NUEL BELNAP
MICHAEL PERLOFF

The Way of the Agent

Abstract. The conditional, *if an agent did something, then the agent could have done otherwise*, is analyzed using *stit* theory, which is a logic of "seeing to it that" based on agents making choices in the context of branching time. The truth of the conditional is found to be a subtle matter that depends on how it is interpreted (e.g., on what "otherwise" refers to, and on the difference between "could" and "might") and also on whether or not there are "busy choosers" that can make infinitely many choices in a finite span of time.

1 Introduction

The complexities of the connections among actions, moral responsibility, and the alternatives open to the agent have long tormented philosophers.[1] Hume, for example, the most famous of all compatibilists, claims that universal determinism is not only consistent with human freedom but necessary for morality.[2] His view is that while agents are able to choose among alternatives, there is the liberty of voluntary action. We think that is right. On the other hand, Hume asserts, morality is without foundation if actions are not fully determined. We think that is wrong. Our intent, however, is not to joust with Hume, but rather to recount a tale that begins with the following proposition, one that might have puzzled even such as Don Quixote of La Mancha and his squire, Sancho Panza.

- If an agent is morally responsible for doing something, then the agent could have done otherwise.[3]

In an effort to unravel the complexities of the displayed proposition, we cleave it in twain, each conjunct seeming essential to its meaning.

- If an agent is morally responsible for doing something, then the agent did it.

- If an agent did something, then the agent could have done otherwise.

The middle term, *doing something*, is thereby revealed. We vouchsafe that for the route from moral responsibility to "could have done otherwise"

[1] We are indebted to the referee for many improvements.

[2] *An Enquiry Concerning Human Knowledge*, Section VIII, Part I.

[3] Modern replies to Moore [13] by Austin [1] and to Frankfurt [9] by Van Inwagen [19] put the issue in just these terms.

Studia Logica **51**: 463–484, 1992.

to be accurately charted, heed must be paid this idea, formerly hidden, that stands at the crossroad. We shall not in this paper further consider the claim of the first conjunct;[4] instead we press attention on the morality–free claim of the second conjunct that doing something implies having been able to do otherwise. There has been little effort to clarify the second conjunct *in isolation from moral considerations*; we deem worthwhile the enterprise of examining the relation between agentive doings and what it means to say that "an agent could, or might, have done otherwise." Hume's own account, "if we choose to remain at rest, we may; if we choose to move, we also may," seems, for example, to suggest a tie between "could" and "might": an agent could have done otherwise just in case there is something else he or she might have done. Is that right? Our view is that the armory of ordinary language is inadequate to the task of deciding such questions. We need the weaponry provided by a theory.

2 Review

Accordingly, we bend our bow with basic proposals of what we call "*stit* theo-ry".[5] The theory begins with the *stit* sentence, a sentence of the form $[\alpha \textit{ stit: } Q]$. We propose this form as an approximation to "α sees to it that Q", with the meaning that Q is true entirely because of a prior choice of the agent α. Among its theses are the following. The *Stit paraphrase thesis*: Q is agentive for α whenever Q is properly paraphrased as $[\alpha \textit{ stit: } Q]$. The *Stit complement thesis*: $[\alpha \textit{ stit: } Q]$ is grammatical and meaningful for any sentence Q. The *Restricted complement thesis*: A variety of constructions concerned with agents and agency — including deontic, imperatival, and others — must take agentives as *their* complements. These prior theses gain much of their force from the *Stit normal form thesis*: In constructions that take agentives as complements, the wiser course is to represent the complements therein as *stit* sentences. Of a more specialized character is the *Stit analysis of refraining*: Where α is the subject of Q, to refrain from Q is $[\alpha \textit{ stit: } \sim Q]$, and to refrain from seeing to Q is $[\alpha \textit{ stit: } \sim [\alpha \textit{ stit: } Q]]$.

[4] But any reader of this paper will note at least the following: there is a deep difference between (a) the first conjunct as here displayed and (b) its sound–alike, "if an agent is morally responsible for an outcome, then the agent saw to that outcome." Also see Bartha [2] for an application of the logic of agency described below to certain problems of deontic logic.

[5] We have elsewhere described various parts of this theory: Belnap and Perloff [6] (general introduction), Belnap [5] (some context), [3] (history and pictures), [4] (more formal development) and Perloff [14] (some comparisons). Since these publications spell out and motivate the theses of this paragraph and the conditions of the next, we provide only a brief restatement here, just for reference.

We evaluate a *stit* sentence with the help of a system of agentive choices in a temporal structure having multiple branches open to the future but only a single route to the past. More precisely, we posit a structure $\langle Tree, \leq, Instant, Agent, choice\rangle$ as follows. (1) *Partial order.* *Tree*, whose members are dubbed "moments," is partially ordered by $\leq$. Let m and w be moments and let M be a set of moments. (2) *No backward branching.* Incomparable moments never share an upper bound. (3) *Historical connection.* A *history* is a maximal chain of moments. Every pair of histories intersects, i.e., every two moments have a common lower bound. Let h be a history, and let H be a set of histories. (4) *Instant* is a partition of *Tree*. The members of *Instant* are dubbed "*instants*." Members of the same instant are dubbed "*co–instantial.*" Let i be an instant, and let $i_{(m)}$ be the set of moments co–instantial with m. (5) Each instant intersects each history in a unique moment. (6) *Instant* never skews historical order. (7) *Agent* is a nonempty set whose members are dubbed "agents." Let α be an agent. (8) *Choice* is a function defined on agents and moments, yielding as value for α and w a partition of all the histories passing through w. A member of the partition is called a *possible choice for* α *at* w. The choice function is subject to the following constraints. (9) *Something happens.* For each w, for every way of selecting one possible choice for each agent at w, the intersection of the selected choices contains at least one history. (10) *No choice between undivided histories.* Two histories are said to be *undivided* at a moment if the two histories share a properly later moment. Agents cannot choose between undivided histories; that is, no possible choice for an agent at w separates histories that are undivided at w.

Next we review certain defined concepts, including some presented semantically.

