DYNAMICS OF COUNTERFACTUAL SUPPOSITION I: THE QUALITATIVE CASE

FRANCESCO BERTO AND SOROUSH RAFIEE RAD

Abstract. This paper studies the dynamics of counterfactual supposition using qualitative belief revision techniques; a second paper will provide a quantitative account using probabilities. We individuate four general features of suppositional thinking; define a *counterfactual dynamics*; show that it displays those features by recapturing the axioms of KGM belief update theory. We then isolate a crucial difference between counterfactual supposition and supposition 'in the indicative mood', namely that the former, but not the latter, can make the supposing agent lose certainties. We propose a Dynamic Logic of Counterfactual Supposition (DLCS) with a sound axiomatization. We prove a Ramsey Test for the logic: one believes the counterfactual that if it were that P, it would be that Q, precisely when one comes to believe that Q on the counterfactual supposition of P.

1. Supposing, Learning, and Belief Dynamics

What would be the case if P was the case? We deal with counterfactual questions all the time, in explanatory reasoning ('Would the patient show such symptoms if it was smallpox?'), to ascertain legal and moral responsibilities ('Would the pedestrian have been hit if the brakes had worked as expected?'), learn from mistakes ('Would Johnson have scored if he had played centre forward?'), choose what to do next ('Would I make more money if I invested in equities?'). Counterfactual thinking pervades intelligent life. How does it work?

It is often remarked that our primary way of assessing conditionals is by supposing the antecedent and checking the believability of the consequent on the supposition (Williamson, 2020). The idea is at the core of 'suppositional accounts' of conditionals (Adams, 1975; Edgington, 1995; Bennett, 2003; Evans and Over, 2004); however, these have been developed mainly in a probabilistic setting and for indicatives, rather than counterfactuals. The view is at least as old as a mythical footnote by Frank Ramsey:

> If two people are arguing 'If p will q' and are both in doubt as to p, they are adding p hypothetically to their stock of

Date: 24 July 2026.

> knowledge and arguing on that basis about q; so that in a sense 'If p, q' and 'If p, $\neg q$' are contradictories. We can say that they are fixing their degrees of belief in q given p. (Ramsey, 1990, 155n)

This we now call the Ramsey Test. It makes of our assessment of conditionals a sort of *simulated belief revision*. It takes our assessment of Q on the supposition of P as structurally similar to our assessing Q in the light of a revision of our beliefs upon learning that P. This makes sense because, just as imagining is a bit like seeing, supposing is a bit like learning. Supposing that Paul is in the kitchen, or imagining a situation where Paul is in the kitchen, we then adjust our beliefs to the supposition, checking what we take to be true in the imagined scenario. We can come to believe that in the imagined scenario, that is, within the supposition, it's dinner time. This is a bit like seeing, and thus learning, that Paul is in the kitchen and then adjusting our beliefs accordingly, coming to believe that it's dinner time.[1] The idea sits well with simulationist accounts of imagination, which have it that imagination imitates counterpart non-imaginative mental states and activities (Currie and Ravenscroft, 2002; Goldman, 2002).

The formal literature on belief revision agrees that revision must be *minimal* (Alchourrón et al., 1985; Grove, 1988; Van Ditmarsch et al., 2008; Baltag and Smets, 2008; Van Benthem, 2011): one ought to change one's beliefs as little as possible upon learning that P. Correspondingly, per the Ramsey Test, one should assess a conditional with antecedent P and consequent Q by assessing Q against a belief system that's changed, in one's mental simulation, as little as possible to accommodate the supposition P. The literature on the epistemic value of imagination agrees, too (Kind, 2016; Langland-Hassan, 2016; Van Leeuwen, 2016; Williamson, 2016): imagination can have epistemic value in hypothetical thinking, to the extent that the development of an imaginative input is constrained by principles of plausibility and adherence to how one takes reality to be. What's difficult is to fine-tune what 'as little as possible' exactly means.

Conditionals give us reasons to think that minimality comes in more than one fashion. For look at Lewis (1973)'s famous example (credited to Adams):

(1) If Oswald did not kill Kennedy, then someone else did.
(2) If Oswald had not killed Kennedy, then someone else would have.

[1] The correspondence is not perfect. Supposing, for instance, that Paul is stealing money from us, we may come to believe we'll never find out, for he is smart. But if we learn that Paul is stealing money from us, we have found out. This links to so-called 'Thomason conditionals' (see Van Fraassen (1980); Bennett (2003), 28-9): conditionals whose consequent has to do with one's ignorance of the antecedent. That they need a special treatment is no objection to the structural similarity in the general case.

(1) sounds very acceptable, (2) doesn't (if one believes in the verdict of the Warren commission). By the Ramsey Test, that's because the consequent of (1) looks believable, the one of (2) doesn't, on supposing the antecedent. But because it's the same antecedent and consequent, the difference must be in how we suppose and, accordingly, revise beliefs. So the literature distinguishes indicative and subjunctive or counterfactual supposition (Joyce, 1999; Byrne, 2005; Leitgeb and Segerberg, 2007; Eva et al., 2021).

Beliefs can be all-or-nothing, or graded (then they're often called 'credences'). So belief revision can be phrased in a qualitative or quantitative fashion. This paper is the first of a two-part work, focusing on the qualitative case and using modal-epistemic logic to capture belief revision on a counterfactual supposition. Part II will deal with the quantitative case, using probabilities.

Here's one further important feature of such mental simulation activity, recognized in the literature: it is *episodic* (see e.g. Wansing, 2017; Badura, 2021). In an episode of hypothetical thought, we suppose that P deliberately, wondering what follows. We usually have specific Qs to focus on. We carry out the exercise, revising our beliefs under the supposition. The exercise unfolds in time. We stop after a (long or short) while, when we reach a stable verdict on the acceptability of Q on the supposition P. (Or, when the real world gets in the way and interrupts our mental exercise!)

Thus, simulated belief revision under supposition has its *dynamics*. Now the dynamics of belief revision under the indicative supposition of P must be very different from that of belief revision under the counterfactual supposition of P. And while the former has been extensively investigated in the formal literature on belief revision and in particular in Dynamic Epistemic Logic (DEL) (see, e.g., Van Ditmarsch et al., 2008; Baltag and Smets, 2008; Van Benthem, 2011), the latter is less well-understood and studied, one notable exception being the wonderful Leitgeb and Segerberg (2007) which, however, rather than DEL, employs the framework of Segerberg's Dynamic Doxastic Logic (DDL). We get back to the Leitgeb-Segerberg work later on.

2. Features of Supposition

We start by listing four general features of the kind of hypothetical thinking we are after:

(Certainty) *One who supposes P believes P on the supposition.*

(Minimality) *If P is already believed, one does not revise one's beliefs on the supposition of P.*

(Consistency) *If Q is believed on the supposition of P, and P is believed on the supposition of Q, then revising by P and by Q must lead to the*

same result.

(Possibility) *If P is possible (and, one's beliefs are consistent), then one's beliefs on the supposition of P will (still) be consistent.*

These surface here and there in the literature and there's some agreement on their holding for both indicative and counterfactual supposition, both for all-or-nothing belief and for credences (Lewis, 1973, 1976; Skyrms, 1987; Joyce, 1999).

Certainty is endorsed, e.g., by Williamson: 'trivially, one accepts A on the supposition A'. 'Anything is to be accepted on the supposition of itself. Therefore, by the Suppositional Rule [i.e. the Ramsey Test], "If A, A" is always to be accepted' (Williamson, 2020, 18 and 53). One finds the idea in formal accounts of belief revision under the label of the 'Success' principle: after updating by P, one believes P (Alchourrón et al., 1985; Van Benthem and Smets, 2015). For supposition, it may not hold unrestrictedly, however: there may be such a thing as *partial supposition*, whereby one treats the supposed content as having an increased degree of plausibility or likelihood but stops short of endorsing it as true within the supposition. This is best dealt with in a probabilistic setting, so we postpone discussion to our Paper II. The idea is investigated in Eva and Hartmann (2021).

Minimality follows from the idea that revision must be, well, minimal. It says: 'no revision when none is needed' (Lewis, 1976, 150). If you are already convinced that Paris is in France, and you're asked to suppose that Paris is in France, you'll feel there is nothing new for your belief system to adjust to on the supposition.

We gave Consistency its name following Lewis, who says that 'the method [of belief revision] must be consistent in its comparisons' (Lewis, 1976, ibid): if it failed, revising by P would be both more and less of a revision than revising by Q.

Possibility encodes the idea that when you suppose a possible P, you'll only look at possible scenarios – unless your beliefs were inconsistent in themselves already, before the supposition, and then it's their fault if your revised beliefs could not be true in any possible scenario.

We consider it a *desideratum* for a good account of the dynamics of counterfactual supposition that it satisfies these conditions. We mentioned that these are (taken as) shared by indicative supposition. We'll come later to a key feature pulling the two kinds apart.

3. Making Things Precise

Start with a nonempty set of possible worlds W and take a proposition $P \in \mathcal{P}(W)$ as represented, in traditional fashion, by the set of worlds where it holds: P is true at w iff $w \in P$.

Definition 1. A *plausibility relation* on W is a family $\{\leqslant_w\}_{w\in W}$ where each $\leqslant_w$ is a total well-preorder on W.

For $P \subseteq W$, let $\leqslant_w^P$ be the restriction of $\leqslant_w$ to P and let $min_P \leqslant_w$ be the set of minimal elements of P with respect to $\leqslant_w$:

$$min_P \leqslant_w = \{w' \in P \,|\, \forall w'' \in P \ , \ w'' \not<_w w'\},$$

where $w'' <_w w'$ if and only if $w'' \leqslant_w w'$ and $w' \not\leqslant_w w''$. We write $min \leqslant_w$ for $min_W \leqslant_w$.

Plausibility will be at work in the characterization of belief and (dispositions to) belief revision. The idea will be, as is common in the literature (see e.g. Grove (1988); Van Benthem and Smets (2015)), that what one believes is what holds at the worlds one finds *most plausible* given one's total information. One could call these the agent's *belief-worlds*.

That plausibility is a preorder, not an order, means there can be ties in the competition for plausibility. That it's total means we can always compare worlds for plausibility. That it's a well-preorder means that there will never be infinite descending chains of worlds, getting more and more plausible, forever, i.e., $min_P \leqslant_w \neq \emptyset$ when $P \neq \emptyset$.

