Of course, I am applying here a weak understanding of justification according to which we can have justified false beliefs.

References

Belnap, N., Perloff, M., & Xu, M. (2001). *Facing the future.* Oxford: Oxford University Press.

Placek, T. (2009). O usiłowaniu. In J. Czetwertyński-Sytnik (Ed.), *Rozważania o filozofii prawdziwej* (pp. 105–115). Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.

Prior, A. (1967). *Time and modality.* Oxford: Oxford University Press.

Semmling, C., & Wansing, H. (2009). A sound and complete axiomatic system of bdi–stit logic. In M. Pelis (Ed.), *Logica Yearbook 2008* (pp. 193–210). London: College Publications.