The Semantics of Ability: *Can* and *Could Have*

Reuben Cohn-Gordon

June 23, 2017

1 Ability Readings of *Can* and *Could Have*

The English modal *can* is often used used to express a claim about an agent's ability[1]. Suppose that Groucho encounters a stream and considers jumping across, rather than using the nearby bridge. In these circumstances, (1) and (2) appear to mean different things:

(1) Groucho can make it across the stream by jumping.

(2) Groucho might make it across the stream by jumping.

While (2) asserts the existence of a series of events that would lead to Groucho crossing the stream by jumping, i.e. that this outcome is *possible*, (1) is the claim that Groucho has the *ability* to do so.

It has been noted repeatedly that modeling the semantics of ability differs from "simple" possibility (see (Cross 1986, Brown 1988)). For instance, *might(P)* does not always entail *canP*, on the ability reading of the latter. (Thomason 2005) observes that while I *might* enter the correct password if I try to log into a stranger's email account, it does not follow that I have the *ability* to log into their account. Thus, at least on one reading of *can*, "I can log into McCarthy's email account." appears to be false.

A similar ability reading exists for *could have.* Suppose Groucho decides against jumping, and crosses the river using the bridge. We can then consider analogues to (1) and (2), namely (3) and (4):

(3) Groucho could have made it across the stream by jumping (though he didn't try to).

(4) Groucho might have made it across the stream by jumping (though he didn't try to).

[1] Other common uses include the deontic "You can leave the room now." and the circumstantial "Hydrangeas can grow in acidic soil."

There seems to be a reading of (3) which is a past tense version of (1): i.e. the claim that in the past, Groucho had the ability to jump across the stream. Again, on such a reading, (4) does not entail (3).

The matter here is complicated by the existence of a reading of *could have* on which it is equivalent to *might have* (on a "metaphysical" rather than epistemic reading of the latter)[2]. That is, there is also a different reading of (3), on which it is equivalent to (4). The existence of this "metaphysical" reading makes the ability reading of *could have* somewhat elusive, and perhaps accounts for the fact that this past tense ability reading of *could have* is seldom discussed, and even ruled out by (Cross 1986) ("...the past tense of 'can' is...never 'could have' - when 'can' is used to mean 'have the ability to'.").

Building on the proposal for the semantics of ability offered by (Mandelkern, Schultheis, and Boylan 2016), I will present a new semantics of ability *can* that resolves a number of outstanding issues (discussed in section (2)) and propose to model ability *could have* as its past tense counterpart. After doing this, I'll offer a characterization of the circumstances in which the ability reading of *could have* appears.

2 The Properties of Ability

A distinction can be made between one-off and generic ability. For instance, Groucho may have the generic ability to juggle, but lack the ability on a particular Friday evening, after a few beers. The present paper is concerned with one-off ability. The motivation for this is the view that genericity is an orthogonal issue, that applies to more than just ability (for an overview of this field of study, see (Carlson 1995)). For this reason it seems sensible to design a semantics for one-off ability, and then account for generic ability through a theory of generics (as (Mandelkern, Schultheis, and Boylan 2016), for instance, attempt to do.). For this reason, the focus of the present paper is to develop a semantics for one-off ability, i.e. the ability to perform a given action at a given time and place.

A semantics for one-off ability *can* has to account for a number of theoretical considerations. Firstly:

(5) "A might ϕ"[3] does not entail "A can ϕ" and "A might not ϕ" does not entail "A can not ϕ".

[2] An example of this "metaphysical" reading of *could have* is found more clearly in (1) and an epistemic reading of *could have* is found in (2):

(1) The bike could have slipped on the ice, but it didn't.

(2) John could have come to this party, but I'm not sure if he did.

[3] We follow (Mandelkern, Schultheis, and Boylan 2016) in treating actions as properties, i.e. functions from an entity at a world and time to a truth value. Given that the abilities in question are one-off, not generic, this seems appropriate.

(5), observed by (Brown 1988), will be the basis for a rejection of simple existential account of ability *can*, as discussed in section (3.2). A typical case is the lottery. Since, in a fair lottery (i.e. one which is not rigged in any way), the winning number is not determined by any participant, no participant has the *ability* to win, even though any participant *might* win.

(6) In certain situations, failure for an agent to ϕ is grounds to infer that the agent does not have the ability to ϕ.

(6) is also observed by (Brown 1988), among others: if Caesar *had* gone to war and lost spectacularly, this could be grounds for rejecting the truth of (1), as uttered at a time before Caesar was killed. One intuitive example relates to betting: if Groucho bets Karl ten dollars that he can hit the bullseye on his next throw, and then misses, he presumably owes Karl ten dollars.

However, note that failure for an agent to ϕ is not always evidence that the agent lacked the ability to ϕ. For instance, if Caesar had *decided* not to go to war, or (as in fact happened) had been assassinated first, the fact that he then would not have gone on to win the war against the Parthians doesn't seem to be evidence against his ability to have done so. A theory of ability *can* needs to explain when failure to ϕ should count as evidence that an agent did not have the ability to ϕ.