Definition 2. A *divergence relation* on W is a family $\{\preccurlyeq_w\}_{w\in W}$ of total preorders on W such that for each $w \neq w' \in W$:

(i) $w \prec_w w'$: each w is the single least divergent world from itself.
(ii) The restriction of $\preccurlyeq_w$ to any non-empty $V \subseteq W$ is a total well-preorder, so there are no infinite descending chains in V.

($w'' \prec_w w'$ if and only if $w'' \preccurlyeq_w w'$ and $w' \not\preccurlyeq_w w''$.) For $P \subseteq W$, let $\preccurlyeq_w^P$ be the restriction of $\preccurlyeq_w$ to P and let $min_P \preccurlyeq_w$ be the set of minimal elements of P with respect to $\preccurlyeq_w$. In words: $min_P \preccurlyeq_w$ are the closest to w (least divergent) worlds in P. Condition (ii) ensures that in general $min_P \preccurlyeq_w \neq \emptyset$.

Divergence is the hallmark of counterfactual supposition. One can take $min_P \preccurlyeq_w$ as the worlds one reaches, starting from w and applying, in one's counterfactual imagination, the minimal changes required to make P true. One who counterfactually supposes that Paul is in the kitchen will imagine a situation that's changed as little as possible from the viewpoint of w, compatibly with making it the case that Paul is in the kitchen.

This becomes obvious as soon as one phrases the idea in terms of *selection functions*:

Definition 3. A *selection function* on W is a mapping $f : W \times \mathcal{P}(W) \to \mathcal{P}(W)$ such that for all $w \in W$ and all $P, Q \subseteq W$:

(1) $f(w, P) \subseteq P$.
(2) If $w \in P$, then $f(w, P) = \{w\}$.
(3) If $f(w, P) \subseteq Q$ and $f(w, Q) \subseteq P$ then $f(w, P) = f(w, Q)$
(4) If $P \neq \emptyset$, then $f(w, P) \neq \emptyset$

Each divergence relation on W imposes a selection function: when $\{\preccurlyeq_w\}_{w \in W}$ is a divergence relation on W, then $f : W \times \mathcal{P}(W) \to \mathcal{P}(W)$, defined as $f(w, P) = min_P \preccurlyeq_w$, is a selection function.

Selection functions are familiar from the literature on conditional logics (Stalnaker, 1968; Lewis, 1973), where one gives the truth conditions for a *ceteris paribus* conditional, say $\varphi > \psi$, by claiming that it's true at w iff the consequent ψ is true at the selected antecedent-worlds, i.e., iff $f(w, ||\varphi||) \subseteq ||\psi||$ ($||\varphi||$ being the proposition expressed by φ: the set of worlds where φ is true).

Now give '$f(w, P)$' a suppositional reading: take it as picking the worlds one looks at, from the viewpoint of w, on the counterfactual supposition of P; that is, the worlds one reaches by implementing, in imagination, the least drastic change required to make P true (an insight already in Stalnaker (1968)). Then our conditions (1)-(4) mirror exactly our four conditions on suppositional thought. (1) maps to Certainty: all the worlds one looks at on supposing P must make P true. (2) maps to Minimality: when P is already true at w, the single closest world to w is w itself, i.e., one moves nowhere away from w (this is often called (Strong) Centering in the literature on conditional logics). (3) maps to Consistency (in Comparisons): if the least divergent P-worlds make true Q and vice versa, one looks at the same worlds when supposing P and when supposing Q (this is sometimes called Uniformity). (4) maps to Possibility: if P is possible, one must look at some possible scenario when supposing it.

It will be useful to compare indicative and counterfactual supposition by mapping the former to belief revision as captured in Dynamic Epistemic Logic (DEL). The following essentially re-purposes what one can find in the lovely introduction to DEL by Van Benthem and Smets (2015).

In dynamic logics in general, one has in the language operators $[a]$ ('after action a has been executed') which work as model-transformers: one claims that $[a]\varphi$ holds in a model when φ holds in the model(s) obtained by transforming the initial one as per the instructions encoded in $[a]$. In DEL, the focus is on epistemic actions such as learning.

Learning a true P with certainty should reduce the set of possibilities W to those consistent with P. This is often called 'hard information': information that irrevocably changes the set of epistemic possibilities for an agent. However, not all information has such drastic effects. For instance, learning information from not fully reliable sources, or that is uncertain in itself, does not necessarily reduce the set of possibilities. What such information does change is what is believed locally, at a world w. Let's unpack.

Definition 4. P is *believed*, $Bel(P)$, at w, iff $min \leqslant_w \subseteq P$.

Definition 5. Q is *believed on the indicative supposition of* P, $Bel^P(Q)$, at w, iff $min_P \leqslant_w \subseteq Q$.

So P is believed when it's true at all of one's belief-worlds, i.e., at the worlds one finds most plausible. Q is believed on the indicative supposition of P when it's true at all of one's belief-worlds where P is true, i.e., at the most plausible P-worlds.

Soft information can prompt a dynamics of belief revision by changing the plausibility ordering between worlds. Two popular revision policies are the *radical* (aka *lexicographic*) and *conservative* upgrades (see Van Benthem and Smets, 2015, 334-6) – in terms of supposition:

Definition 6. Q is believed on the *radical supposition* of P, ${}^{\Uparrow P}Bel(Q)$, at w, iff $min \leqslant'_w \subseteq Q$ where (with $\bar{P}$ the complement of P):

$$w' \leqslant'_w w'' \iff \begin{cases} w' \in P \ \& \ w'' \in \bar{P} \\ w', w'' \in P \ \& \ w' \leqslant_w w'' \\ w', w'' \in \bar{P} \ \& \ w' \leqslant_w w'' \end{cases} \tag{1}$$

Definition 7. Q is believed on the *conservative supposition* of P, ${}^{\uparrow P}Bel(Q)$, at w, iff $min \leqslant'_w \subseteq Q$ where

$$w' \leqslant'_w w'' \iff \begin{cases} w' \in min_P \leqslant_w \\ w', w'' \notin \min_P \leqslant_w, \text{and} w' \leqslant_w w'' \end{cases} \tag{2}$$

So the radical setting makes any P-world more plausible than any not-P world while leaving the old plausibility ordering in place within the two groups. The conservative setting replaces the current ordering with one that puts the most plausible P-worlds ahead, leaving the rest unchanged.

Other formats have been studied, but this should suffice to get an idea. What matters for us is that the *counterfactual* supposition of P is quite different from its indicative counterpart:

Definition 8. Q is believed on the *counterfactual supposition* of P, $Bel_P(Q)$, at w, iff

$$\bigcup_{w' \in \min_{\leqslant_w}} min_P \preccurlyeq_{w'} \subseteq Q$$

So Q is believed on the counterfactual supposition of P when it's true at all of the P-worlds least divergent from each of one's belief-worlds. Thus, notice the difference: on the indicative supposition of P, we look at *the most plausible* P-worlds; on the counterfactual supposition of P, we look at the P-worlds *closest to* the most plausible worlds.

When we suppose that P indicatively, we pretend to be getting new information about actuality. Hence, only plausibility is at work: there is no departure from *here* to check how things would be if P, i.e., to check how things are *there*, i.e., at the minimally divergent world(s) where P. So we work globally: we take all of our belief-worlds, representing how we take reality to be, and we have them interact with P by intersecting them with the P-worlds, or by variously re-ordering their plausibility in the light of P.

When we suppose that P counterfactually, instead, we depart from how we believe reality to be so that we can make nontrivial comparisons with it from within the supposition: divergence enters into play. We work locally: for each of our belief-worlds, we switch to its closest P-world(s), i.e., to what we reach by minimal change imposed, in our imagination, in order to make P true; and we lump those together.

4. Counterfactual Dynamics Recaptures KGM

The celebrated AGM theory of belief revision by Alchourrón et al. (1985) provides the standard framework when the revision corresponds to learning new information; it is naturally mapped to indicative supposition. For counterfactual supposition, the theory to look at is the KGM account (so-named in Leitgeb and Segerberg (2007) after the seminal paper by Katsuno and Mendelzon (1992), complemented by the work of Grahne, e.g., Grahne (1998)). Let's explore why this is so.

The kind of revision captured by KGM is often called 'belief update', with 'belief revision' reserved to AGM in contexts of comparison between the two – but we need not quibble about the terminology. Katsuno and Mendelzon marked the difference with AGM by claiming that KGM models how one changes one's beliefs in response to relevant worldly change, whereas AGM models how one does it in response to new information on a relevantly unchanged world, but this way of capturing their difference is disputed – see e.g. Friedman and Halpern (1999).

What matters for us, is that KGM is good for counterfactual supposition, a point investigated in Eva et al. (2021). We will now show this by revisiting in an abstract setting the dynamics we introduced in the previous section

for counterfactual supposition and showing that it satisfies the KGM belief update postulates one can find in Katsuno and Mendelzon (1992).

The KGM postulates include principles that match our four general features of supposition, singled out in section 3: Certainty, Minimality, Consistency (in Comparisons), and Possibility. By recapturing KGM, our dynamics will therefore comply with them as well. In the section 5, we will then see that the dynamics of counterfactual supposition fails a principle which is, instead, satisfied by indicative supposition and its dynamics. This failure marks their crucial difference.

Definition 9. Let w be a world.

- We call $Bel(w) = \min \leqslant_w$ the *belief set* at w, i.e., the set of most plausible worlds giving what one believes at w.
- We call $Bel_P(w) = \bigcup_{v \in Bel(w)} \min_P \preccurlyeq_v$ the *P-counterfactual belief set* at w.
- We say that (counterfactual) belief at w is complete if $Bel(w)$ (res. $Bel_P(w)$) is a singleton.

Notice that the belief and counterfactual belief sets characterise the set of beliefs and the set of P-counterfactual beliefs, in the sense that they define the strongest belief and P-counterfactual belief at w. In particular, $Bel(P)$ and $Bel_P(Q)$ hold at w iff $Bel(w) \subseteq P$ and $Bel_P(w) \subseteq Q$ respectively.