- *Choice equivalence.* We provision ourself with three set–theoretical representations of the fact that even though there are many histories, there are all too few choices. (i) $H^w_\alpha(h)$, which is the possible choice for α at w to which h belongs (i.e., the member of the partition delivered by *choice* for the arguments α and w to which h belongs), is fundamental. When $w < m$, (ii) we let $H^w_\alpha(m) = H^w_\alpha(h)$, where h is any arbitrary history through m, and (iii) we let $M^w_\alpha(m)$ be the set of moments at which histories in $H^w_\alpha(m)$ intersect $i_{(m)}$. $M^w_\alpha(m)$ is thus the set of moments that are "*choice equivalent to* m *for* α *at* w."

- *Truth and settled truth.* We adopt from Prior via Thomason [15] the fundamental principle that it is impossible to make sense of branching

time unless one sees truth as relative to moment/history pairs m/h with $m \in h$. Also from Thomason comes the notion that Q is *settled true (false)* at m (rather than "at m/h") just in case Q is true (false) at m/h for all h to which m belongs (or, pictorially, which pass through m). In fact Thomason's own words for expressing this idea do not include the modifier "settled." We adopt his policy in part. To make our usage clear, we first say that Q is *moment–determinate* if for each m, Q is either settled true or settled false at m. Our usage is this: we allow ourselves to say that Q is true (false) at m, dropping "settled," only in the special case when you can see by easy inspection that Q is moment–determinate. Otherwise, we think we avoid confusion by either explicitly inserting "settled" or by explicitly mentioning a history as well as a moment. In a similarly motivated departure from Thomason, by "implication" we mean preservation of truth (not just settled truth) at all moment/history pairs. But note that implication between moment–determinate sentences can be simplified to just truth–preservation at all moments, and we sometimes rely on this equivalence.[6]

- *The horizon from w at i* is the set of moments lying on the intersection of i and some history through w.

- $[\alpha$ *stit:* $Q]$ is true at m/h just in case there is a choice point w — we call w a "*witness*" — satisfying the following conditions.

  Priority condition: $w < m$.

  Positive condition: Q must be settled true at each member of $M^w_\alpha(m)$, i.e., at each moment that is choice equivalent to m for α at w.

  Negative condition: There must be at least one member of the horizon from w at i at which Q is not settled true.

The Negative condition says that when $[\alpha$ *stit:* $Q]$ is true at m/h, α really did make a choice with respect to Q (the falsity of Q was risked). The Positive condition says that the agent's prior choice guarantees the truth of Q.

[6] We and others have tried conflicting expository policies for reducing the confusion between truth and settled truth, and we can only hope that we have maintained a large measure of consistency within the confines of this paper.

$[\alpha$ *stit:* $Q]$ is evidently moment–determinate. We shall therefore feel free either to include or to drop "settled" in "$[\alpha$ *stit:* $Q]$ is true (false) at m."

- *Busy Choosers and witness by chains.* The preceding semantic narrative suffices if the agent is not a *Busy Chooser*; if, that is, there is no pair of moments between which the agent has an infinite number of nonvacuous choices. We do not know if there are any Busy Choosers, but if there are any, then the semantics for *stit* must be complicated to admit *witness by chains* as follows: $[\alpha$ *stit:* $Q]$ is true at m/h just in case there is a wholly prior nonempty chain of moments, c, dubbed a *witness*, that satisfies a version of the Positive and Negative conditions. The *Positive condition* for witness by chains states that Q must be settled true at every moment in $i_{(m)}$ that is choice equivalent to m at c for α. The latter phrase is defined as follows (assuming m is later than every moment in c): m_1 is choice equivalent to m at c for α provided they are co–instantial, provided m_1 is after some moment in c, and provided that for every $w \in c$ that is a common lower bound for m and m_1, $H^w_\alpha(m) = H^w_\alpha(m_1)$. In other words, no choice that α makes in the course of c distinguishes m_1 from m. The *Negative condition* for witness by chains says that the falsehood of Q is risked throughout the chain: for every w in c there is a later moment co–instantial with m at which Q is not settled true.

- $[\alpha$ *dstit:* $Q]$. *Stit* is not the only useful approximation to "seeing to it that." The following concept, in some respects simpler because the moment of witness is identified with the moment of evaluation of the complement, is due independently to von Kutschera and Horty (see notes 12 and 17 of Belnap [3]). With homage to Thomason [16], we sometimes call it the *deliberative stit*, which accounts for the notation (the "*d*" is silent). $[\alpha$ *dstit:* $Q]$ is true at a moment/history pair w/h_1 just in case (*Positive condition*) Q is true at w/h_2 for every $h_2 \in H^w_\alpha(h_1)$, and (*Negative condition*) Q is not settled true at w. Evidently *dstit* sentences are *not* in general moment–determinate. In fact, although they can be settled false at m, the combination of the Positive and Negative conditions forbids that they ever be settled true. One has to keep track of this crucial difference between *stit* and *dstit* when considering examples.

Further Concepts

In this section we outfit ourselves with further concepts needed to complete this inquiry, most of which adapt tense and modal notions to the theory of branching time.[7] Because we shall be involved more in analysis than in calculation, we adopt helpful English words as our symbols.

- *Will:Q* (*Was:Q*) is true at m/h iff Q is true at some m_1/h, with $m_1 \in h$ and with $m_1 > m$ ($m_1 < m$). Neither *Will:Q* nor *Was:Q* is moment–determinate.

- *Settled:Q* is true at m/h iff Q is settled true at m. *Settled:Q* is moment–determinate.

- *Was–always–inevitable:Q* is true at m/h iff Q is settled true throughout $i_{(m)}$. Historical connection guarantees that the English expression is helpful.[8] *Was–always–inevitable:Q* is moment–determinate.

- *Might–have–been:Q* is true at m/h iff there is some member of $i_{(m)}$ at which Q is not settled false. Since *Might–have–been:Q* $\leftrightarrow \sim$*Was–always–inevitable:*$\sim Q$, it is again Historical connection that warrants that our English is appropriate.[9] *Might–have–been:Q* is moment determinate.

We next represent the "all in" ability of Austin [1], present or absent on a particular occasion for a particular agent and with respect to a particular complement. Such ability statements can be tensed either as of the moment of witness or as of the moment of evaluation of the complement. We use "can" for the former and "could have" for the latter :

- $Can_i[\alpha$ *stit:* $Q]$ is true at w/h iff there is a moment, m, lying on the horizon from w at i such that w witnesses that $[\alpha$ *stit:* $Q]$ is true at m. $Can_i[\alpha$ *stit:* $Q]$ is moment–determinate.