(7) If A has the ability to ϕ then A has the ability to *not* ϕ.

An example of (7) is the action of crossing a river on a bridge. Groucho has the ability to do so, and also the ability to not do so. This relates to the notion of intention: if Groucho *chooses* to cross the bridge, he will succeed (all else being equal), and if he chooses not to, he will succeed in not crossing.

(8) The dual of ability is compulsion. An agent is compelled to ϕ iff they do not have to ability to not ϕ.

(8) makes a connection between ability modals and so-called compulsion modals (as coined by (Mandelkern, Schultheis, and Boylan 2016)). Typical English compulsion modals include "cannot but" and one reading of "have to", as in:

(9) Groucho cannot but sneeze.

(10) Groucho has to sneeze.

These appear to be the duals of ability modals. To say that Groucho is compelled to sneeze is to say that he does not have the ability to *not* sneeze.

(11) An ability to ϕ may require an agent to perform a sequence of actions, rather than just one.

Unlike (5), (6) and (8), (11) is rarely discussed in the literature, and will require a separate innovation to account for. (11) concerns cases where doing some ϕ requires a series of actions. For ϕ="defeating the Parthians", Caesar would have had to perform a series of actions, after each of which, a number of different events could transpire. For instance, Caesar might first have to move his troops to Parthia. If he is then ambushed, he will have to maneuver his troops to encircle the Parthians. If not, he will have to employ a different strategy. If we consider Caesar to have had the ability to defeat the Parthians, it seems that we are claiming that there is a series of actions that he can take that will lead to his victory, no matter what happens after each action.

To capture these five key properties, I propose an adaption of an existing semantics for ability modals, proposed by (Mandelkern, Schultheis, and Boylan 2016). Their analysis, termed the *act-conditional analysis*, treats claims of ability as hypothetical claims about the consequences of performing certain actions. Consequently, the semantics they propose can be characterized in their words as follows:

> "An agent has an ability to do ϕ iff there exists an action A available to them such that in *the closest world* in which they perform the action, they do ϕ."

This account avoids a number of difficulties faced by simpler accounts relating to properties (5) and (8) (see sections (3.2, 3.3 and 3.4) for a detailed overview). However, it also has a number of shortcomings relating to (6) and (11).

In response to these shortcomings, I will propose a new semantics, which I term the *game-semantic analysis*[4]. Informally:

> An agent has an ability to do ϕ iff there exists an action A available to them such that in *all the closest worlds which are fair* in which they perform the action, they *either* do ϕ *or* they can do ϕ.

A useful example is a game of chess. At a given point in the game, an agent has a set of available actions, constrained by the rules of what constitute a valid chess move. To say that an agent has the ability to win means that there is some move that, *for any move* the opponent makes in response, either the agent wins the game, or the agent has the ability to win the game (via a further succession of moves).

[4]This name is inspired by an informal analogy between the modeling of ability through iterated actions, and the notion of a winning strategy in a game. The application of games to logical semantics has a long history (see Hintikka and Sandu 1997), in which a proposition is held to be true if a verifier has a winning strategy against a falsifier in a two player game. Here, a similar notion is applied to natural language, to model ability.

3 The Shortcomings of Existing Accounts of Ability *Can*

3.1 Background Assumptions

The account of ability I will go on to propose will be built on top of a modal-temporal world model. For the sake of consistency, I will present the other accounts I consider and reject in the same setting. It will therefore be useful to outline the assumptions that the semantics I propose, as well as the proposals I reject, rest on.

Firstly, a world is a function from times to maximal consistent sets of propositions. I abuse notation so that $< w, t >$ denotes the set of propositions true in w at t. To model context, I add an extra parameter to the interpretation function, so that P is valued at a context C and a world-time pair $< w, t >$. I will write $\mathrm{P}_{<w,t>}$ to denote that P holds at $< w, t >$.

Two worlds w and w$'$ are *historically equivalent* up to t iff $\forall$t$'$ $\leq$t, $< w, t' > = < w', t' >$. (Here equality is to be understood as equality of sets.). (13) defines a set HistEq at $< w, t >$, denoting the historically equivalent worlds to w at t.

We say that a proposition P is undetermined at $< w, t >$ if $\exists$w$'$,w$''$ such that $< w', t >$ and $< w'', t >$ are both historically equivalent to $< w, t >$, and $\exists$t$'$>t such that $\mathrm{P}_{<w',t>}$, while $\forall$ t$'$>t, $\neg\mathrm{P}_{<w'',t>}$ (see (13)). Informally, there are two historically equivalent worlds to w at t, such that P only goes on to hold in one but not the other[5].

We can understanding this model as a way of interpreting "branching worlds". At a given point in time, we evaluate the truth of a proposition at a world w. However, w is indistinguishable from (on account of being historically equivalent to) a set of worlds {w',w"...}. The futures of these worlds represent the possible continuations of the future at w.