Counterfactual dynamics transforms a proposition (a set of worlds) into another:

Definition 10. The *P-counterfactual dynamics* is a mapping $(\cdot)_P : \mathcal{P}(W) \to \mathcal{P}(W)$ defined for all $V \subseteq W$ as:

$$V_P = \bigcup_{v \in V} f(v, P)$$

This essentially rephrases in our abstract setting the idea originally formulated by Katsuno and Mendelzon in terms of 'possible models': one updates V on P by (a) moving from each point/world/model $v \in V$ to those closest to v where P holds, and (b) collecting all of these points/worlds/models together: see Katsuno and Mendelzon (1992), 188 and compare Grahne (1998), 94.

In particular, the P-counterfactual dynamics transforms a belief set $Bel(w) = \min \leqslant_w$ to the belief set $Bel_P(w) = \bigcup_{v \in \min \leqslant_w} f(v, P) = \bigcup_{v \in \min \leqslant_w} \min_P \preccurlyeq_v$.

Theorem 1. *The P-counterfactual dynamics satisfies the KGM belief update axioms.*[2]

[2]Katsuno and Mendelzon worked with a simple propositional language and had belief states ('knowledge bases', in their terminology) represented by single formulas of the

Proof. Let $v, u, w \in W$ and $P, Q \subseteq W$.

(i) $Bel_P(w) \subseteq P$

[*As per Certainty*: P is believed on the counterfactual supposition of P.]

To see this, remember that $f(v, P) \subseteq P$, so

$$Bel_P(w) = \bigcup_{v \in \min_{\leqslant_w}} f(v, P) \subseteq P$$

(ii) If $Bel(w) \subseteq P$, then $Bel_P(w) = Bel(w)$

[*As per Minimality*: if P is believed already, then the counterfactual supposition of P changes nothing.]

First, if $v \in Bel(w) = \min \leqslant_w \subseteq P$ then $f(v, P) = \{v\}$. Thus $Bel_P(w) = \bigcup_{v \in \min_{\leqslant_w}} f(v, P) = \min \leqslant_w = Bel(w)$.

(iii) If $Bel(w) \neq \emptyset$ and $P \neq \emptyset$ then $Bel_P(w) \neq \emptyset$

[*As per Possibility*: if one's beliefs are consistent and P is possible, then one's beliefs are consistent, i.e., one looks at some possibility after the counterfactual supposition of P.]

If $Bel(w) \neq \emptyset$, let $v \in Bel(w)$. Then since $P \neq \emptyset$, we have $f(v, P) \neq \emptyset$ and thus $\bigcup_{v \in \min_{\leqslant_w}} f(v, P) \neq \emptyset$ and so $Bel_P(w) \neq \emptyset$

(iv) If $Bel(w) = Bel(v)$ and $P = Q$ then $Bel_P(w) = Bel_Q(v)$

[Equivalent belief sets give equivalent belief sets after supposing equivalents.]

---

language. Denoting with '$\psi \diamond \varphi$' the result of updating belief state ψ by φ, they had the following axioms (Katsuno and Mendelzon, 1992, 189):

(1) $\psi \diamond \varphi$ implies φ.
(2) If ψ implies φ, then $\psi \diamond \varphi$ is equivalent to ψ.
(3) If both ψ and φ are satisfiable, then $\psi \diamond \varphi$ is also satisfiable.
(4) If $\psi_1 \leftrightarrow \psi_2$ and $\varphi_1 \leftrightarrow \varphi_2$ then $\psi_1 \diamond \varphi_1 \leftrightarrow \psi_2 \diamond \varphi_2$.
(5) $(\psi \diamond \varphi) \wedge \theta$ implies $\psi \diamond (\varphi \wedge \theta)$.
(6) If $\psi \diamond \varphi_1$ implies φ_2 and $\psi \diamond \varphi_2$ implies φ_1 then $\psi \diamond \varphi_1 \leftrightarrow \psi \diamond \varphi_2$.
(7) If ψ is complete then $(\psi \diamond \varphi_1) \wedge (\psi \diamond \varphi_2)$ implies $\psi \diamond (\varphi_1 \vee \varphi_2)$.
(8) $(\psi_1 \vee \psi_2) \diamond \varphi \leftrightarrow (\psi_1 \diamond \varphi) \vee (\psi_2 \diamond \varphi)$.

A complete belief state ψ is one that for each formula θ implies either θ or $\neg\theta$. So a complete belief state is 'maximally opinionated', deciding each issue that can be expressed in the language.

Since $Bel(w) = Bel(v)$

$$Bel_P(w) = \bigcup_{u \in \min_{\leqslant_w}} f(u,P) = \bigcup_{u \in \min_{\leqslant_v}} f(u,P) = \bigcup_{u \in \min_{\leqslant_v}} f(u,Q) = Bel_Q(v)$$

(v) $Bel_P(w) \cap Q \subseteq Bel_{P \cap Q}(w)$

[Conjoining the revision by P with Q is the same as revising by $P\&Q$.]

Let $v \in Bel_P(w) \cap Q$, choose $u \in Bel(w)$ with $v \in f(u,P)$, and notice that $v \in P \cap Q$. So no $P \cap Q$ world can be closer to u than v, otherwise it would also be P world that is closer to u than v, and that contradicts $v \in f(u,P)$. So $v \in f(u, P \cap Q) \subseteq Bel_{P \cap Q}(w)$.

(vi) If $Bel_P(w) \subseteq Q$ and $Bel_Q(w) \subseteq P$ then $Bel_P(w) = Bel_Q(w)$.

[*As per Consistency in Comparisons*: if the revision by P implies Q and vice versa, revising by P is the same as revising by Q.]

To see this, note that for all $u \in \min \leqslant_w$, $f(u,P) \subseteq Bel_P(w) \subseteq Q$ and $f(u,Q) \subseteq Bel_Q(w) \subseteq P$. So by Definition 3 (3), for all $u \in \min \leqslant_w, f(u,P) = f(u,Q)$. Taking union over all $u \in \min \leqslant_w$ we get $Bel_P(w) = Bel_Q(w)$.

(vii) If $Bel(w)$ is complete then $Bel_P(w) \cap Bel_Q(w) \subseteq Bel_{P \cup Q}(w)$.

[Intersecting the revision by P and that by Q implies revising by $P \vee Q$.]

Let $Bel(w) = \{u\}$ and $v \in Bel_P(w) \cap Bel_Q(w) = \bigcup_{s \in \min_{\leqslant_w}} f(s,P) \cap \bigcup_{s' \in \min_{\leqslant_w}} f(s',Q) = f(u,P) \cap f(u,Q)$.
Now suppose, towards a contradiction, that $v \notin Bel_{P \cup Q}(w)$ and thus $v \notin \bigcup_{u \in \min_{\leqslant_w}} f(u, P \cup Q) = f(u, P \cup Q)$. So there exists $v_u \prec_u v$ such that $v_u \in f(u, P \cup Q) \subseteq P \cup Q$. Since $v_u \in P \cup Q$, either $v_u \in P$ or $v_u \in Q$. If $v_u \in P$ then $v \notin f(u,P)$ and thus $v \notin f(u,P) \cap f(u,Q)$. Similarly, if $v_u \in Q$ then $v \notin f(u,Q)$ and thus $v \notin f(u,P) \cap f(u,Q)$ which gives the necessary contradiction.

(viii) $(Bel(w) \cup Bel(v))_P = Bel_P(w) \cup Bel_P(v)$

[Finally: first fusing two belief sets and then revising the result by P is the same as first revising each by P and then fusing the results.]

To see this, observe that

$$(Bel(w) \cup Bel(v))_P = \bigcup_{u \in Bel(w) \cup Bel(v)} f(u,P) = \bigcup_{u \in Bel(w)} f(u,P) \cup \bigcup_{u' \in Bel(v)} f(u',P)$$

$$= \bigcup_{u \in \min_{\leqslant_w}} f(u, P) \cup \bigcup_{u' \in \min_{\leqslant_v}} f(u', P) = Bel_P(w) \cup Bel_P(v).$$

□

5. Preservation of Certainties

It is not difficult to show that (counterpart formulations of) Certainty, Minimality, Consistency in Comparisons, and Possibility, hold for AGM and indicative supposition. So while it's good for our counterfactual dynamics to satisfy them as well, insofar as they're conditions on the adequacy of an account of supposition in general, this tells us little on how counterfactual supposition differs from its indicative cousin. Let's delve into this by starting with what we take to be a mistaken way of drawing the distinction.

It is a piece of common wisdom that we only suppose that P indicatively when we are not certain that P is false. When we are so certain, this is the territory of counterfactual supposition. For instance, glossing on the Ramsey Test for indicatives, Bennett says:

> [Y]ou cannot say what the upshot is of adding to your belief system something you actually regard as having *no* chance of being true [...]. There is abundant intuitive evidence that *nobody has any use for* $A \rightarrow C$ *when for him* $P(A) = 0$. (Bennett, 2003, 55)

(Bennett makes the point for degrees of belief represented by probabilities, but of course it carries over to all-or-nothing belief.)

Common wisdom is wrong, we think. One can non-trivially reason with indicatives whose antecedent one disbelieves: one can assess Indicative Oswald also when one is sure that Oswald killed Kennedy (Gillies, 2004). Thinkers who don't deem themselves infallible can suppose in the indicative mood that P also when they utterly disbelieve P, and assess whether Q is the case under that supposition:

> [I]t is often assumed that any form of probabilistic belief revision that involves 'raising the dead' by increasing the probabilities of certainly false propositions must involve counterfactual beliefs. This is not so. It is logically consistent both to be certain that some proposition is false and yet to speculate about what the world is like if one is in fact wrong. To be subjectively certain of something is, after all, not the same as regarding oneself infallible on the matter. (Joyce, 1999, 203)

One example is given by Williamson (2020): you can imagine that you don't exist and reasonably assess the indicative 'If I do not exist now, then thinking occurs without a thinker' although, presumably, you have Cartesian certainty in the falsity of the antecedent. One doesn't need to get so philosophical anyway, for mathematics provides a plethora of cases. The maths teacher in the business of a proof by *reductio* can felicitously assert: 'If there is a largest prime, then some number is both prime and composite'.[3]

If the difference between indicative and counterfactual supposition is not what common wisdom claims it is, then what *is* it? We think, following Joyce (1999), that it has to do with the difference in the background beliefs that can be preserved, or lost, in the respective dynamics. And this will be the right way to mark the crucial aspect whereby indicative and counterfactual supposition differ.