---

[7] See Thomason [17], which includes access to the literature.

[8] Let *Was–always:Q* be true at m/h iff Q is true at every m_1/h with $m_1 \in h$ and with $m_1 < m$. Let *Inevitable–at–i:Q* be true at w/h iff Q is settled true at every moment in the horizon from w at i. Note that i is the time of evaluation of Q, not the moment of the inevitability. Then by Historical connection, *Was–always–inevitable:Q* is true at m/h just in case *Was–always:Inevitable–at–*$i_{(m)}$*:Q* is true there. (On the other hand, defining *Inevitable:Q* as *Settled:Will:Q*, you can see that *Was–always–inevitable:Q* is by no means equivalent to *Was–always:Inevitable:Q*.)

[9] Let *Might–be–at–i:Q* be defined as $\sim$*Inevitable–at–i:*$\sim$ Q, and let *Might–be:Q* be defined as $\sim$*Settled:*$\sim$*Will:Q*. In parallel with the preceding note, we observe (i) that *Might–have–been:Q* is true at m/h just in case *Was:Might–be–at–*$i_{(m)}$*:Q* is true there, and (ii) that it is not equivalent to *Was:Might–be:Q*.

- *Could-have*[α *stit:* Q] is true at m/h iff *Was:Can*$_{i_{(m)}}$[α *stit:* Q] is true there. *Could-have*[α *stit:* Q] is moment-determinate.

These definitions illustrate the Restricted complement thesis. Because [α *stit:* Q] occurs as a unit on the right sides of these definitions, you can see that [α *stit:* Q] is indeed a complement. You can also see that the right sides would make no sense if [α *stit:* Q] were replaced by an arbitrary sentence — which is just to say that the complement position is restricted.

4 Questions and Conjectures

Our preparations complete, we venture forth to battle a variety of questions in the topic of agency and "could have done otherwise." The assembled technical apparatus, though weighty, is neither more nor less than is needed to complete the task. On the one side we avoid the perils induced by the vagaries of ordinary language, for example the differences between "might" and "could" in this context, and the exact target of the anaphoric reference of "otherwise". With *stit* normal forms to guide us, we sharpen the contrasts and bring each topic into clear relief. On the other side, having rendered a question or conjecture in the language of *stit*, we employ *stit* theory to render formal judgment (an upshot). That, in turn, requires clarity about alternatives, about matters temporal and historical, and about possibilities in branching time.

4.1 Could have been and might have been

First a simple question with a simple answer.

QUESTION. Is what could have been the same as what might have been?

UPSHOT. We think there is no difference.

Branching time, however, encourages one to see that both are oft times context-dependent, since the question of *when* something was possible is generally in order but hard to deal with. For both "could have been" and "might have been" we use a (tensed but) rigorous context-independent reading that works well for appropriately idealized cases:

Stit VERSION. *Might-have-been:Q.*

We hope it is obvious that our suggestion is not empty: "might have been" and "could have been" are English, while "*Might-have-been:*" is formally characterized symbolism.

4.2 Could have done and might have done

Now a conjecture not quite so simple.

CONJECTURE. What an agent might have done is different from he or she could have done.

By "the agent might have done it" we intend an impersonal modality, expressed perchance less ambiguously (and less idiomatically) by "it might have been that the agent did it."

We confess it is easy to run together "could have done" and "might have done." As a first advance in distinguishing them, replace "do" by "*stit*," yielding "could have *stit*" vs. "might have *stit*." As a second advance, interpret "could have *stit*" as "it was (simple past) in the agent's power to *stit*," and interpret "might have *stit*" as "there was a way that things could have gone such that it would have been in the agent's power to *stit*."

These ordinary language statements remain, however, obscure, and discourage further advance. *Stit* theory suggests that the critical difference concerns the witness to the *stit*–ing. When we say that "the agent could have done," we demand that the witness to the *stit*–ing be in the past of the moment m, at which Q is evaluated. On the other hand, when we say that "the agent might have done," we allow the witness to be in the future of possibilities of a moment that is in the past of m. These differences suggest the following.

Stit VERSION. *Could-have*[α *stit:* Q] is not equivalent to *Might-have-been*:[α *stit:* Q].

UPSHOT. The conjecture is correct; *Could-have*[α *stit:* Q] implies *Might-have-been*:[α *stit:* Q], but not conversely. The semantics for the two locutions are different: although the truth at m_0 of each of *Could-have*[α *stit:* Q] and *Might-have-been*:[α *stit:* Q] requires the truth of [α *stit:* Q] at some moment co-instantial with m_0, the former alone requires that the witness be in the direct past of m_0. Here is a picture.

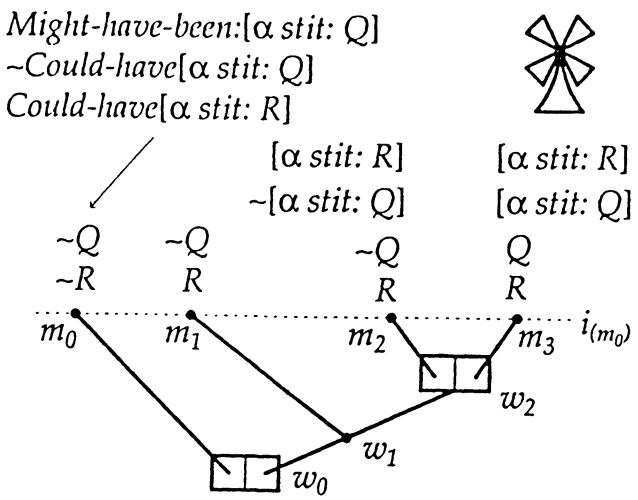

Fig. 1: Don Quixote attacks the windmill

The rectangles are choices for α, whereas what happens at w_1 is not up to the agent. The salient feature is that while $[\alpha$ *stit:* $Q]$ at m_3 has a witness, namely, w_2, that witness does not stand in the past of m_0. So at m_0 it is true that α might have seen to it that Q, but it is false that α could have seen to it that Q. In contrast, at m_0 it is true that α could have seen to it that R, since w_0 witnesses $[\alpha$ *stit:* $R]$ at each of $m_1 - m_3$.

EXAMPLE. Don Quixote attacks the windmill.