It will also be worth outlining what is meant by an ordering source - a common construction in Kratzerian modal semantics. Given a set of worlds, the ordering source is a way to take the subset most in accordance with some set of propositions, which I'll refer to as the *test conditions*, since for present purposes, the ordering source will be used to choose worlds which provide fair tests of an ability. Formally, (12) defines the ordering source as a function which takes a set and returns the optimal worlds, given a set of test conditions, TEST. TEST, in turn, is world and context dependent.

(12) $\mathrm{O}_{(TEST)}(\mathrm{S}) = \{w' \in \mathrm{S} \mid \forall \mathrm{w}'' \in \mathrm{S}, \{\mathrm{R} \in \mathrm{TEST} \mid \mathrm{R} \in < w', t >\} \supseteq \{\mathrm{R} \in \mathrm{TEST} \mid \mathrm{R} \in < w'', t >\}\}$

[5]Whether any propositions are genuinely undetermined in the real world, is a question we do not address: as far as the model is concerned, events like the lottery are genuinely undetermined, in the sense defined above.

(13) $\text{HistEq}^{<w,t>} = \{w | \forall t' \leq t, < w', t' > = < w, t' >\}$

(Mandelkern, Schultheis, and Boylan 2016) provides an account of the shortcomings of naive existential, universal and conditional accounts of ability *can*. It will prove useful to summarize these arguments as well as adding further ones, before moving on the shortcomings of the analysis of (Mandelkern, Schultheis, and Boylan 2016) itself and my proposed solution.

3.2 Attempt 1: Existential Quantification for Ability *Can*

A first attempt to model ability *can* might assume that it is a possibility operator, i.e. an existential quantifier over some set of worlds. This is a reasonable approach, given that other uses of *can* can be analyzed in this way (e.g. deontic *can*). More formally:

(14) $[\![\text{A can } \phi]\!]^{<w,t>,O,C}$ iff $\exists w' \in O_{(TEST^{<w,t>},C)}(HistEq^{<w,t>})$, $\exists t' > t$: $[\![\text{A does } \phi]\!]^{<w',t'>,C}$

Here, C denotes a context. The ordering source O is parametrized by a set of test conditions TEST, which is chosen by the context and the world and time[6]. Precisely what should be in TEST is not specified in (14). A reasonable choice would be to have the ordering source restrict consideration to worlds which satisfy physical laws, so that the worlds whose futures we consider are not aberrant in this regard. This would be achieved by having the set of test conditions for the ordering source, TEST, describe the laws of physics. We could, in a similar vein, have a stereotypical ordering source, obtained by a set of test conditions which described "normal" worlds.

Under (14), then, having an ability to ϕ would amount to performing ϕ in some possible future of the world, provided it followed, at the least, the physical laws of the world. To give an example of this semantics in action, suppose Groucho, an expert dart player, is about to throw a dart. On (14), we evaluate whether Groucho is able to hit the bullseye by seeing if there is a possible continuation of the future that accords to some restrictions (i.e. which is maximal in the ordering source) in which he hits it.

The problem with such a semantics for ability *can* arises in cases of undetermined events. For example, suppose Groucho is about to buy a lottery ticket. On an ability reading of *can*, we would expect (15) to be false, provided that the lottery is not rigged in any way. Of course, Groucho *might* win the lottery - the point is that he lacks the ability to:

(15) Groucho can win the lottery.

[6] In Kratzerian terms, the set of historically equivalent worlds to w at t forms the modal base, from which the ordering source O selects a subset.

I now formulate a simple argument to demonstrate that we will not be able to obtain the falsity of (15) from the existential semantics of ability *can* proposed in (14), no matter what ordering source we choose to employ.

First Premise Faced with the need to value (15) as false, the existential account of ability *can* has only one recourse: to use the ordering source to constrain the domain of worlds quantified over such that in *none* Groucho loses the lottery. Otherwise, (15) will be true.

Second Premise The outcome of the lottery is undetermined. For simplicity, let us suppose that there are 100 numbers that could win the lottery, of which Groucho could buy any one. Supposing the lottery is random, there are 100 worlds, w1 - w100, which are historically identical up to the time t′ at which the winning number is drawn.

Third Premise Without loss of generality, suppose that Groucho's number is 35. Then there is a world, let's say w40, in which the number 40 is drawn and he loses. But there is also a world w35 in which his number is drawn and he wins. There is no principled way to exclude w35 from the quantification while including w40 and the others worlds, w1, w2, etc. This is because, by construction, at every point before t′, w35 and w40 are identical.

Conclusion Whatever ordering source we choose, there will be a world in the ensuing domain of quantification in which Groucho goes on to win the lottery. Because the quantification is existential, this guarantees the truth of (15), no matter which ordering source we choose.

If these premises are accepted, the conclusion is a fatal flaw for at least this version of an existential semantics for ability *can*.