Start with another definition:

Definition 11. Let $\star : W \times \mathcal{P}(W) \to \mathcal{P}(W)$ be a belief revision dynamics that, for each world w and proposition $P \subseteq W$, maps the belief set $Bel(w)$ to a new belief set $Bel(w) \star P$, the P-revised belief set at w.

Say $\star$ *preserves certainties* if for all $P, Q \subseteq W$ and $w \in W$, if $Bel(w) \subseteq Q$, and $Bel(w) \not\subseteq \neg P$, then $Bel(w) \star P \subseteq Q$.

So a belief revision dynamics preserves certainties when, if one already believes that Q (and doesn't disbelieve that P), one will retain one's belief in Q on supposing that P.

Indicative supposition preserves certainties – look at Indicative Oswald again: we are certain that Q, that Kennedy has been killed. So when we wonder, What if P?, What if Oswald didn't kill Kennedy?, we retain our belief in Q. Since Kennedy has been killed, it must have been someone else. Hence we believe the consequent on indicatively supposing the antecedent.

Proposition 1. *Indicative supposition, radical supposition and conservative supposition all preserve certainties. That is, if Q is believed at w, ($Bel(Q)$ holds at w) and the negation of P is not believed at w ($Bel(w) \not\subseteq \neg P$), then at w we have*

(i) $Bel^P Q$ *(Q is believed at w under indicative supposition that P)*

[3]Various philosophers, e.g., Lewis (1973), Williamson (2007), also think that counterfactual conditionals are ill-named because there can be true counterfactuals with true antecedents. (We ourselves are sticking with the terminology, rather than speaking of, say, 'subjunctive supposition', just because it's widespread.) But are counterfactuals that are not counterfactual ever acceptable and assertable, in particular, by one who believes the antecedent? Yes that can happen, too, for instance, in contexts of abductive reasoning. I firmly believe that the atom has been ionized, for this would explain, if true, the track left by a particle in an experiment we've just carried out. I can felicitously assert 'If the atom had been ionized, we'd see exactly the particle track we see', trying to convince you.

(ii) $\Uparrow^P BelQ$ *(Q is believed at w under radical supposition that P)*
(iii) $\uparrow^P BelQ$ *(Q is believed at w under conservative supposition of P)*

Proof. For (i) notice that since Q is believed at w, $min \leqslant_w \subseteq Q$ and since $min \leqslant_w = Bel(w) \not\subseteq \neg P$ we have $min_P \leqslant_w \subseteq min \leqslant_w \subseteq Q$. Thus by definition we have $Bel^P Q$ at w. For (ii) let $\leqslant'_w$ be as defined in Definition 6 and notice that since $min \leqslant_w = Bel(w) \not\subseteq \neg P$, we have $min \leqslant'_w \subseteq min_P \leqslant_w \subseteq Q$ and thus by definition we have $\Uparrow^P BelQ$ at w. And (iii) follows in a similar manner. □

Counterfactual supposition, instead, can make us lose certainties – look at Counterfactual Oswald again: when we wonder, What if it had been the case that P?, What if Oswald hadn't killed Kennedy?, we look, for each of our belief-worlds w, at the minimally divergent scenarios where P, i.e., we switch to $f(w, P)$, and lump them together. If we are Warrenite enough, these may include worlds with no killing. Hence we don't believe the consequent on counterfactually supposing the antecedent.

We think the opposition between preserving and losing certainties marks the fundamental difference between indicative and counterfactual supposition (thanks to an anonymous referee of this Journal for pressing us to make this explicit). The reason is that, in the indicative way of supposing that P, we *simulate getting new information about the actual world*, or *the actual world as unchanged in the relevant respects*. So things taken as facts, i.e., that we are certain of as actually holding, such as that Kennedy has been killed, are left untouched. In the counterfactual way of supposing that P, *we simulate that the world has been changed in the relevant respects, no more than needed to make it true*. So we may twist (things taken as) facts in our counterfactual mental simulation. For a different example used (in a quantitative, degrees-of-belief setting) by Hannes Leitgeb, let P = *The banana is not in the basket*. Then the difference is between 'whether the evidence is meant to express some additional information about an unchanged world (*the banana is not in the basket, i.e., it has not been there*) or rather information about how the world has changed (*the banana is not in the basket, i.e., it was removed if it had been there at all*)' (Leitgeb, 2017, 474).

Our sticking to actuality in the former case, not in the latter, is mirrored in the famously different behaviour of the two kinds of conditionals when actuality operators ('Actually P' is true at world w iff P is true at the actual world) are embedded in their consequent. The indicative ('If Midori is working, she is actually working') is trivially true; the counterfactual ('If Midori was working, she would actually be working') is not.

So it would be good for our counterfactual dynamics to deserve its name by not preserving certainties. And indeed, it doesn't:

Example 1 (Failure of Certainties Under Counterfactual Supposition)**.** Let $u, u', v, w \in W$ with $u, v \in Bel(w) \subseteq Q$, $v \in P$, $u \notin P$, $u' \notin Q$, and $f(u, P) = \{u'\}$. Then $u' \in Bel_P(w)$ and thus $Bel_P(w) \not\subseteq Q$.

So we have shown both how counterfactual dynamics satisfies the four general features of suppositional thinking i.e. Certainty, Minimality, Consistency, and Possibility, and how it diverges from indicative supposition by failing Preservation of Certainties. We now put our counterfactual dynamics to work in a full-fledged logical system. We record here our debt to Leitgeb and Segerberg (2007) which presents what is, to the best of our knowledge, the closest antecedent (pun intended) to our logic. We will comment on some similarities and differences at the end of the next section.

6. Dynamic Logic of Counterfactual Supposition

6.1. **Syntax and Semantics.** Given a set of propositional variables $\mathfrak{P}$, the sentences of DLCS, $S\mathcal{L}_{DLCS}$ are recursively defined as:

$$\varphi := p \,|\, \neg\varphi \,|\, \varphi \wedge \varphi \,|\, K\varphi |\, B\varphi \,|\, [^{\curvearrowleft}\varphi]\varphi \,|\, \varphi > \varphi$$

Read 'K' as 'one knows that', 'B' as 'one believes that', '$[^{\curvearrowleft}\varphi]$' as 'after the counterfactual supposition of φ', and '$>$' as a counterfactual conditional, 'if it were ... then it would be that'.

Definition 12. A *dynamic counterfactual model (DCM)* is a tuple $M = (W, \{\leqslant_w\}_{w\in W}, \{\preccurlyeq_w\}_{w\in W}, v)$: W is a non-empty set of worlds, $\{\leqslant_w\}_{w\in W}$ is a plausibility relation on W, $\{\preccurlyeq_w\}_{w\in W}$ is a divergence relation on W, and $v : \mathfrak{P} \to \mathcal{P}(W)$ is a valuation function that maps each propositional variable to a proposition taken as a set of worlds.

For any $P \subseteq W$, let $f_P(w) = min_P \preccurlyeq_w$ denote the set of minimal elements of P with respect to $\preccurlyeq_w$. In words, $f_P(w)$ denotes the set of closest or least divergent P-worlds to w.

We now define the model-transforming operation for counterfactual revision. Let $M = (W, \{\leqslant_w\}_{w\in W}, \{\preccurlyeq_w\}_{w\in W}, v)$ be a model and $P \subseteq W$.

Definition 13. The *counterfactual revision* of M with $P \subseteq W$ is a model $M^{\curvearrowleft P} = (W, \{\leqslant'_w\}_{w\in W}, \{\preccurlyeq_w\}_{w\in W}, v)$ where

$$w' \leqslant'_w w'' \iff \begin{cases} w' \in f_P(v) \text{ for some } v \in min \leqslant_w \\ w' \notin f_P(v) \;\&\; w'' \notin f_P(u) \text{ for all } v, u \in min \leqslant_w \;\&\; w' \leqslant_w w'' \end{cases} \tag{3}$$

So counterfactual revision turns M into a $M^{\curvearrowleft P}$ that differs from M in that the closest P-worlds to the most plausible worlds in M, i.e., intuitively, the

worlds reachable from each of the belief-worlds of the agent by minimal change required to make P true, become the most plausible worlds.

The truth conditions for formulas of the language go thus:

Definition 14. With model $M = (W, \{\leqslant_w\}_{w\in W}, \{\preccurlyeq_w\}_{w\in W}, v)$, $w \in W$:

(Sp) $M, w \vDash p \iff w \in v(p)$
(S$\neg$) $M, w \vDash \neg\varphi \iff M, w \nvDash \varphi$
(S$\wedge$) $M, w \vDash \varphi \wedge \psi \iff M, w \vDash \varphi$ and $M, w \vDash \psi$
(SK) $M, w \vDash K\varphi \iff M, v \vDash \varphi$ for all $v \in W$
(SB) $M, w \vDash B\varphi \iff M, v \vDash \varphi$ for all $v \in \min \leqslant_w$
(S$\curvearrowleft$) $M, w \vDash [\curvearrowleft \varphi]\psi \iff M^{\curvearrowleft\varphi}, w \vDash \psi$
(S$>$) $M, w \vDash \varphi > \psi \iff M, v \vDash \psi$ for all $v \in \min_{||\varphi||} \preceq_w$

(Again '$||\varphi||$' refers to the truth set of formula φ; also, we use '$M^{\curvearrowleft\varphi}$' as a shorthand for the clunky '$M^{\curvearrowleft||\varphi||}$'.)

So (SK) has it that knowledge is truth at all worlds. (SB) has it that belief is truth at the most plausible worlds. (S$\curvearrowleft$) has it that counterfactual supposition works as a model-transformer: after the counterfactual supposition of φ, ψ is the case iff ψ is true in the transformed model at w.

(S$>$) has it that φ counterfactually implies ψ, at w in M, iff ψ is true in all of w's closest φ-worlds, i.e., iff $f_{||\varphi||}(w) \subseteq ||\psi||$. So this is a variably strict conditional à la Stalnaker (1968), Lewis (1973).

Definition 15. We say that φ is a logical consequence of a set of formulas Γ, and write $\Gamma \vDash \varphi$, if for every model M, and every world w in M, if $M, w \vDash \psi$ for all $\psi \in \Gamma$ then $M, w \vDash \varphi$. We say that φ is a logical truth if for every model M and every world w in M, $M, w \vDash \varphi$.