> Commending himself most devoutly to his lady, Dulcinea, whom he begged to help him in this peril, he covered himself with his buckler, couched his lance, charged at Rozinante's full gallop, and rammed the first mill in his way.

At the moment, w_0, that ends his commending, the Knight of the Mournful Countenance had the choice either to stand down or to ride on. Once having begun his charge, however, there was a slightly later moment w_1, at which Rozinante might by chance have collapsed. In the case of no collapse, there was then a later moment, w_2, at which Quixote had the choice either to swerve towards or to swerve away from the disastrous confrontation. Let m_3 be the moment at which he rammed the windmill. Consider the history on which, as he finished his commending, the Knight regained his wits and stood down, and follow that history to a moment, m_0, that is co–instantial with the moment, m_3, at which he rammed the windmill. At the moment,

m_0, it would be true to say that he might have attacked the windmill, but false to say that he could have. What decides the matter is that there is nothing he could have chosen at the end of his commending (w_0 — which is the only choice point in the past of the moment of non–attack under consideration) that would *guarantee* his attack. Both chance (at w_1) and the uncertainty of the outcome of a future choice (at w_2) stand in the way of such a guarantee.

4.3 Might have been otherwise

To appreciate this next conjecture, consider a *stit* with a *non*–agentive state of affairs as complement, and let the anaphor, "otherwise," refer to just that non–agentive complement rather than to the entire *stit* sentence.

CONJECTURE. If yon fellow sees to some state of affairs, then it might have been that the state of affairs not obtain — at that very instant.

The final phrase accomplishes a task more easily than idiomatic English: make sure that the "might" means that the absence of the state of affairs obtains as a co–instantial alternative to the very moment in question. In notation it is unambiguous;

Stit VERSION 1. $[\alpha$ *stit:* $Q]$ implies *Might–have–been:*$\sim Q$.

This is evidently equivalent to

Stit VERSION 2. $[\alpha$ *stit:* $Q]$ is inconsistent with *Was–always–inevitable:*Q.

The *stit* version says that if $[\alpha$ *stit:* $Q]$, then it has not been inevitable (determined) from all eternity that Q should obtain at the instant in question.

UPSHOT. The conjecture is, in its *stit* version, true. It is an obvious consequence of the Negative condition.

EXAMPLE. If it was inevitable from all eternity that the hog gelder's reed flageolet sounded four times while Don Quixote was at his meal, then the hog gelder did not see to it that his reed flageolet sounded four times while Don Quixote was at his meal. A hard determinist valiantly endorses the consequent; a soft determinist becomes cross, changes the topic, and exits the lists.

4.4 Might not have done it

Next an important conjecture with a straightforward disposition.

CONJECTURE. If α does something, then it might have been otherwise; i.e., α might not have done it.

Here let the "otherwise" refer anaphorically to the entire *stit* sentence, not just to its complement.

Stit VERSION. $[\alpha$ *stit:* $Q]$ implies *Might-have-been*:$\sim[\alpha$ *stit:* $Q]$.

UPSHOT. True.

We belabor the obvious by offering two proofs. First, since $[\alpha$ *stit:* $Q]$ implies Q (by the Positive condition), so that $\sim Q$ implies $\sim[\alpha$ *stit:* $Q]$, this is an immediate consequence of Upshot 4.3. The second and more important proof is this: *the consequent is a truth of logic*, so that the implication is vacuous! This is related to the Triponodo principle of Makinson [12], except here instead of the "trivial (legal) power not to do," we have the "Trivial possibility of not doing." The argument that it is logically true is an easy reductio. If $[\alpha$ *stit:* $Q]$ were true throughout an instant, then by the Positive condition, Q would be settled true throughout that same instant — which would leave no room for satisfaction of the Negative condition.

EXAMPLE. Consider the apparently agentive statement: Don Quixote "armed himself cap-a-pie, mounted Rozinante, placed his ill-constructed helmet on his head, braced on his buckler, grasped his lance, and through the door of his back yard sallied forth into the open country." If the statement is taken at face value as agentive, then it must possibly be false. For example, if one looks to the description of Don Quixote as "having lost his wits completely" in order to judge the agentive statement false (on the grounds that a man without wits cannot see to anything), one sees that its possible falsity is trivial. If, however, one uses the fact that Don Quixote's brains have dried up as an excuse to reinterpret the apparently agentive content of the statement as really non-agentive, perhaps something like a metaphor, so that Quixote mounted Rozinante as the storm mounts a distant hill, then the description is not agentive and we concede that *stit* theory has no strategem for probing the statement in question.[10]

[10] At this point we might have included a discussion of "might not have refrained from preventing," but we refrained, noting only that even though "refrain from preventing,"

4.5 Could not have avoided doing

This conjecture, inserted here because its disposition involves an application of the preceding result, is a proposal for a sufficient condition of doing.

CONJECTURE. "The fact that a person could not have avoided doing something is a sufficient condition of his having done it" (Frankfurt [9], p. 150).

This appears to be an instance of "necessity implies truth"; but analysis reveals that the conjecture is plausible and interesting only because it is ambiguous. Its status depends on whether "avoided doing it" means just "didn't do it," and so is non–agentive; or whether, agentively, it means "refrained from doing it," i.e., "saw to it that he or she didn't do it." This ambiguity is difficult to detect in ordinary English; but when it is revealed by normal forming with *stit*, either the plausibility or the interest of the conjecture disappears, as we see by considering the following two versions.

Stit VERSION 1. $\sim$*Might–have–been*:$\sim$[α *stit:* Q], i.e., *Was–always–inevitable*:[α *stit:* Q], implies [α *stit:* Q].

Stit VERSION 2. $\sim$*Could–have*[α *stit*:$\sim$[α *stit:* Q]] implies [α *stit:* Q].

UPSHOT. *Stit* version 1 is true, but *vacuously* so, since by the Trivial possibility of not doing principle of §4.4, its antecedent is a logical falsehood. *Stit* version 2 is evidently false; for a counterexample, choose Q as any tautology. It is then past doubt that the antecedent of version 2 is trivially true and the consequent trivially false. There is thus no reading of the conjecture on which it is both interesting and true.