3.3 Attempt 2: Universal Quantification for Ability *Can*

The failure of the existential account suggests a simple alternative: dualizing the modal quantifier, so that it has universal force. With this, the definition of ability *can* would look as follows:

(16) $[\![\text{A can } \phi]\!]^{<w,t>,O,C}$ iff $\forall w' \in O_{(TEST^{<w,t>},C)}(HistEq^{<w,t>})$, $\exists t' > t : [\![\text{A does } \phi]\!]^{<w',t'>,C}$

Here, TEST cannot be the same set as it was in (14). If it were, many true claims of ability would come out false: for instance, take the case of Groucho's ability to hit the bullseye in a game of darts. If worlds where, in the future, Groucho does not hit the bullseye are included in the domain of quantification, "Groucho can hit the bullseye." will come out false, even if Groucho is an expert dart player and *does* have the ability to hit the bullseye. An example of a world that needs to be ruled out is one where someone pushes Groucho as he is throwing the dart, causing him to miss.

For this reason, it is necessary to introduce a much stricter ordering source, and therefore, a larger set of test conditions, than would be used in (14). We can still assume, however, that the set from which the maximal worlds under the ordering source are chosen is a set of historically equivalent worlds to the base world and the time of utterance[7]. It is worth outlining what this ordering source would look like, as it will play a key part in the eventual proposal I offer for ability *can*.

This ordering source will have to rule out worlds where an intervention prevents the agent from performing an action ϕ, for instance, Groucho hitting the bullseye. To do this, the ordering source must specify the conditions under which an action can be tested fairly. Thus, TEST will include a proposition describing the conditions under which Groucho makes the shot, and stating that no-one physically interferes. It will also contain propositions like "Groucho has arms." and so on - in short, anything that is a contextually reasonable test condition for Groucho's ability.

The Benefits of the Universal Account The universal account of *can* satisfies (5), since an existential claim (*might*) does not entail a universal one (which *can* is, on this account).

As well as (5), the universal analysis of *can* addresses (6), the observation that the failure to perform an action can witness against the agent having a certain ability. For instance, if Groucho tries to hit the bullseye in darts and fails, all else being equal, this seems to be evidence that he lacked the ability to hit the bullseye on that throw.

(17) Groucho can hit the bullseye on this throw.

The universal approach in fact captures the connection between the failure to hit the bullseye and the lack of ability to do so as an entailment: suppose that at $< w, t >$, the claim (17) is made. Let w$'$ be a historically equivalent world to w up to t. In w$'$, Groucho is presented with a fair test of his ability (no-one cheers or pushes him, for example). At time t$' > t$ in w$'$, the dart misses the bullseye.

Note that the modal in (17), under the semantics of (16), is quantifying over a set of worlds which includes w$'$, so long as w$'$ is maximal in the ordering source, which it is, supposing, as we have, that it is a fair

[7]Restricting the modal base to the historically equivalent worlds seems reasonable: it amounts to saying that when judging whether an agent has an ability, the worlds we look in are those with the same pasts and different futures.

test of Groucho's ability. This means that if Groucho misses the bullseye in w′, (17) is false. Intuitively, this makes sense: w′ counts as a world which provides a valid test of Groucho's ability, and in it, Groucho fails to hit the bullseye.

By contrast, if w′ were *not* maximal in the ordering source, as a result of some factor that meant that it was not a fair test of Groucho's ability, it would not be in the domain of quantification for *can* in (17), and so, Groucho's failure to hit the bullseye in w′ would not witness the falsity of (17), uttered about the throw in question. For instance, if Groucho was pushes and subsequently missed the bullseye, this would not entail that he lacked the ability to hit it.

The problems with the universal semantics for *can* Though the universal approach avoids the fatal flaw faced by the existential approach, and satisfies both (5) and (6), it fails on (8). Consider (18):

(18) Groucho cannot but hit the bullseye.

The natural way to model "A cannot but ϕ" is as the dual of "A can ϕ", i.e. as equivalent to "It is not the case that A can do (not ϕ)". The problem with this is that if ability *can* is universal, "cannot but" is existential.

The consequence of an existential semantics for "cannot but" is that it will then be logically possible to have both "A cannot but ϕ" be true and "A cannot but not ϕ". As an example, we can once again make use of the case of Groucho, a skilled darts player, who is about to throw a dart. There is a historically equivalent world that is maximal in the ordering source in which he hits the bullseye. So (18) is true, on the assumption that *cannot but* is dual to ability *can*.

However, there are also historically equivalent worlds in which Groucho does not even try to throw the dart. Some such world will be maximal in the ordering source, provided that the test conditions are fair for Groucho trying to *not* hit the bullseye (e.g., no-one is physically compelling him to hit it). In this ordering-source-maximal world he does not hit the bullseye, so "Groucho cannot but not hit the bullseye." is true. Under a plausible semantics for *cannot but*, neither of the above sentences should be true.

This problem for a universal semantics of ability *can* is closely related to another: because of the universal quantifier in (16), this semantics for ability *can* does not observe property (7). That is, under this semantics, for any action ϕ, if an agent has the ability to ϕ, then they do not have the ability to not ϕ, and vice versa.