We use the following shorthand notation:

$$\bot := p \wedge \neg p \qquad \top := p \vee \neg p \qquad \varphi \leftrightarrow \psi := (\varphi \rightarrow \psi) \wedge (\psi \rightarrow \varphi)$$

Via the belief operator B and dynamic operator $[\curvearrowleft \varphi]$ we can define in DLCS belief under counterfactual supposition:

- Belief Under the Counterfactual Supposition of φ

$$B_\varphi\psi := [\curvearrowleft \varphi]B\psi$$

Notice that by definition, $B_\varphi\psi$ if and only if $\bigcup_{v\in\min\leqslant_w} f_{||\varphi||}(v) \subseteq ||\psi||$.

Definition 16. For sentences $\varphi, \psi, \theta \in S\mathcal{L}_{DLCS}$, with $||\varphi|| \neq \emptyset$, we define $(\psi)_\varphi$, inductively as

$$(p)_\varphi = p$$
$$(\neg\psi)_\varphi = \neg(\psi)_\varphi$$
$$(\psi \wedge \theta)_\varphi = (\psi)_\varphi \wedge (\theta)_\varphi$$

$$(K\psi)_\varphi = K(\psi)_\varphi$$
$$(B\psi)_\varphi = B(\varphi > (\psi)_\varphi)$$
$$(\psi > \theta)_\varphi = (\psi)_\varphi > (\theta)_\varphi$$
$$([\curlywedge\, \theta]\psi)_\varphi = ((\psi)_\theta)_\varphi)$$

For $\varphi, \psi \in S\mathcal{L}_{DLCS}$, let $|\varphi|$ denote the length of φ define the the dynamic depth of φ, $d(\varphi)$ inductively as

$$d(p) = 0,$$
$$d(\neg\varphi) = d(K\varphi) = d(B\varphi) = d(\varphi)$$
$$d(\varphi \wedge \psi) = d(\varphi > \psi) = \max\{d(\varphi), d(\psi)\}$$
$$d([\curlywedge\, \varphi]\psi) = 1 + \max\{d(\varphi), d(\psi)\}.$$

Next let $\mu(\varphi) = (d(\varphi), |\varphi|)$ and order the sentences of $S\mathcal{L}_{DLCS}$ lexicographically by μ.

Lemma 1 (Depth bound)**.** *For all formulas φ, ψ, with $||\varphi|| \neq \emptyset$*

$$d((\psi)_\varphi) \leq \max\{d(\varphi), d(\psi)\}. \tag{D}$$

Proof. We show simultaneously, by well-founded induction on the lexicographic rank of ψ, $\mu(\psi)$, that $(\psi)_\varphi$ is well-defined and satisfies (D), uniformly for all φ.

If $\psi = p$, then

$$d((p)_\varphi) = d(p) = 0 \leq \max\{d(\varphi), d(p)\}.$$

For negation,

$$\begin{aligned} d((\neg\theta)_\varphi) &= d(\neg(\theta)_\varphi) \\ &= d((\theta)_\varphi) \\ &\leq \max\{d(\varphi), d(\theta)\} \\ &= \max\{d(\varphi), d(\neg\theta)\}. \end{aligned}$$

For knowledge,

$$\begin{aligned} d((K\theta)_\varphi) &= d(K(\theta)_\varphi) \\ &= d((\theta)_\varphi) \\ &\leq \max\{d(\varphi), d(\theta)\} \\ &= \max\{d(\varphi), d(K\theta)\}. \end{aligned}$$

For conjunction,

$$\begin{aligned} d((\theta \wedge \zeta)_\varphi) &= \max\{d((\theta)_\varphi), d((\zeta)_\varphi)\} \\ &\leq \max\{d(\varphi), d(\theta), d(\zeta)\} \\ &= \max\{d(\varphi), d(\sigma \wedge \tau)\}. \end{aligned}$$

The conditional case is analogous:

$$\begin{aligned} d((\theta > \zeta)_\varphi) &= \max\{d((\theta)_\varphi), d((\zeta)_\varphi)\} \\ &\leq \max\{d(\varphi), d(\theta), d(\zeta)\} \\ &= \max\{d(\varphi), d(\theta > \zeta)\}. \end{aligned}$$

For belief,

$$\begin{aligned} d((B\theta)_\varphi) &= d\big(B(\varphi > (\theta)_\varphi)\big) \\ &= \max\{d(\varphi), d((\theta)_\varphi)\} \\ &\leq \max\{d(\varphi), d(\theta)\}. \end{aligned}$$

Finally for the dynamic formulas $\psi = [\restriction \theta]\zeta$, notice that the proper subformula ζ has lower rank than ψ. Let $\eta := (\zeta)_\theta$, then by induction hypothesis η is well-defined and

$$\begin{aligned} d(\eta) &\leq \max\{d(\theta), d(\zeta)\} \\ &< 1 + \max\{d(\theta), d(\zeta)\} \\ &= d(\psi). \end{aligned}$$

So $\mu(\eta) < \mu(\psi)$, and by induction hypothesis we have

$$\begin{aligned} d((\psi)_\varphi) &= d((\eta)_\varphi) \\ &\leq \max\{d(\varphi), d(\eta)\} \\ &\leq \max\{d(\varphi), d(\theta), d(\zeta)\} \\ &\leq \max\{d(\varphi), d(\psi)\}. \end{aligned}$$

This completes the proof. □

As the Proposition 2 shows, $(\psi)_\varphi$ is the sentence that will be translated to ψ under the counterfactual assumption of φ.

Proposition 2. *For any DCM, M, world w, and formulas φ, ψ with $||\varphi|| \neq \emptyset$, and ψ only involves consistent counterfactual dynamics,*

$$M^{\restriction \varphi}, w \vDash \psi \iff M, w \vDash (\psi)_\varphi.$$

Proof. By induction on the lexicographical order rank of ψ, $\mu(\psi)$.

(i) $\psi = p$. Follows immediately.
(ii) $\psi = \theta \wedge \zeta$. Then $\mu(\theta) < \mu(\psi)$, and $\mu(\zeta) < \mu(\psi)$ so using the induction hypothesis we have:

$$M^{\restriction \varphi}, w \vDash \theta \wedge \zeta \iff M^{\restriction \varphi}, w \vDash \theta \& M^{\restriction \varphi}, w \vDash \zeta \iff$$

$$M, w \vDash (\theta)_\varphi \& M, w \vDash (\zeta)_\varphi \iff M, w \vDash (\theta)_\varphi \wedge (\zeta)_\varphi \iff M, w \vDash (\theta \wedge \zeta)_\varphi.$$

(iii) $\psi = \neg\theta$. Then $\mu(\theta) < \mu(\psi)$, and we have

$$M^{\curvearrowleft\varphi}, w \vDash \neg\theta \iff M^{\curvearrowleft\varphi}, w \nvDash \theta \iff$$

$$M, w \nvDash (\theta)_\varphi \iff M, w \vDash \neg(\theta)_\varphi \iff M, w \vDash (\neg\theta)_\varphi.$$

(iv) $\psi = K\theta$. Then $\mu(\theta) < \mu(\psi)$, and we have

$$M^{\curvearrowleft\varphi}, w \vDash K\theta \iff M^{\curvearrowleft\varphi}, v \vDash \theta \text{ for all } v \in W \iff$$

$$M, v \vDash (\theta)_\varphi \text{ for all } v \in W \iff M, w \vDash K(\theta)_\varphi \iff M, w \vDash (K\theta)_\varphi.$$

(v) $\psi = B\theta$. Then $\mu(\theta) < \mu(\psi)$, and

$$M^{\curvearrowleft\varphi}, w \vDash B\theta \iff M^{\curvearrowleft\varphi}, v \vDash \theta \text{ for all } v \in \min \leqslant_w^{\curvearrowleft\varphi} \iff$$

$$M, v \vDash (\theta)_\varphi \text{ for all } v \in \min \leqslant_w^{\curvearrowleft\varphi} \iff$$

$$M, v \vDash (\theta)_\varphi \text{ for all } v \in \bigcup_{u \in \min \leqslant_w} f_{||\varphi||}(u) \iff$$

$$\text{for all } u \in \min \leqslant_w, \big(M, v \vDash (\theta)_\varphi \text{ for all } v \in f_{||\varphi||}(u)\big) \iff$$

$$\text{for all } u \in \min \leqslant_w, (M, u \vDash \varphi > (\theta)_\varphi) \iff M, w \vDash B(\varphi > (\theta)_\varphi)$$

(vi) $\psi = \theta > \zeta$. Then $\mu(\theta) < \mu(\psi)$, and $\mu(\zeta) < \mu(\psi)$

$$M^{\curvearrowleft\varphi}, w \vDash \theta > \zeta \iff M^{\curvearrowleft\varphi}, v \vDash \zeta \text{ for all } v \in \min_{\prec_w} ||\theta||_{M^{\curvearrowleft\varphi}}$$

$$\iff \text{for all } v \in \min_{\prec_w} ||(\theta)_\varphi||_M \; v \in ||(\zeta)_\varphi||_M$$

$$\iff M, w \vDash (\theta)_\varphi > (\zeta)_\varphi \iff M, w \vDash (\theta > \zeta)_\varphi$$

(vii) $\psi = [\curvearrowleft \theta]\zeta$. Let $\gamma = (\zeta)_\theta$ and notice that$\mu(\zeta) < \mu(\psi)$ and $\mu(\gamma) < \mu(\psi)$. Then

$$M^{\curvearrowleft\varphi}, w \vDash [\curvearrowleft \theta]\zeta \iff (M^{\curvearrowleft\varphi})^{\curvearrowleft\theta}, w \vDash \zeta$$

$$\iff M^{\curvearrowleft\varphi}, w \vDash \gamma \iff M, w \vDash (\gamma)_\varphi \iff M, w \vDash ([\curvearrowleft \theta]\zeta)_\varphi.$$

□

6.2. **Ramsey Test.** Now we get a Ramsey Test for our counterfactual conditional – the proof is simple:

Proposition 3. *For all $\varphi, \psi \in S\mathcal{L}_{DLCS}$ with $||\varphi|| \neq \emptyset$,*

$$B_\varphi(\psi) \iff B(\varphi > (\psi)_\varphi)$$

Proof. Follows from Proposition 2 immediately. □

Corollary 1. *For any propositional sentence $\psi \in S\mathcal{L}_{DLCS}$,*

$$B_\varphi(\psi) \iff B(\varphi > \psi).$$

Proof. Notice that for a propositional sentence ψ without knowledge or belief operators, $(\psi)_\varphi = \psi$.