EXAMPLE. Since the first version is an easy application of §4.4, we illustrate only *stit* version 2. On the side of the antecedent, it is evident that not even the great Mameluke of Persia, either before or after his nine–hundred–year enchantment, could have avoided (refrained from) seeing to it that if the golden helmet of Mambrino was made of brass, then it was made of brass; but on the side of the consequent, that dignitary certainly did not in fact see to that, nor to any tautology.

---

in the sense of [α *stit:* $\sim$[α *stit:* $\sim$ Q]], is a distinct agentive modality, "might not have refrained from preventing" does not create distinct analytical problems.

4.6 Could have prevented

The following conjecture is confusing in ordinary language but easy to settle correctly when expressed symbolically.

CONJECTURE. That we are responsible for some state of affairs implies that it must have been possible for us to have been responsible for its absence.

Stit VERSION. $[\alpha \textit{ stit: } Q]$ implies *Could-have*$[\alpha \textit{ stit: } \sim Q]$.

Examples of this conjecture sound plausible in English: it appears to follow from the fact that if Sancho Panza remained at rest beneath the cork tree, then he could have seen to it that he moved (Hume).

UPSHOT. But as all contemporary logicians of action know, the most elementary story tells us that the conjecture is false. In *stit* theory, the relevant point emerges through the Negative condition, which requires only that the falsity of Q be risked, not that it be guaranteed.

EXAMPLE. La Tolosa, the fair cobbler's daughter from Toledo, saw to it that Don Quixote was girded with his sword; but given the rough company of carriers, to say nothing of La Molinera, that poor wench was evidently in no position to see to it that the knight *failed* to be girded.

4.7 Could have refrained

A subtler question is the following :

QUESTION. If α saw to something, could α have refrained from seeing to it?

Some think yes. Chisholm [7], for instance, says that if some varlet loosed his firelock, then "there was a moment at which it was true, both that he could have fired the shot and also that he could have refrained from firing it." (Observe that Chisholm's "also" is *not* "could have seen to it that the shot was not fired"; he does not make the superficial mistake of supposing that Conjecture 4.6 is true.)

Some think no. Frankfurt [9] supposes it possible that there should be such a thing as "the fact that a person who has done something could not have done otherwise."

On our view this question is not to be happily represented without taking into consideration the *stit* analysis of refraining, so that we are not surprised

to find that on those few occasions that the literature notices the existence of the question, it resorts to sheer postulation.

Stit VERSION. Does $[\alpha$ *stit:* $Q]$ imply *Could-have*$[\alpha$ *stit:* $\sim [\alpha$ *stit:* $Q]]$?

We interpret the question as asking whether or not the fact that $[\alpha$ *stit:* $Q]$ is true at m implies that there is a moment *in the past of* m that stands as witness to $[\alpha$ *stit:* $\sim [\alpha$ *stit:* $Q]]$ at a moment co–instantial with m. That is, is there some single choice point in the past of m that, had a different choice been made, would have guaranteed the agent's failure to *stit* Q?

UPSHOT FOR *stit*. The implication fails, with an easy example, though not quite so easy as the counterexample to Conjecture 4.6. In the following picture, Q is settled true at m_0 and m_1, and settled false at m_2, all of which are co–instantial.

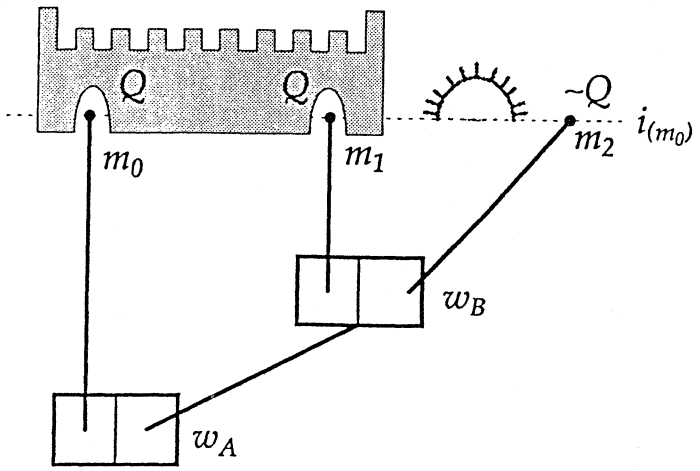

Fig. 2: Journey to the ducal castle

Each of w_A and w_B picture a choice for α; you can see from the diagram that $M_\alpha^{w_A}(m_0) = \{m_0\}$, so that w_A witnesses the truth of $[\alpha$ *stit:* $Q]$ at m_0 (looking to m_2 for satisfaction of the Negative condition). On the other hand, $M_\alpha^{w_A}(m_1) = \{m_1, m_2\}$, so that since $[\alpha$ *stit:* $Q]$ is true at m_1, w_A does not satisfy the Positive condition for witnessing that $[\alpha$ *stit:* $\sim [\alpha$ *stit:* $Q]]$ is true at m_2. The choice point w_A is "too soon." (The choice point, w_B, however, does that job.) Therefore, *Could-have*$[\alpha$ *stit:* $\sim [\alpha$ *stit:* $Q]]$ fails at m_0, and therewith the implication stated in the *stit* version.

EXAMPLE. We take some literary license in the following idealization, recounted in such a way that Figure 2 serves as both spatial map and a model of choices in branching time. One afternoon The Knight of the Lions and his squire, Sancho Panza, journeyed to the castle of the Duke and Duchess.[11] At a certain point A they chose[12] the north path, which (ideally) guaranteed their arrival at the castle by sunset. We suppose that their only other choice at point A was the northeast path, which itself split, after a few minutes, at point B. At point B they could have either elected the north path from B, which *also* would (ideally) have guaranteed their arrival by sunset, or they could have chosen the northeast path from B, which would have led them astray with no possibility of arrival by sunset.

Thus, when Don Quixote and Sancho actually arrived at the castle by sunset, there was no choice point in the past of their entrance to the fortress at which they could have guaranteed refraining from arriving by sunset. That is, no choice in the past of their arrival could have positively prevented them from choosing to arrive by sunset. Take notice that the moment of departure from B, at which indeed they could have chosen to refrain from arriving by sunset, is not in the past of the moment of their actual arrival. Heed also that the example is purely structural — the desires, belief, and intentions of the agents are irrelevant.

For this conjecture, it makes a difference whether one considers *stit* or *dstit*.