Both of these shortcomings of the universal account stem from the failure to model intention. That is, the universal account predicts that if it is undetermined whether an agent does ϕ or does not do ϕ, as a result of the agent having the *choice* to do either, then the agent does not have the ability to ϕ.

To avoid this unwanted conclusion, it should be a prerequisite for demonstrating that an agent lacks an ability to ϕ that an agent *tries* to ϕ, or at least, takes some action in order to effect ϕ.

To capture the notion of *trying* in (16), the obvious approach would be to include "The agent tries to ϕ." in the test conditions for "The agent can ϕ". The problem here is that not every proposition in the test conditions for the ordering source will always be satisfied in the maximal worlds. This is a simple consequence of the way that the ordering source is defined: if the test conditions are incompatible, the maximal worlds will not satisfy them all.

3.4 Attempt 3: A Conditional Analysis for Ability *Can*

The natural way to improve (16) so that it models "trying" is to make ability explicitly conditional. Thus, "A can ϕ" is true iff "A does ϕ if A tries to ϕ" is true. Following (Mandelkern, Schultheis, and Boylan 2016), we can formalize this with a selection function. The idea is that the selection function takes a proposition P and a world w, and returns the closest world to w in which P is true[8]:

(19) $[\![\text{A can } \phi]\!]^{<w,t>,C}$ iff $\exists t' > t : [\![\text{A does } \phi]\!]^{<f(A-tries-to-\phi,w),t'>,C}$

The simplest version of a conditional account, however, faces a problem with compulsion modals, as detailed by (Mandelkern, Schultheis, and Boylan 2016). In short, the dual of (19) is such that "A cannot but ϕ." means that if A tries to not ϕ, A will ϕ. This is not quite right for the semantics.

To see this, suppose that one night, Groucho only falls asleep when he tries to stay awake. By (19), it follows that he was compelled to fall asleep. However, it also happens that if he had tried to sleep, he would have not slept. Given this fact, to say that Groucho couldn't but sleep seems too strong a claim. What this example suggests is that a semantics for compulsion should require that for Groucho to be compelled to sleep, for *each* action Groucho might perform, he would fall asleep.

The natural modification to the conditional analysis is to explicitly model the different actions an agent could hypothetically take, in order to avoid the above problem with compulsion modals. This leads to the act-conditional analysis of (Mandelkern, Schultheis, and Boylan 2016), which we discuss next.

[8]The selection function performs a similar role to the ordering source in the Kratzerian architecture. However, it returns a single world.

3.5 Attempt 4: Act-Conditional Analysis for Ability *Can*

In response to the problems outlined above, (Mandelkern, Schultheis, and Boylan 2016) proposes an "act-conditional" semantics. It works the same way as the conditional analysis, but quantifies existentially over a set of *available actions*[9]. The type of actions can be understood as ($\mathrm{A} \rightarrow \langle W, T\rangle \rightarrow \mathrm{Bool}$), where A is the type of agents, $\langle W, T\rangle$ is the type of world-time pairs, and Bool is the type of truth values. In other words, an action maps agents to truth values at world-time pairs.

(20) $[\![\text{A can } \phi]\!]^{<w,t>,C}$ iff $\exists \psi \in \mathrm{ACT}_{(A,<w,t>,C)} : \exists t' > t \; [\![\text{A does } \phi]\!]^{<f(A-tries-to-\psi,w),t'>,C)}$

ACT denotes the set of actions available to an agent at a world w, time t and in a context C. The benefit of this analysis is that it directly models the notion of choice: the actions available to an agent represent to choices they can make. An agent has an ability to ϕ if there exists an available action ψ such that in the closest world where they try to perform ψ, they do ϕ.

What constitutes an available action depends on context. One case where this is particularly clear is (21)[10], where Groucho has been invited to dinner by his neighbor but has to decline on account of prior commitments:

(21) Groucho: I am not able to come to dinner.

(22) Karl: Yes you are, you just would prefer to do something else!

Groucho can only be considered to be speaking truthfully if walking over the Karl's house at the time he was invited is not an available action. Karl's point is that this *was* an available action, but it seems reasonable to have a semantics which allows it either to be counted or not; this way, we can say that the controversy between Groucho and Karl over the truth of (21) comes down to what each considers to be the set of available actions.

Benefits of the Act-Conditional Analysis As noted by (Mandelkern, Schultheis, and Boylan 2016), the act-conditional analysis addresses the problem of compulsion modals, since the dual of (20) is as follows:

(23) $[\![\text{A cannot but } \phi]\!]^{<w,t>,C}$ iff $\forall \psi \in \mathrm{ACT}_{(A,<w,t>,C)} : \exists t' > t \; [\![\text{A does } \phi]\!]^{<f(A-tries-to-\psi,w),t'>,C)}$

[9]Note that the version of the act-conditional analysis I present here differs in certain details from the formulation provided in (Mandelkern, Schultheis, and Boylan 2016); in particular, I model times.