□

So counterfactuals are nicely mapped to counterfactual supposition in the way envisaged by Ramsey: one believes that if it were that φ, it would be that ψ, precisely when one comes to believe that ψ after revising one's belief system under the counterfactual supposition of φ.

6.3. **Axiomatisation.** We will now take the first steps towards axiomatising DLCS. Although we do not provide completeness results for the logic in this paper, we propose some axioms, prove soundness, and investigate a few sample theorems.

The axioms of DLCS are grouped into five categories: the first covers all instances of propositional tautologies; the second consists of axioms for knowledge; the third is the axioms capturing the behaviour of the Stalnaker conditional; the fourth covers axioms for belief; the fifth deals with axioms for the counterfactual dynamics.

Axiom System for DLCS:

- All instances of propositional tautologies

Knowledge

- S5 axioms for K[4]

Stalnaker Conditional

Ax. 1. $\varphi > \varphi$
Ax. 2. $(\varphi > \psi \wedge \varphi > (\psi \rightarrow \theta)) \rightarrow \varphi > \theta$
Ax. 3. $\varphi > \bot \leftrightarrow K\neg\varphi$
Ax. 4. $(\varphi > \psi \wedge K(\varphi \leftrightarrow \theta)) \rightarrow \theta > \psi$
Ax. 5. $\varphi \rightarrow (\varphi > \psi \leftrightarrow \psi)$
Ax. 6. $(\varphi > \theta \wedge \psi > \theta) \rightarrow ((\varphi \vee \psi) > \theta)$
Ax. 7. $((\varphi \vee \psi) > \theta) \rightarrow \varphi > \theta \vee \psi > \theta$
Ax. 8. $(\varphi > \psi \wedge \psi > \varphi) \rightarrow (\varphi > \theta \leftrightarrow \psi > \theta)$

Belief

Ax. 9. $B\top$
Ax. 10. $B\varphi \rightarrow \neg B\neg\varphi$
Ax. 11. $B(\varphi \wedge \theta) \leftrightarrow B\varphi \wedge B\theta$

[4]A helpful referee of this Journal asked us about our making the logic of knowledge as strong as S5. We are aware, of course, that one is thereby modelling strongly idealized agents, fully introspective with respect to what they do and do not know. However, on the one hand, full S5 strength streamlines the interaction of the operator with the belief and counterfactual operators in axioms 3, 4 and 12 below. On the other hand, having the semantics of K as that of an unrestricted quantifier over worlds above allowed us to keep our DCMs relatively simple, sparing us from the need for an additional accessibility relation. Since knowledge is not the focus of our investigation, making it S5-strong for the purposes of simplicity and technicality looks like an innocent idealisation to us.

Ax. 12. $(B\varphi \wedge K(\varphi \to \theta)) \to B\theta$

Counterfactual Supposition

Ax. 13. $[\curlywedge\, \psi]p \leftrightarrow p$ for all propositional variables p
Ax. 14. $[\curlywedge\, \psi]\neg\varphi \leftrightarrow \neg[\curlywedge\, \psi]\varphi$
Ax. 15. $[\curlywedge\, \psi](\varphi \wedge \theta) \leftrightarrow [\curlywedge\, \psi]\varphi \wedge [\curlywedge\, \psi]\theta$
Ax. 16. $[\curlywedge\, \psi]B\varphi \leftrightarrow (K\neg\psi \wedge B\varphi) \vee (\neg K\neg\psi \wedge B(\psi > [\curlywedge\, \psi]\varphi))$

Inference Rules

R. 1. From φ and $\varphi \to \psi$ infer ψ
R. 2. From φ infer $K\varphi$
R. 3. From θ infer $\varphi > \theta$

Theorem 2 (Soundness)**.** *The set of axioms and inference rules of DLCS is sound with respect to all DCM models.*

Proof. Let $M = (W, \{\leqslant_w\}_{w\in W}, \{\preccurlyeq_w\}_{w\in W}, v)$ be any dynamic counterfactual model and $w \in W$. Soundness of instances of propositional tautologies and for S5 axioms is straightforward. As for the rest:

Ax.1. Since $\min_{||\varphi||} \preceq_w \subseteq ||\varphi||$ by definition we have $M, w \vDash \varphi > \varphi$.

Ax.2. Let $M, w \vDash (\varphi > \psi \wedge \varphi > (\psi \to \theta))$. Then $M, w \vDash \varphi > \psi$ and $M, w \vDash \varphi > (\psi \to \theta))$. Then $\min_{||\varphi||} \preceq_w \subseteq ||\psi \to \theta||$. On the other hand, $\min_{||\varphi||} \preceq_w \subseteq ||\psi||$, so $\min_{||\varphi||} \preceq_w \subseteq ||\ \theta||$, and thus $M, w \vDash \varphi > \theta$.

Ax.3. By definition $M, w \vDash \varphi > \bot$ if and only if $\min_{||\varphi||} \preceq_w \subseteq \emptyset$ if and only if $\min_{||\varphi||} \preceq_w = \emptyset$ which by Definition 2[(ii)] is true if and only if $||\varphi|| = \emptyset$ if and only if $||\neg\varphi|| = W$ if and only if for all $v \in W$, $M, v \vDash \neg\varphi$ if and only if $M, w \vDash K\neg\varphi$.

Ax.4. Let $M, w \vDash (\varphi > \psi \wedge K(\varphi \leftrightarrow \theta))$. Then $M, w \vDash \varphi > \psi$ and $M, w \vDash K(\varphi \leftrightarrow \theta))$. Then for every $v \in W$, $M, v \vDash \varphi \leftrightarrow \theta$, and thus $||\varphi|| = ||\theta||$, and $\min_{||\varphi||} \preceq_w \subseteq ||\psi||$. Then $\min_{||\theta||} \preceq_w = \min_{||\varphi||} \preceq_w \subseteq ||\psi||$ and thus $M, w \vDash \theta > \psi$.

Ax.5. If $M, w \nvDash \varphi$ then $M, w \vDash \varphi \to ((\varphi > \psi) \leftrightarrow \psi)$. So suppose $w \in ||\varphi||$. Then $\min_{||\varphi||} \preceq_w = \{w\}$, and we have $M, w \vDash \varphi > \psi \iff \min_{||\varphi||} \preceq_w \subseteq ||\psi|| \iff w \in ||\psi|| \iff M, w \vDash \psi$.

Ax.6. Let $M, w \vDash (\varphi > \theta \wedge \psi > \theta)$. Then $M, w \vDash \varphi > \theta$ and $M, w \vDash \psi > \theta$. And thus $\min_{||\varphi||} \preceq_w \subseteq ||\theta||$, and $\min_{||\psi||} \preceq_w \subseteq ||\theta||$. Then we have $\min_{||\varphi\vee\psi||} \preceq_w \subseteq \min_{||\varphi||} \preceq_w \cup \min_{||\psi||} \preceq_w \subseteq ||\theta||$, and so $M, w \vDash (\varphi \vee \psi) > \theta$.

item[Ax.7.] Let $M, w \vDash (\varphi \vee \psi) > \theta$, and thus $\min_{||\varphi\vee\psi||} \preceq_w \subseteq ||\theta||$. Then if $||\varphi|| = \emptyset$ or $||\psi|| = \emptyset$ the result follows immediately. Otherwise, let $u \in \min_{||\varphi||} \preceq_w$ and $v \in \min_{||\psi||} \preceq_w$ and notice that either $u \preceq_w v$ in which case $\min_{||\varphi||} \preceq_w \subseteq \min_{||\varphi\vee\psi||} \preceq_w \subseteq ||\theta||$ and

$M, w \vDash \varphi > \theta$. Or $v \preceq_w u$ in which case $\min_{||\psi||} \preceq_w \subseteq \min_{||\varphi \vee \psi||} \preceq_w \subseteq ||\theta||$ and $M, w \vDash \psi > \theta$.

Ax.8. Let $M, w \vDash \varphi > \psi$ and $M, w \vDash \psi > \varphi$. Then $\min_{||\varphi||} \preceq_w \subseteq ||\psi||$ and $\min_{||\psi||} \preceq_w \subseteq ||\varphi||$ thus by Definition 3, we have $\min_{||\varphi||} \preceq_w = \min_{||\psi||} \preceq_w$ and thus for any θ, $\min_{||\varphi||} \preceq_w \subseteq ||\theta||$ if and only if $\min_{||\psi||} \preceq_w \subseteq ||\theta||$.

Ax.9. Since $\min \leqslant_w \subseteq W = ||\top||$ by definition we have $M, w \vDash B\top$.

Ax.10. If $M, w \vDash B\varphi$ then $\min \leqslant_w \subseteq ||\varphi||$ and thus $\min \leqslant_w \not\subseteq W \setminus ||\varphi|| = ||\neg\varphi||$ so $M, w \nvDash B\neg\varphi$ and thus $M, w \vDash \neg B\neg\varphi$.

Ax.11. We have $M, w \vDash (B\varphi \wedge B\theta)$ if and only if $M, w \vDash B\varphi$ and $M, w \vDash B\theta$ if and only if $\min \leqslant_w \subseteq ||\varphi||$ and $\min \leqslant_w \subseteq ||\theta||$ if and only $\min \leqslant_w \subseteq ||\varphi|| \cap ||\theta||$ if and only if $\min \leqslant_w \subseteq ||\varphi \wedge \theta||$ if and only if $M, w \vDash B(\varphi \wedge \theta)$.

Ax.12. Suppose $M, w \vDash B\varphi \wedge K(\varphi \to \theta)$ then since $M, w \vDash B\varphi$, we have $\min \leqslant_w \subseteq ||\varphi||$, and since $M, w \vDash K(\varphi \to \theta)$ we have $||\varphi|| \subseteq ||\theta||$. Thus $\min \leqslant_w \subseteq ||\theta||$ and thus $M, w \vDash B\theta$.

Ax.13. Let p be a propositional variable. Then $M, w \vDash [\curvearrowleft \psi]p \iff M^{\curvearrowleft\psi}, w \vDash p \iff w \in v(p) \iff M, w \vDash p$.