Dstit VERSION. Does $[\alpha \textit{ dstit: } Q]$ imply $\mathit{Can}[\alpha \textit{ dstit: } \sim [\alpha \textit{ dstit: } Q]]$?

UPSHOT FOR *dstit*. In contrast to the *stit* version, the implication holds.

We take $\mathit{Can}[\alpha \textit{ dstit: } Q]$ as $\sim \mathit{Settled}{:}\sim[\alpha \textit{ dstit: } Q]$, noting that the formal countenance of "can" for *dstit* can be less wrinkled that "can" with *stit*, because one need not worry about a double temporal reference. The implication then comes to this: $[\alpha \textit{ dstit: } Q]$ implies $\sim \mathit{Settled}{:}\sim[\alpha \textit{ dstit: } \sim [\alpha \textit{ dstit: } Q]]$.

[11] We follow Pellicer in identifying the unnamed Duke and Duchess with Don Carlos de Borja and Maria Luisa de Aragón, whose ducal descendant celebrated the third centenary of *Quixote* in Pedrola in 1905.

[12] Anyone familiar with Thomason [15] (this is, alas, by no means the universal class) will appreciate that the past tense requires a standpoint and *makes no sense without one.* In using the phrase "they chose," we place ourselves at some moment later than m_0. By No backward branching and No choice between undivided histories, it is then uniquely determined what "they chose" at w_A — relative to that standpoint.

PROOF. Grant $[\alpha$ *dstit:* $Q]$ at w/h. By the Negative condition, Q is false at w/h_1, some h_1 through w. But by the failure of the Positive condition, $[\alpha$ *dstit:* $Q]$ is false, and so $\sim[\alpha$ *dstit:* $Q]$ true, at w/h_2 for every $h_2 \in H^w_\alpha(h_1)$. Hence, the Positive condition for $[\alpha$ *dstit:* $\sim [\alpha$ *dstit:* $Q]]$ at w/h_1 is satisfied; and so also — in virtue of our original supposal — is the Negative condition. Since $[\alpha$ *dstit:* $\sim [\alpha$ *dstit:* $Q]]$ is true at w/h_1, it follows that $\sim$*Settled:*$\sim[\alpha$ *dstit:* $\sim [\alpha$ *dstit:* $Q]]$ is true at w/h.

EXAMPLE. We redescribe the journey to the castle in term of *dstit*, changing only the Q, which is now to the effect that (*Will*:Don Quixote and Sancho Panza arrive at the castle by sunset), which we evaluate at w_A and the history on which they choose to go north. On this construal the later moment w_B is irrelevant: at w_A, given the choice they did make, they *dstit* (*Will*:they arrive by sunset); also at w_A they could have chosen to *dstit* they did not *dstit* (*Will*:they arrive by sunset). Moment w_B is relevant in considering whether or not at w_A they can *dstit* (*Will*:they fail to arrive by sunset). They can't.

4.8 Might have refrained

We pray your close attention to a question whose answer depends on whether an infinite number of choices is made in a finite time (Busy Choosers).

QUESTION. Suppose that α sees to it that Q; does it follow that α might have refrained from seeing to it that Q in the sense that there is a co-instantial alternative at which α refrains from seeing to it that Q?

Stit VERSION. Does $[\alpha$ *stit:* $Q]$ imply *Might-have-been*:$[\alpha$ *stit:* $\sim [\alpha$ *stit:* $Q]]$?

The answer to this question depends on whether or not there are Busy Choosers.

UPSHOT WITHOUT BUSY CHOOSERS. If there are no Busy Choosers, the implication is valid.

PROOF. Grant $[\alpha$ *stit:* $Q]$ true at m_0/h_0 and let w_0 be the witness in question. Let m_1 be some moment in $i_{(m_0)}$ at which $[\alpha$ *stit:* $Q]$ is true, and which has the further feature that it "has a closest witness" in the sense that there is a witness, w_1, to $[\alpha$ *stit:* $Q]$ at m_1 such that between w_1 and $i_{(m_0)}$ there are no further witnesses to the truth of $[\alpha$ *stit:* $Q]$ at any moment in $i_{(m_0)}$. Because of no Busy Choosers, m_1 must exist.

By the Negative condition, there is a moment, $m_{1'}$, lying in $i_{(m_0)}$ and above w_1 at which Q is not settled true. We claim that w_1 is a witness to $[\alpha \textit{ stit: } \sim [\alpha \textit{ stit: } Q]]$ at $m_{1'}$. The Negative condition is easy: $[\alpha \textit{ stit: } Q]$ at m_1 is just what is required. Suppose, for reductio, that the Positive condition failed; i.e., suppose that $[\alpha \textit{ stit: } Q]$ were true at some moment $m_2 \in M_\alpha^{w_1}(m_{1'})$, with witness at, say, w_2. Where could w_2 be? Because both w_1 and w_2 precede m_2, by No backward branching, either $w_1 < w_2$ or $w_2 \leq w_1$. The first alternative is impossible, because w_1 is a closest witness. The second alternative is equally impossible, because then the Positive condition of w_2 witnessing $[\alpha \textit{ stit: } Q]$ at m_2 would conflict with the failure of Q to be settled true at $m_{1'}$. It cannot therefore be gainsaid that $[\alpha \textit{ stit: } \sim [\alpha \textit{ stit: } Q]]$ is true at $m_{1'}$, which is co–instantial alternative to m_0.

EXAMPLE WITHOUT BUSY CHOOSERS. The journey to the ducal couple's castle illustrates the subtle difference between "could have refrained" and "might have refrained". In that adventure the travelers might have refrained from arriving before sunset, though it was false that they could have refrained from doing so, since they were not Busy Choosers. And thus it is: the moment at which the wayfarers might have departed from point B is an excellent witness to the truth of $[\alpha \textit{ stit: } \sim [\alpha \textit{ stit: } Q]]$ at a moment co–instantial with the one in question, thus verifying "might have refrained." Since, however, that moment is not in the past of the moment of their actual arrival at the castle, it does not help verify "could have refrained."

UPSHOT WITH BUSY CHOOSERS. In the presence of Busy Choosers and witness by chains, the implication fails.[13]

PROOF. The following busy picture (Fig. 3) supplies a counterexample.