[10]This example is adapted from (Mandelkern, Schultheis, and Boylan 2016)

Informally, an agent is compelled to do ϕ iff for all available actions ψ, if they do ψ, they do ϕ. This seems intuitively convincing: an agent is compelled to do ϕ iff no matter which available action the agent chooses, they do ϕ. Having a suitable semantics for compulsion which is dual to ability, as in (23), is the result of the single world selection function employed by (Mandelkern, Schultheis, and Boylan 2016).

Despite these advantages, there are number of shortcomings of (20), which prompt the need for a new proposal for ability *can* (see section (4)).

3.6 The Shortcomings of the Act-Conditional Analysis

First Shortcoming The act-conditional analysis makes use of a selection function which returns a single closest world. That this is a single world is crucial to the analysis, as the authors themselves admit[11]. This results in a problem for their semantics, relating to property (5).

The problem with this single world function arises in cases where an ability claim should come out false, as a result of an *undetermined outcome*, i.e. cases relating to property (5). To see this, recall the case of the lottery with 100 possible outcomes. (15) should be false, but on (20), the selection function will have to pick some world as the closest. Without loss of generality, let's say this is w30, the world in which the 30 is the number that wins the lottery. This means that if buying the ticket with the number 30 is an available action, (15) will come out true on an ability reading.

To avoid this, (Mandelkern, Schultheis, and Boylan 2016) are forced to argue that picking 30 (or whichever number corresponds to the world that the selection function selects, *mutatis mutandis*) is not an available action in the context C, a move which makes the already vague notion of "available action" yet more slippery.

By contrast, if the lottery is rigged, and Groucho knows that 30 will be the winning number no matter what, we would want (15) to be true. As a result, in *this* context, buying the ticket with the number 30 *should* be an available action.

The issue of the problem here is that non-determinism is not modeled by the Stalnackerian single-world selection function. As a result, the account of (Mandelkern, Schultheis, and Boylan 2016) fails to express the situation in which, after performing or trying to perform some action ψ, what will go on to happen is still open.

Thus, for the account of (Mandelkern, Schultheis, and Boylan 2016), (5) does not hold unless the set of

[11] "Note that the plausibility of our predictions here stems in part from our choice to use the selection function from Stalnaker (1968), which selects a single world..."

available actions omits what seem to be reasonable actions (e.g. purchasing a given lottery ticket in a lottery).

A Second Shortcoming The act-conditional analysis does not account for (6). This is the property of ability that if an agent fails to ϕ in fair conditions, the failure is evidence that they lack the ability to ϕ.

The act-conditional analysis does not relate this failure, i.e. the existence of a historically equivalent world in which an agent does not ϕ, to the falsity of the agent possessing the ability to ϕ, as defined by the act-conditional analysis. The problem is that the act-conditional analysis relies on a selection function. Even if we stipulate that the selection function must pick a historically equivalent world to the present one, a failure in the actual future to perform ϕ (e.g. to hit the bullseye in a game of darts), does not provide evidence for the lack of ability to do so.

Third Shortcoming Finally, the act-conditional analysis cannot deal with the property of ability *can* described in (11), that a series of actions may be required to perform an ability.

Consider the case of a game of chess, raised in section (1): at a certain point in a game of chess against Karl, Groucho has secured a winning position. An observer then utters (24):

(24) Groucho can win this game of chess.

Supposing that the actions available to Groucho correspond to the legal moves of chess[12], the act-conditional analysis seems too weak. It is not the case that there is any single move that leads to Groucho's victory. Instead, he must make a move, after which his opponent moves, and so on, until he has won. The key point is that (24) should be true just in case he has an answer at each successive stage to whatever move Karl makes, that eventually leads to victory.

4 A New Proposal for Ability *Can*: The Game-Semantic Analysis

In response to the shortcomings of the act-conditional analysis, I propose two changes, resulting in the *game-semantic analysis*, as follows:

[12] One foreseeable response is to question whether this is the correct choice of a set of actions to assign to Groucho at this point in time. If, instead, one of Groucho's actions were to play to the best of his abilities, then it would only take this action for him to win the game.

(25) $[\![\text{A can } \phi]\!]^{<w,t>,C}$ iff $\exists\psi \in \text{Act}_{(A,<w,t>,C)}$: $\forall \text{w}' \in \text{O}_{TEST^{<w,t>,A,\phi,C}}(\{\text{h} \in \text{HistEq}_{<w,t>} | [\![\text{A tries to } \psi]\!]^{(<h,t>,C)}\})$, $\exists t' > t$: $([\![\text{ A does } \phi]\!]^{<w',t'>,C} \vee [\![\text{ A can } \phi]\!]^{<w',t'>,C})$

The above definition is recursive, in the sense that the ability to do ϕ is defined partial in terms of the ability to do ϕ. The purpose of this recursion is to address property (11): an agent may have to perform a series of actions of unbounded length in order to execute a given ability, and between each action, the nondeterminism of the world means that a number of things can happen.