Ax.14. By definition $M, w \vDash [\curvearrowleft \psi]\neg\varphi \iff M^{\curvearrowleft\psi}, w \vDash \neg\varphi \iff M^{\curvearrowleft\psi}, w \nvDash \varphi \iff M, w \nvDash [\curvearrowleft \psi]\varphi \iff M, w \vDash \neg[\curvearrowleft \psi]\varphi$.

Ax.15. By definition $M, w \vDash [\curvearrowleft \psi](\varphi \wedge \theta) \iff M^{\curvearrowleft\psi}, w \vDash (\varphi \wedge \theta) \iff M^{\curvearrowleft\psi}, w \vDash \varphi$ & $M^{\curvearrowleft\psi}, w \vDash \theta \iff M, w \vDash [\curvearrowleft \psi]\varphi$ & $M, w \vDash [\curvearrowleft \psi]\theta \iff M, w \vDash [\curvearrowleft \psi]\varphi \wedge [\curvearrowleft \psi]\theta$.

Ax.16. Follows from the definition of the update rule, that ensures that model update by an inconsistent counterfactual supposition is the identity operation, and Proposition 3.

R1-2. Are straightforward.

R 3. Let $M = (W, \{\leqslant_w\}_{w \in W}, \{\preceq_w\}_{w \in W}, v)$ be any DCM model. Since θ is valid, for every $v \in W$ we have $M, v \vDash \theta$. Thus in particular, for every $w \in W$ and every $v \in \min_{||\varphi||} \preceq_w$, $M, v \vDash \theta$, and thus $M, w \vDash \varphi > \theta$. Since this is true for every world w in every model M, $\varphi > \theta$ is valid.

□

Definition 17. Let $M = (W, \{\leqslant_w\}_{w \in W}, \{\preceq_w\}_{w \in W}, v)$ be a DCM.

- We say that $v, w \in W$ are *modally indistinguishable* and write $w \leftrightarrows v$ if for every sentence $\varphi \in S\mathcal{L}_{DLCS}$,

$$M, w \vDash \varphi \iff M, v \vDash \varphi.$$

- We say that M is *deterministic* if for every $w \in W$, $\varphi \in S\mathcal{L}_{DLCS}$, and every $u, v \in \min_{||\varphi||} \preceq_w$, $u \leftrightarrows v$.

The next axiom characterises deterministic DCM models.

Ax. 17. $\neg K \neg \varphi \rightarrow (\neg(\varphi > \psi) \leftrightarrow \varphi > \neg\psi)$

Proposition 4. *Ax.17 is valid in a DC model, M, if and only if M is deterministic.*

Proof. Let M be deterministic, and assume $M, w \vDash \neg(\varphi > \psi)$. Then there is $v \in \min_{||\varphi||} \preceq_w \not\subseteq ||\psi||$. Since M is deterministic, for all $u \in \min_{||\varphi||} \preceq_w$, $u \notin ||\psi||$. Hence $\min_{||\varphi||} \preceq_w \subseteq W \setminus ||\psi|| = ||\neg\psi||$, and thus $M, w \vDash \varphi > \neg\psi$.

In the other direction, suppose M is not deterministic, and let w, φ be such that $u, v \in \min_{||\varphi||} \preceq_w$ and $u \not\leftrightarrows v$. Then there is a $\psi \in S\mathcal{L}_{DLCS}$ such that $M, v \vDash \psi$, and $M, u \vDash \neg\psi$. Then $M, w \vDash \neg(\varphi > \psi)$ but $M, w \nvDash \varphi > \neg\psi$. □

Axiom 17 says that for deterministic models we get a Commutativity with Negation principle: provided φ is possible, the negation of a whole counterfactual is equivalent to the counterfactual with negated consequent. Ramsey himself said that 'in a sense, "If p, q" and "If p, $\neg q$" are contradictories' (Ramsey, 1990, 155n). Dorothy Edgington suggested that 'A is to $\neg A$ as "If A, B" is to "If A, $\neg B$"', adding that "'It is not the case that if A, B" has no clear established sense distinguishable from this' (Edgington, 1995, 283). Our proof points out that for this to hold one needs to consider deterministic models.

The set of axioms Ax. 1-16, and inference rules R. 1-3, provide a proof system for DLCS:

Definition 18. For $\Gamma \subseteq S\mathcal{L}_{DLCS}$ and $\varphi \in S\mathcal{L}_{DLCS}$, we say that φ is provable from Γ, and write $\Gamma \vdash_{DLCS} \varphi$ if there is a sequence $\varphi_1, \ldots, \varphi_n$ of sentences of DLCS such that $\varphi_n = \varphi$ and each φ_i, $1 \leq i \leq n$ is either in Γ, or is an instance of an axiom, or can be derived from $\varphi_1, \ldots, \varphi_{i-1}$ using an inference rule. We say that φ is a theorem of DLCS, and write $\vdash \varphi$, if $\emptyset \vdash \varphi$.

Although we fall short of establishing completeness for the logic in this paper, there is some indication that the above axiomatisation is indeed complete. Remember that in our setting, the closeness relation between worlds is given by a divergence relation on the set of worlds from which we extract, for each w and φ, the set of closest φ-worlds to w. The main difficulty in getting completeness through standard techniques for constructing canonical models, is to define the appropriate divergence relation on the canonical model that satisfies the necessary constraints in Definition 2 and accommodates the necessary validities.

On the other hand, it is not too difficult to construct canonical models and prove completeness with respect to simplified models, in which the divergence relation is replaced with a selection function that specifies, for each world w and each φ, the set of closest φ-worlds to w without having to specify the full order relation.

A natural strategy is then to use models with selection functions as an intermediary step and construct DCM models from models with selection functions. Such a construction may be achievable by, for example, a technique similar to the one used in Leitgeb and Segerberg (2007), the details of which, of course, need to be worked out, which we leave for future work.

We now provide a few sample theorems of DLCS, just to give a flavour of the workings of the proof system.

Proposition 5. *The following are theorems of DLCS:*

(i) $(\varphi > \psi \wedge K(\psi \rightarrow \theta)) \rightarrow \varphi > \theta$
(ii) $(\varphi > \psi \wedge \varphi > \theta) \rightarrow (\varphi > (\psi \wedge \theta))$
(iii) $(\varphi > (\psi \wedge \theta)) \rightarrow (\varphi > \psi \wedge \varphi > \theta)$
(iv) $\varphi > \alpha$ *for any instance of a propositional tautology* α.
(v) $\varphi > \psi \rightarrow (\varphi \rightarrow \psi)$
(vi) $((\varphi > \psi) \wedge (\varphi \wedge \psi) > \theta) \rightarrow \varphi > \theta$

Proof.

(i) We show that $\varphi > \psi \wedge K(\psi \rightarrow \theta) \vdash \varphi > \theta$

(1) $\varphi > \psi \wedge K(\psi \rightarrow \theta)$	
(2) $\varphi > \psi$	$Prop.Tau.$ & MP
(3) $K(\psi \rightarrow \theta)$	$Prop.Tau.$ & MP
(4) $\psi \rightarrow \theta$	(3) & $S5Axioms$
(5) $\varphi > (\psi \rightarrow \theta)$	(4) & $R.3$
(6) $\varphi > \psi \wedge \varphi > (\psi \rightarrow \theta)$	(2) &(5) &$Prop.Tau$
(7) $\varphi > \theta$	(6) & $Ax.2$ & MP

(ii) We show that $\varphi > \psi \wedge \varphi > \theta \vdash \varphi > (\psi \wedge \theta)$

(1) $\varphi > \psi \wedge \varphi > \theta$	
(2) $\varphi > \psi$	$Prop.Tau.$ & MP
(3) $\varphi > \theta$	$Prop.Tau.$ & MP
(4) $\psi \rightarrow (\theta \rightarrow (\psi \wedge \theta))$	$Prop.tau$
(5) $K(\psi \rightarrow (\theta \rightarrow (\psi \wedge \theta)))$	(4) & $R.2$
(6) $\varphi > (\theta \rightarrow (\psi \wedge \theta))$	(2) & (5) & (i) & MP
(7) $\varphi > (\psi \wedge \theta)$	(3) &(6) & $Ax.2$ & MP

(iii) We show that $\varphi > (\psi \wedge \theta) \vdash \varphi > \psi \wedge \varphi > \theta$

(1) $\varphi > (\psi \wedge \theta)$	
(2) $(\psi \wedge \theta) \rightarrow \psi$	$Prop.Tau.$
(3) $(\psi \wedge \theta) \rightarrow \theta$	$Prop.Tau.$
(4) $K((\psi \wedge \theta) \rightarrow \psi)$	(2) & $R2$
(5) $K((\psi \wedge \theta) \rightarrow \theta)$	(3) & $R2$
(6) $\varphi > \psi$	(4) & (i) & MP
(7) $\varphi > \theta$	(5) & (i) & MP
(8) $\varphi > \psi \wedge \varphi > \theta$	(6) & (7) & $Prop.Tau.$ & MP

(iv)

(1) α	$Prop.Tau.$
(2) $\varphi \rightarrow \alpha$	$Prop.Tau.$
(3) $K(\varphi \rightarrow \alpha)$	(2) & $R2$
(4) $\varphi > \varphi$	$Ax.1.$
(5) $\varphi > \alpha$	(4) & (i)

(v)

(1) $\varphi > \psi$	
(2) $\varphi \to \varphi$	$Prop.Tau.$
(3) $\varphi \to (\varphi > \psi)$	(1) & $Prop.Tau.$ & MP
(4) $\varphi \to (\varphi > \psi \to \psi)$	$Ax.5.$ & $Prop.Tau.$ & MP
(5) $\varphi \to (\varphi \to (\varphi > \psi \to \psi))$	(4) & $Prop.Tau.$ & MP
(6) $\varphi \to (\varphi \wedge (\varphi \to (\varphi > \psi \to \psi)))$	(2) & (5) & $Prop.Tau.$ & MP
(7) $\varphi \to (\varphi > \psi \to \psi)$	(6) & $Prop.Tau.$ & MP
(8) $\varphi \to (\varphi > \psi \wedge (\varphi > \psi \to \psi))$	(1) & (7) & $Prop.Tau.$ & MP
(9) $\varphi \to \psi$	(8) & $Prop.Tau.$ & MP