The facts are these. We are ultimately interested only in a certain moment, m_0, but we consider also (i) moments in $i_{(m_0)}$, which is represented by the dashed horizontal, and (ii) choice points of two sorts: those for α, represented by rectangles, and those for β, represented by circles. Each choice point for α is binary, with the right possible choice containing a single history on which Q is assigned settled true where it intersects $i_{(m_0)}$.

[13] See Xu [20] to obtain an understanding of the modal structure of *stit* theory when there are Busy Choosers.

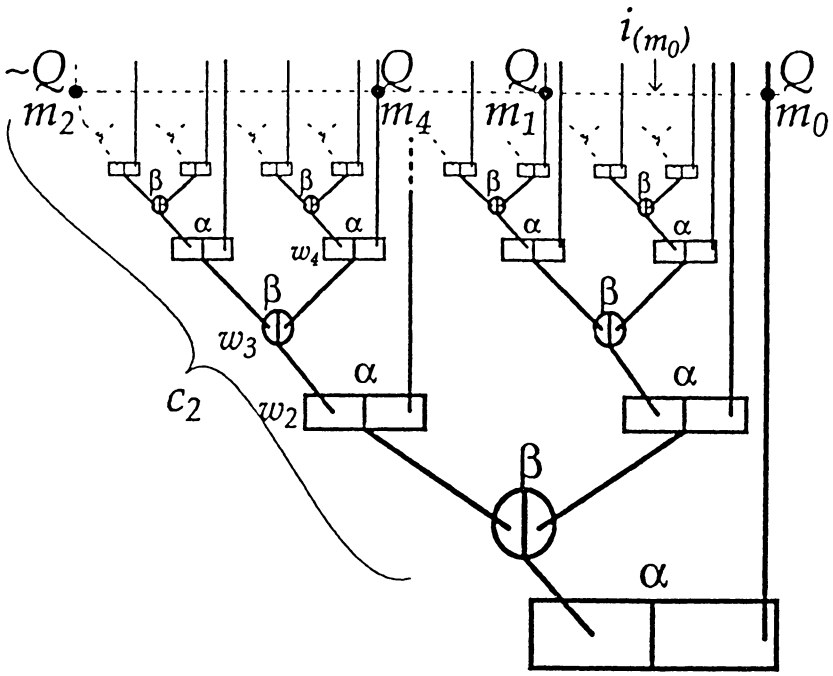

Fig. 3: The Mirror game

The left possible choice for α at the same choice point leads immediately to a choice point for β, at which there are two possible choices, left and right, each of which leads immediately to a choice point for α of exactly the same kind as before. We suppose the temporal distance between choices for α is halving, and that each entire denumerable historical series of choice points for α and β approaches a unique member of $i_{(m_0)}$; and we assign Q settled false at such members of $i_{(m_0)}$. The moment, m_0, is the one lying above the right side of the first choice for α.

We claim first that each choice point, w_1, for α witnesses the truth of $[\alpha$ *stit:* $Q]$ at the moment, m_1, in which the history belonging to the right hand possible choice for α at w_1 intersects $i_{(m_0)}$. The Positive condition is easy, since we assigned Q settled true at m_1 and since there is but a single history contained in that possible choice. The Negative condition is satisfied by our having assigned Q settled false at those members of $i_{(m_0)}$ approached by a denumerable historical series of choice points; one (and indeed many) of those members of $i_{(m_0)}$ must be properly later than w_1. As a special case, $[\alpha$ *stit:* $Q]$ is true at m_0.

We claim next that *nowhere* in $i_{(m_0)}$ is it true that $[\alpha$ *stit:* $\sim [\alpha$ *stit:* $Q]]$. This is obvious for the members of $i_{(m_0)}$, such as m_0 itself, that lie above some right hand possible choice for α. Now suppose for *reductio* that $[\alpha$ *stit:* $\sim [\alpha$ *stit:* $Q]]$ is true at some moment m_2 in $i_{(m_0)}$ that is approached by a denumerable series of choice points. There must then be a witness, and since we are allowing witness by chains, let the chain be c_2. The Positive condition implies that $\sim[\alpha$ *stit:* $Q]$ be settled true at every moment in $i_{(m_0)}$ that is choice equivalent to m_2 for α at c_2. Choose some member, w_2, of c_2. Properly between w_2 and m_2 there will be a choice point, w_3, for β, and properly after w_3 there will be a choice point, w_4, for α that is *not* in the past of m_2 ($w_4 \not< m_2$). Let m_4 be the member of $i_{(m_0)}$ lying above the right hand possible choice for α at w_4. The critical point is that m_4 is choice equivalent to m_2 for α at c_2, since — by the No choice between undivided histories principle — no choice for α in c_2 distinguishes m_4 from m_2. So, since $[\alpha$ *stit:* $Q]$ is certainly settled true at m_4, we have contradiction.

Thus $[\alpha$ *stit:* $Q]$ is true at m_0 but *Might-have-been:*$[\alpha$ *stit:* $\sim [\alpha$ *stit:* $Q]]$ is not, so that the implication fails.

EXAMPLE WITH BUSY CHOOSERS: THE MIRROR GAME.[14] The Knight of the Mournful Countenance and the Knight of the Mirrors engage in serious combat. They play a busy game that begins at noon and ends at sunset — at which time the vanquished is to remain entirely at the mercy of the victor.[15] Some plays of the game consist of infinitely many moves, which our knights-errant manage by halving the time spent on each successive move. Though busy, it is still a simple game. Don Quixote has the first move. On his turn the Manchegan has the following choice: press on or retire. If he retires then at sunset he is the vanquished. If he decides to press on, it is the turn of the Knight of the Mirrors, whose move always consists in selecting a phantasmical replica of one of two giants for the Manchegan to face: either Pandafilando of the Malignant Eye, or Briareus with many arms, each later phantasm being half as tall as its predecessor. Whatever he of the Mirrors selects, the next turn belongs again to the Knight of the Mournful Countenance. At sunset there are but two relevant possibilities: either Don Quixote has retired, in which case he is the vanquished, or he has succeeded in facing some denumerable sequence of phantasms, in which case he is the victor.

The curious fact to be illustrated is this: it is possible for Don Quixote to retire from the contest, but it is not possible for him to refrain from retiring.

[14] We thank S. Sterrett for supplying a basic idea of this game.

[15] Thomsen [18] reminds us that "there is nothing more serious than play" (p.171).