Intuitively, (25) can be understood as follows: to evaluate a claim of ability to ϕ at $< w, t >$ for an agent A, we first consider whether, if A tries to ψ at $< w, t >$, A does ϕ in every continuation of the future which satisfies the test conditions for ϕ (i.e. the maximal worlds under the ordering source.). If A does ϕ in every one of these continuations of the future, we say that A has the ability. If *not*, we then consider what happens if the agent performs a *subsequent* action to ψ, say ψ', and repeat the procedure. This iterative procedure is expressed recursively in (25).

An agent A tries to perform an action ψ. Afterwards, some number of continuations of the future are possible, namely the futures of the historically equivalent worlds which are maximal in the ordering source. In all of these continuations, the agent has either succeeded in performing ϕ or continues by performing another action ψ'. This procedure continues indefinitely. If, no matter which

As with (16), there is a set of test conditions which parametrizes the ordering source. What these test conditions are is determined by the world and time, the context C, the action ψ and the agent A. The idea is that, as in (16), they should be a set of criteria defining the conditions which need to be met in order for a test to be fair. For instance, one condition, in the case of Groucho's ability to hit the bullseye, will be that he is not physically hampered from making his throw.

The name of the game-semantic analysis is inspired by the resemblance of (25) to the definition of a winning strategy in a two player game. A has a winning strategy if there exists a move M available to A such that for every move N available to B after M is made, either A wins after N or has a winning strategy. This analogy to a game maps an agent to a player of a game, and their opponent to the world; an agent has an ability to ϕ if they have a winning strategy to accomplish ϕ[13].

I now review the two main changes distinguishing the game-semantic from the act-conditional analysis, and explain their motivation and benefit.

[13]See the concept of games against nature, e.g. Harsanyi 2004

The first innovation The first change in (25) from the act-conditional analysis is the presence of a *set* of closest worlds in which the agent tries to ψ, rather than a single closest such world. This set is a quantified over universally.

This change has two advantages. Firstly, consider (5) and the problem of the lottery, in which buying the winning ticket was not an available action for the act-conditional analysis:

(26) Groucho can win the lottery.

Suppose the winning number, at time t', is 30 and that the lottery is not rigged in any way. Under the act-conditional analysis, in order for (26) to be false, the action of purchasing the ticket with the number 30 cannot be in the set of available actions. This is no longer a problem: purchasing any ticket can now count as an available action for Groucho, with (26) remaining false.

Suppose, for instance, that Groucho takes the action of purchasing ticket 30. It is not the case that in every one of the closest worlds, he is now able to win. So this action (and indeed all actions ψ) do not lead to Groucho winning in *every* fair world in which he tries to ψ. This means that we obtain the wanted result, that (26) is false, while still having the purchasing of any particular ticket be an available action to Groucho.

The second advantage of the universal quantification over historically equivalent worlds is that we now have some connection between a failure to perform an action and a lack of ability (i.e. the property outlined by (6)): in particular, under certain circumstances, the failure entails the lack of ability. Consider the case of Groucho trying and failing to hit the bullseye. Suppose that in the actual world w, nothing has intervened to make the test conditions unfair - Groucho has just missed. In such a case, the world w offers itself as a witness against the claim that in *every* world in which Groucho performed ψ (let us say that ψ is the action of throwing the dart), he either hit the bullseye or had the ability to go on to do so.

Note, however, that it is still possible that had Groucho performed a different action to ψ, he would have succeeded. So the failure of Groucho to hit the bullseye only is grounds to suggest that he lacked the ability to hit it *if* we believe that no other action ξ would have lead to his success.

The universal quantification, as a result of the test condition ordering source, also handles cases where in the actual world, Groucho is not presented with a fair test of his ability. Worlds which aren't fair in this sense will not appear in the domain of quantification, since they are not maximal in the ordering source, and as such, will not witness the falsity of a given ability claim.

This explains why "Groucho can hit the bullseye." can be judged to have been uttered truthfully even if Groucho goes on to miss the bullseye as a result of someone pushing him. In this case, the real world doesn't provide fair test conditions, and thus isn't an element of the domain of quantification of ability *can.*

A third advantage of the universal quantification over actions is an interesting consequence for compulsion modals. Dualizing (25), we obtain (27):

(27) $[\![$A cannot but $\phi]\!]^{<w,t>,C}$ iff $\forall\psi \in \text{Act}_{(A,<w,t>,C)}$: $\exists$w$' \in \text{O}_{TEST^{<w,t>,A,\phi,C}}(\{\text{h} \in \text{HistEq}_{<w,t>} |\ [\![\text{A tries to } \psi]\!]^{(<h,t>,C)}\ \})$, $\exists t' > t$: $([\![$ A does $\phi]\!]^{<w',t'>,C} \vee [\![$ A cannot but $\phi]\!]^{<w',t'>,C})$

Informally, this means that an agent is compelled to ϕ iff there is nothing an agent can do to rule out the possibility of ϕ. If this seems too weak, note that compulsion is defeasible. If Groucho has to sneeze at time t, and he is in a car crash at time t$'$ resulting in a concussion before he sneezes, his failure to sneeze doesn't mean that it was false that he had to sneeze[14].