(vi) We show $\varphi > \psi \wedge (\varphi \wedge \psi) > \theta \vdash \varphi > \theta$

(1) $\varphi > \psi \wedge (\varphi \wedge \psi) > \theta$	
(2) $\varphi > \psi$	(1) & $Prop.Tau.$ & MP
(3) $(\varphi \wedge \psi) > \theta$	(1) & $Prop.Tau.$ & MP
(4) $\varphi > \varphi$	$Ax.1$
(5) $\varphi > \varphi \wedge \psi$	(2) & (4) & $Prop.Tau$ & (ii)
(6) $\varphi \wedge \psi > \varphi \wedge \psi$	$Ax.1$
(7) $\varphi \wedge \psi \to \varphi$	$Prop.Tau.$
(8) $K(\varphi \wedge \psi \to \varphi)$	$R2$
(9) $\varphi \wedge \psi > \varphi$	(6) & (8) & (i)
(10) $\varphi > \theta$	(3) & (5) & (9) & $Ax.8$

□

We take the closest forerunner of our logic to be, to the best of our knowledge, the system presented in section 3.4 of Leitgeb and Segerberg (2007) ('L&S'), which inspired much of our DLCS. We therefore comment on some debts/similarities, and differences. Our formal language features, just as the language of the L&S logic, knowledge and belief operators, a counterfactual, and an update operator. Next, the L&S semantics for their language is based on a topology (U, T), where the items in U work essentially like possible worlds and T is a (Stone) topology on U. Not all sets of points from U count as 'propositions' in the setting, but just the sets which are clopen (both closed and open) in the topology, whereas the topologically closed sets are called 'theories'.

While the L&S setting is very elegant, it might be found a bit opaque. Our semantics for DLCS, while still recapturing KGM as shown, brings it closer

to the classical DEL semantics recapturing AGM as presented in Van Benthem and Smets (2015) and may look more transparent to readers familiar with the standard possible worlds semantics of modal-epistemic logics. E.g., while both we and L&S give standard Lewis-Stalnaker style truth conditions for the counterfactual conditional, L&S need to add a closure operation to the clause for the update operator (see their p. 182), whereas we don't.

7. Conclusion and Further Work

We have studied supposition as simulated belief revision in a qualitative setting. We have focused on counterfactual supposition, contrasting it with supposition 'in the indicative mood': we have come up with a counterfactual dynamics that, by recapturing KGM, complies with the four principles governing supposition in general, namely Certainty, Minimality, Consistency (in Comparisons), Possibility. We have then marked the difference between indicative and counterfactual supposition by establishing that the former, but not the latter, preserves certainties. We have introduced a Dynamic Logic of Counterfactual Supposition (DLCS): a language featuring knowledge and belief operators, a dynamic operator for counterfactual supposition, and a counterfactual conditional. We have then given a semantics for the DLCS language, out of which we have established a Ramsey Test: one comes to believe a counterfactual 'If it were that φ, it would be that ψ' when one comes to believe that ψ after dynamically revising one's beliefs on the counterfactual supposition of φ. We have provided a set of axioms for DLCS, for which we have proved soundness.

One area of further investigation is: belief in Q on the counterfactual supposition of P from section 4, $Bel_P(Q)$, is the counterpart of belief in Q on the indicative supposition of P, $Bel^P(Q)$, whereby one simulates updating one's beliefs on learning P with certainty. Are there natural counterfactual counterparts of indicative updates on (simulating) learning *soft* information, as per, e.g., radical-lexicographic and conservative upgrades? How would such 'soft' counterfactual updates look like?

Another area for further work is tied to our DLCS: coming up with a completeness proof for the proposed axiomatic system. As explained above, what so far has been standing in our way in the construction of the canonical model is the difficulty of coming up with a divergence relation that is a total well-order and satisfies the required conditions. The counterfactual supposition of P is more complex to deal with than its indicative counterpart because, in addition to involving plausibility, i.e., the notion of what situations an agent finds most plausible given P, it also involves precisely divergence, i.e., the notion of minimal departure required to make P true.

Work we have already carried out has to do with investigating counterfactual supposition as simulated revision of credences or degrees of belief. This will be the topic of our sequel: *Dynamics of Counterfactual Supposition II*.

8. Acknowledgments

Between 2023 and 2025, ideas included in this paper have been presented during talks at the Choice Group of the London School of Economics, at the conference *Fiction, Imagination, & Knowledge* at the University of Luxembourg, at the national conference of the Italian Society for Analytic Philosophy (SIFA) in Turin, at LOGOS-University of Barcelona, at the Logic & Interactive Rationality (LIRa) seminar, of the ILLC, University of Amsterdam, at the conference of the Dutch Research School of Philosophy (OZSW) in Delft, at the Stapledon Colloqium, University of Liverpool, at various seminars of the Arché Research Centre at the University of St Andrews. We are grateful to all audiences for useful questions and discussions. We would also like to thank the anonymous referees for a careful consideration of the paper and for several insightful suggestions and comments that improved the presentation of the paper considerably.

9. Funding

This research is funded by a Leverhulme Trust Research Project Grant RPG-2023-236, *What If ...? Knowing by Imagining* [WIKI]*: The Logic and Rationality of Imaginatio*n.

References

Adams, E. (1975). *The Logic of Conditionals*. Reidel, Dordrecht.

Alchourrón, C., Gärdenfors, P., and Makinson, D. (1985). On the logic of theory change: Partial meet functions for contraction and revision. *Journal of Symbolic Logic*, 50:510–30.

Badura, C. (2021). How imagination can justify. In Badura, C. and Kind, E., editors, *Epistemic Uses of Imagination*. Routledge, New York.

Baltag, A. and Smets, S. (2008). A qualitative theory of dynamic interactive belief revision. In Bonanno, G., van der Hoek, W., and Wooldridge, M., editors, *Logic and the Foundations of Game and Decision Theory*, pages 9–58. Amsterdam University Press, Amsterdam.

Bennett, J. (2003). *A Philosophical Guide to Conditionals*. Oxford University Press, Oxford.

Byrne, R. (2005). *The Rational Imagination. How People Create Alternatives to Reality*. MIT Press, Cambridge, MA.

Currie, G. and Ravenscroft, I. (2002). *Recreative Minds*. Oxford University Press, Oxford.

Edgington, D. (1995). On conditionals. *Mind*, 104:235–329.

Eva, B. and Hartmann, S. (2021). The logic of partial supposition. *Analysis*, 81:215–24.

Eva, B., Shear, T., and Fitelson, B. (2021). Four approaches to supposition. *Ergo*, 8:58–99.

Evans, J. and Over, D. (2004). *If*. Oxford University Press, Oxford.

Friedman, N. and Halpern, J. (1999). Modeling belief in dynamical systems, part ii: Revision and update. *Journal of Artificial Intelligence Research*, 10:117–67.

Gillies, A. (2004). Epistemic conditionals and conditional epistemics. *Noûs*, 38:585–616.

Goldman, A. (2002). *Simulating Minds*. Oxford University Press, Oxford.

Grahne, G. (1998). Updates and counterfactuals. *Journal of Logic and Computation*, 8:87–117.

Grove, A. (1988). Two modellings for theory change. *Journal of Philosophical Logic*, 17:157–70.

Joyce, J. (1999). *The Foundations of Causal Decision Theory*. Cambridge University Press, Cambridge.

Katsuno, H. and Mendelzon, A. (1992). Belief revision. In Gärdenfors, P., editor, *On the Difference Between Updating a Knowledge Base and Revising It*. Cambridge University Press, Cambridge.

Kind, A. (2016). Imagining under constraints. In Kind, A. and Kung, P., editors, *Knowledge through Imagination*, pages 145–59. Oxford University Press, Oxford.

Langland-Hassan, P. (2016). On choosing what to imagine. In Kind, A. and Kung, P., editors, *Knowledge through Imagination*, pages 61–84. Oxford University Press, Oxford.

Leitgeb, H. (2017). *The Stability of Belief. How Rational Belief Coheres with Probability*. Oxford University Press, Oxford.

Leitgeb, H. and Segerberg, K. (2007). Dynamic doxastic logic: Why, how, and where to? *Synthese*, 155:167–90.

Lewis, D. (1973). *Counterfactuals*. Blackwell, Oxford.

Lewis, D. (1976). Probabilities of conditionals and conditional probabilities. *Philosophical Review*, 95:581–9.

Ramsey, F. (1990). General propositions and causality. In Mellor, D., editor, *Philosophical Papers*, pages 145–63. Cambridge University Press, Cambridge.

Skyrms, B. (1987). Updating, supposing, and maxent. *Theory and Decision*, 22:225–46.

Stalnaker, R. (1968). A theory of conditionals. In Rescher, N., editor, *Studies in Logical Theory (American Philosophical Quarterly Monographs 2)*, pages 98–112. Blackwell, Oxford.

Van Benthem, J. (2011). *Logical Dynamics of Information and Interaction*. Cambridge University Press, Cambridge.

Van Benthem, J. and Smets, S. (2015). Dynamic logics of belief change. In Van Ditmarsch, H., Halpern, J., van der Hoek, W., and Kooi, B., editors,

Handbook of Epistemic Logic, pages 313–93. College Publications, London.

Van Ditmarsch, H., van der Hoek, W., and Kooi, B. (2008). *Dynamic Epistemic Logic*. Springer, Dordrecht.

Van Fraassen, B. (1980). Review of Brian Ellis, *Rational Belief Systems*. *Canadian Journal of Philosophy*, 10:497–511.

Van Leeuwen, N. (2016). The imaginative agent. In Kind, A. and Kung, P., editors, *Knowledge through Imagination*, pages 85–111. Oxford University Press, Oxford.

Wansing, H. (2017). Remarks on the logic of imagination. A step towards understanding doxastic control through imagination. *Synthese*, 194:2843–61.

Williamson, T. (2007). *The Philosophy of Philosophy*. Blackwell, Oxford.

Williamson, T. (2016). Knowing by imagining. In Kind, A. and Kung, P., editors, *Knowledge through Imagination*, pages 113–23. Oxford University Press, Oxford.

Williamson, T. (2020). *Suppose and Tell. The Semantics and Heuristics of Conditionals*. Oxford University Press, Oxford.

University of St Andrews and University of Amsterdam, ORCID: 0000-0003-3246-657X

University of St Andrews, ORCID: 0000-0001-5338-902X