

The Developing Group 11-12 September, 2021

Living in the Blend: Counterfactual Thinking

The topic for the next Developing Group workshop is *counterfactuals*. It may well be that you've not come across this term before, or only have a vague idea what they are. Penny's Texan mother used to regularly trot out this saying: "If the dog hadn't stopped to pee he would have caught the rabbit". It's only recently that we realised she was using a counterfactual to make her point – stop whinging about things that *didn't* happen. While they might seem like an obscure linguistic category, there is a lot more to counterfactuals than would first appear.

What are counterfactuals?

Let's start with a 'classic' (if limited) definition of counterfactual thinking:

Considering what would have been the result if events had happened in a different way to how they actually happened.

As modellers, the best way to comprehend counterfactuals is to see lots of examples. Let's start with four obvious examples (later we'll progress into more complex, compressed and nuanced forms):

1. If we had done more preparation it would have been a better presentation.
2. I could've got there earlier and then I'd have been in a right mess.
3. Missing that connection means I avoided the accident.
4. If I hadn't taken them out of the oven they would have been burned to a cinder.

These examples clearly have the classic structure:

If the world had/hadn't been the way it was,
then something better/worse would have happened.

The *If* part is the antecedent. The *then* part is the consequence.

Let's unpack what is involved in this kind of statement. Counterfactuals require thinking about something that happened, and then imagining an alternate reality and the consequences that would have followed from that. The imagined event or situation requires a mental simulation 'contrary to fact' and would have 'caused' the world to take a different pathway.

Our Purpose

Although counterfactuals mostly manifest in language, we are less interested in them as linguistic devices, and more interested in their cognitive construction; their effect on our lives; and how we can work with them in Symbolic Modelling. In other words:

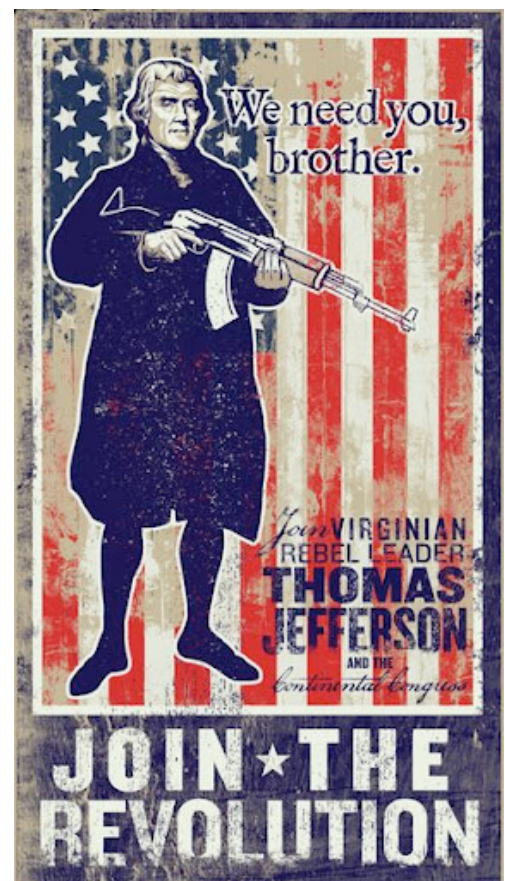
- How to recognise their various forms and, more important, how to 'grok' their nature (understand them at a deep intuitive mind-body level)?
- What do people have to do in their mind in order to say and comprehend these commonplace expressions – what is the cognitive structure of a counterfactual?
- What is the effect of 'living in the blend' of a counterfactual? How do we do that and how do we not? And what determines if that is problematic or productive?
- How do we model counterfactual thinking and how can that be applied in Symbolic Modelling?

Advertising and Films

Advertisers often use counterfactuals to make their point:



Does Advil really relieve being struck by lightning?



Did Thomas Jefferson carry an AK47 assault rifle?

And some films and TV series are entirely based on counterfactual thinking, e.g.

- *Sliding Doors* (Intermedia, 1998) wonders, what would have happened if a person does or doesn't get on an underground train. Watch the official trailer here: [youtube.com/watch?v=Da-Mizk86AE](https://www.youtube.com/watch?v=Da-Mizk86AE)
- *For All Mankind* (Apple TV, 2019). The series dramatises an alternate history depicting, "What if the Soviets had won the space race to the moon?" Official trailer at: [youtube.com/watch?v=HZS9M52Bd_w](https://www.youtube.com/watch?v=HZS9M52Bd_w)
- *The Man in the High Castle* (Amazon Prime, 2015) asks the viewer to consider life if the Allied powers had lost World War II. Official trailer here: [youtube.com/watch?v=zzayf9GpXCI](https://www.youtube.com/watch?v=zzayf9GpXCI)

It's worth watching the trailers and noting that while you are doing so, you are living in the counterfactual blend.

Living in the blend

One of the most sophisticated examinations of counterfactuals from a cognitive linguistics viewpoint is to be found in Gilles Fauconnier and Mark Turner's book, *The Way We Think* (2002). They use their "conceptual blending network model" to unpack counterfactuals (as well as metaphor and a host of other juicy topics). The examples in this section come from their book.