The second innovation The game-semantic analysis accounts not only for properties (5)-(7) of ability, but also (11). This is achieved through the recursivity of (25) (i.e. the fact that under (25), ability is defined partially in terms of itself, in the final disjunction of the definition), which allows us to model cases in which an agent must perform a series of actions in order to achieve some result.

This is important if no available action is sufficient to achieve a given outcome, but instead, one action makes another available, and so on until the desired outcome is reached. Recalling the case of (24), repeated as (28), we can see that the game-semantic analysis can handle the need for just such a sequence of actions:

(28) Groucho can win this game of chess.

Suppose that the available actions are a set of legal moves that Groucho is aware of at this point in the game. To say that Groucho can win is to say that there is either a move which will win him the game, or a move after which it is the case that he can win (by playing a move after which he wins or it is the case that he can win, and so on...). Note, however, that for (28) to be true, it need not be the case that a single move wins Groucho the game.

For (27), the recursive definition allows us to consider complex cases of compulsion, where an agent tries to avoid an outcome through a series of actions, but fails. A famous case of this are Greek myths, where

[14](Davis, Matthewson, and Rullmann 2009) tests this particular case empirically in their fieldwork on Salish compulsion/ability modal circumfix *ka'...a* and finds that in this language, at least, speakers do share this judgment of defeasibility

no matter what an agent tries in order to avoid ϕ, a (usually improbable) series of events leads to ϕ anyway.

4.1 Ability *Could Have*

In section (1), we proposed that ability *could have* is a reading that needs to be modeled, and can be treated as past-tense ability *can*. The idea is that A could have ϕ iff at a past point in time, "A can ϕ" was true (for ability *could have* and *can*).

However, there is an important caveat: if at time t we make an ability claim of the form "A could have done ϕ.", then this claim can only be true *either* if the agent has done ϕ by t, *or* if some circumstance intervened between t$'$ and t such that the agent A did not have a fair test of their ability.

Furthermore, as noted by (Condoravdi 2001), "A *could have* done ϕ" carries that implication that A did not ϕ. This means that, if this implication goes through, the second disjunct above, that some circumstance intervened between t$'$ and t, has to hold in order for "A could have done ϕ." to be true.

To make this point clearer, consider (29)

(29) Groucho could have hit the bullseye.

Suppose Groucho had the ability to hit the bullseye in the past, and nothing prevented him from doing so in the actual world (i.e. the world from which we are backtracking into the past): it follows that he would have gone on to hit the bullseye. However, since *could have* bears the implication that Groucho has not hit the target,(29) is infelicitous on the ability reading of *could have* - after all, if he had the ability and nothing prevented him, why didn't he the bullseye?

This means that the only situations in which there is not an implication of falsity for statements of the form "A could have ϕ." (on an ability reading of *could have*) are those where the ability was *not* fairly tested. These are the cases where it makes sense to use ability *could have*; the case of (30) is one such example:

(30) Caesar could have won the war against the Parthians.

The fact that Caesar was assassinated before embarking on this war means that his ability to defeat the Parthians was never fairly tested. For this reason, the fact that he did not defeat the Parthians is not indicative of his inability to have done so.

So, in summary, ability *could have* only appears in cases where the world in question does not provide optimal test conditions for the given ability.

5 Concluding Remarks

Since this investigation of the semantics of *could-have*$_{able}$ has spanned a number of complex issues, it is worth concluding with a few key takeaways. Previous accounts of ability *can* will fail on one or another of the five properties described in (5)-(11). The game-semantic analysis satisfies all five, and presents a model of ability which can be intuitively summarized as follows: an agent can ϕ iff they have a sequence of actions that results in their doing ϕ, no matter what happens after each action, provided that what happens after each action is a reasonable test of ϕ.

References

Brown, Mark A (1988). "On the logic of ability". In: *Journal of philosophical logic* 17.1, pp. 1–26.

Carlson, Gregory N (1995). *The generic book*. University of Chicago Press.

Condoravdi, Cleo (2001). "Temporal Interpretation of Modals - Modals for the Present and for the Past". In: *The Construction of Meaning*. CSLI Publications, pp. 59–88.

Cross, Charles B (1986). "Can and the logic of ability". In: *Philosophical Studies* 50.1, pp. 53–64.

Davis, Henry, Lisa Matthewson, and Hotze Rullmann (2009). "Out of controlmarking as circumstantial modality in Státimcets". In: *Cross-linguistic semantics of tense, aspect, and modality* 148, p. 205.

Harsanyi, John C (2004). "Games with incomplete information played by Bayesian players, i–iii: part i. the basic model&". In: *Management science* 50.12_supplement, pp. 1804–1817.

Hintikka, Jaakko and Gabriel Sandu (1997). "Game-Theoretical Semantics-Chapter 6". In:

Mandelkern, Matthew, Ginger Schultheis, and David Boylan (2016). "Agentive Modals". In:

Thomason, Richmond (2005). "Ability, action and context". In: *Ms. University of Michigan*.