Contrary to our untutored intuitions, not even an entire chain of choices to press on, right up to sunset, can witness that Don Quixote refrains from retiring; for such a chain of choices does not establish that *it was entirely up to him* that he persevere. The choices that in fact were made by the Knight of the Mournful Countenance bestow no hard information about "what he would have chosen" had the Knight of the Mirrors confronted him with an unrelenting sequence of replicas of Pandafilando of the Malignant Eye. Quixote's famous victory, however, does allow him to wrest from the fallen Knight of the Mirrors the confession that "the torn and dirty shoe of Lady Dulcinea of El Toboso is better than the ill–combed though clean beard of Casildea."[16]

4.9 Summary

The proposition, *if an agent is morally responsible for doing something, then the agent could have done otherwise*, conceals connections among actions, moral responsibility, and alternatives open to an agent. We simplify by dividing the proposition in two: (1) *if an agent is morally responsible for doing something, then the agent did it*, and (2) *if an agent did something, then the agent could have done otherwise.* This division isolates the idea of *doing something*, validating the use of a logic of agency. We use *stit* theory to clarify proposition (2).

CONJECTURE. Proposition (2) can be disambiguated by means of a logic of agents who make choices against a background of branching time.

UPSHOT. The conjecture is true. If we interpret "could have done otherwise" as "might have been otherwise," the implication holds; if we interpret it as "might not have done it," the implication still holds, but vacuously. If we read it as "could have prevented," the implication fails. If "could have done otherwise" means "could have refrained" then it fails for *stit*, but holds for *dstit*. If "could have done otherwise" is taken as "might have refrained," then without Busy Choosers the implication holds, but with Busy Choosers it fails.

[16]Trust the concreteness of this fable, we beseech you, only to the extent that you understand its structural properties. Symmetrically, if you think our chronicle is wrong, please try to find an alternative rigorous account of witnessing, refraining, etc., and not just another picturesque story without a precisely described structure.

References

[1] J. L. AUSTIN, *Ifs and cans*, in **Philosophical Papers**, 2nd ed., J. O. Urmson and G. J. Warnock (eds), Oxford University Press, Oxford 1970, pp. 205 - 232.

[2] P. BARTHA, *Conditional obligation and the logic of agency.* Unpublished manuscript, University of Pittsburgh, Department of Philosophy, 1989.

[3] N. BELNAP, *Backwards and forwards in the modal logic of agency*, **Philosophy and phenomenological research** 51 (1991), pp. 777 - 807.

[4] N. BELNAP, *Before refraining: concepts for agency*, **Erkenntnis** 34 (1991), pp. 137 - 169.

[5] N. BELNAP, *Declaratives are not enough*, **Philosophical studies** 59 (1990), pp. 1 - 30.

[6] N. BELNAP and M. PERLOFF, *Seeing to it that: a canonical form for agentives*, **Theoria** 54 (1988), pp. 175 - 199. W. Edelberg pointed out to us that the informal semantic account of *stit* is garbled in this paper; the account is correct in the version of this paper republished in **Knowledge representation and defeasible reasoning**, H. E. Kyburg, Jr, R. P. Loui and G. N. Carlson (eds), Kluwer Academic Publishers, Dordrecht/Boston 1990, pp. 167 - 190; a correct account also occurs in [4].

[7] R. M. CHISHOLM, *Human freedom and the self*, The Lindley Lecture, Department of Philosophy, University of Kansas, 1964, pp. 3 - 15. Reprinted in **Freedom and determinism**, K. Lehrer (ed.), Random House, New York 1966, pp. 11 - 44; and in **On Metaphysics**, R. M. Chisholm (ed.), University of Minnesota Press, Minneapolis 1989, pp. 5 - 15.

[8] J. M. FISCHER, (ed.), **Moral Responsibility**, Cornell University Press, Ithaca and London 1986.

[9] H. G. FRANKFURT, *Alternate possibilities and moral responsibility*, **Journal of Philosophy** 66 (1969), pp. 828 - 839. Reprinted in **Moral Responsibility**, J. M. Fischer, (ed.), Cornell University Press, Ithaca and London 1986, pp. 143 - 152.

[10] D. GABBAY and G. GUENTHNER, (eds.), **Handbook of Philosophical Logic**, vol. II: **Extensions of Classical Logic**, Synthese library, Studies in epistemology, vol. 165, D. Reidel Publ. Co., Dordrecht 1984.

[11] R. HILPINEN, (ed.), **New studies in deontic logic: norms, actions, and the foundations of ethics**, D. Reidel Publ. Co., Dordrecht 1981.

[12] D. MAKINSON, *On the formal representation of rights relations*, **Journal of Philosophical Logic** 15 (1986), pp. 403 - 425.

[13] G. E. MOORE, **Ethics**, Home university library of modern knowledge, no. 52, H. Holt and Company, New York 1912. Reprinted Oxford University Press, New York 1965.

[14] M. PERLOFF, Stit *and the language of agency*, **Synthese** 86 (1991), pp. 379 - 408.

[15] R. H. THOMASON, *Indeterminist time and truth-value gaps*, **Theoria** 36 (1970), pp. 264 - 281.

[16] R. H. THOMASON, *Deontic logic and the role of freedom in moral deliberation*, in **New studies in deontic logic: norms, actions, and the foundations of ethics**, Risto Hilpinen (ed.), D. Reidel Publ. Co., Dordrecht 1981, pp. 177 – 186.

[17] R. H. THOMASON, *Combinations of tense and modality*, in **Handbook of Philosophical Logic**, vol. II: **Extensions of Classical Logic**, D. Gabbay and G. Guenthner, (eds.), D. Reidel Publ. Co., Dordrecht 1984, pp. 135 – 165.

[18] M. THOMSEN, **The Saddest Pleasure**, Graywolf Press, Saint Paul, Minnesota 1990.

[19] P. VAN INWAGEN, *Ability and Responsibility*, **Philosophical Review** 87 (1978), pp. 201 – 224. Reprinted in **Moral Responsibility**, J. M. Fischer, (ed.) Cornell University Press, Ithaca and London 1986, pp. 153 – 173.

[20] M. XU, *Modalities in* stit *theory without the refref conjecture*. Unpublished manuscript, University of Pittsburgh, Department of Philosophy, December, 1989.

UNIVERSITY OF PITTSBURGH

Received February 20, 1991.