To give you an idea of how blending works, take two concepts: "a woman" and "a strong metal" and then blend them into the "the Iron Lady" (the name given to ex-British Prime Minister, Margaret Thatcher). Notice how this creates ideas that were not contained in either of the two original concepts.

A counterfactual *blends* a 'factual' event or situation with a 'counter-to-fact' feature which prompts emergent logic and entailments. Here's an example: a childminder says to a child in her care:

If I were your father I would smack you.

This is clearly a counterfactual since the childminder is not the child's father. It gets interesting when you try to figure out what message the childminder is trying to get across? Without the full context it's impossible to know, but some possible answers are:

- Your father should be stricter.
- You're lucky I'm not as strict as your father.
- The only reason I am not smacking you is because I am not allowed to.

The meaning of the sentence is not 'in the words', it relies on cognitive processes that make connections and inferences between real and imaginative situations prompted by the linguistic utterance.

Fauconnier and Turner coined the term "living in the blend" for when our conscious experience comes entirely from a blend and in those moments we have little awareness that our perception is based on a counter-to-fact premise. People can live in a counterfactual blend for anywhere between seconds and decades!

Good theatre seduces us to live in a counterfactual blend.

More subtle and compressed examples

Let's look at a few more examples that are not so obvious because they are more 'compressed'. In these examples more of the counterfactual structure is presupposed and has to be teased out by reverse engineering the statement.

Have a go at working out the counterfactuality of these four examples (our thoughts are at the end of these notes):¹

5. What a goal that would have been.
6. Not that it would have done any good if I'd said that.
7. What a shame we took the bet.
8. That operation saved her life.

Types of counterfactuals

Linguists classify counterfactuals in a number of ways. Two of the most common are to notice whether the imagined scenario:

Adds to, or removes something from the fact.

Leads to a *better* or *worse* outcome (than what actually happened).²

By combining these two distinctions we get four possibilities. The counterfactual mental simulation:

- a. *Adds something and the outcome is improved*
- b. *Adds something and the outcome is worse*
- c. *Removes something and the outcome is improved*
- d. *Removes something and the outcome is worse.*

For example:

- a. If I had had the vaccine I wouldn't be ill now.
- b. If I had had the vaccine I might have had a massive reaction.
- c. If I hadn't had the vaccine I'd still be talking to my family.
- d. If I hadn't had the vaccine I'd have caught Covid by now.

If you go back and look at the eight numbered sentences above you'll see:

- | | |
|-------|-------------------|
| 1 & 5 | Adding improves |
| 2 & 6 | Adding worsens |
| 3 & 7 | Removing improves |
| 4 & 8 | Removing worsens |

Productive and Problematic counterfactuals

Counterfactuals are neither inherently productive nor problematic – they are a neutral cognitive process. Like all human capacities, whether we regard a counterfactual as ‘positive’ or ‘negative’ depends on our perspective and the *effects* it has in the world.

Counterfactuals can have major advantages, for example they enable us to learn from the past. By imaging possibilities that *didn't* happen – what if we had done this, and what if we hadn't done that – we can better prepare ourselves for similar situations in the future.

Being able to recognise counterfactuals can give us choice about whether we want to join someone in their counterfactual world or not. If asked “Why didn't you [behaviour]?”, we can decide whether we want to live in their proposed blend or decline the offer with something like “Who knows?”.

A quintessential case of using a counterfactual to good effect is Albert Einstein imaging he was riding on a light-beam. Without this counterfactual he may never have come up with the theory of relativity!

Randomised control trials are another valuable use of counterfactuals.³

Counterfactuals can also have serious disadvantages when a person effectively lives in an unproductive blend. For example continually wishing the past could have been different – “If only I hadn't ...” or “If only I had ...”. It's not the counterfactual thinking per se that is problematic, it's the effect of repeating the thought process. This will involve what David Grove called ‘a replicating mechanism’.

We've noticed some people have great difficulty engaging in certain types of counterfactual thinking. They find it hard to live in the blend even for a few seconds. Often we think this is because they intuitively know that if they spend time considering potential alternatives they won't like the consequences! This might keep them from ‘feeling bad’ or making a difficult decision in the short term, but it can prevent them from learning and improving things in the long run.

Counterfactual and emotions

Certain emotions often require counterfactual thinking:

- guilt
- remorse
- regret
- resentment
- jealousy/envy
- certain kinds of anxiety
- blame
- and the many forms of self-reproach/recrimination.

And likewise:

- relief
- gratitude or feeling lucky for what didn't happen
- and hope for the future based on a different past

are also likely to require some counterfactual thinking.

Present-time counterfactuals:

So far we have concentrated on classic counterfactuals that imagine an alternate history but counterfactuals can also be about the present (and as you'll see, the future too):

- If I were you I'd be working on finishing my book.
- If D H Lawrence were alive today, he would be horrified by the amount of pornography openly on sale.
- I'd be much happier today if I'd had better parents.

And take a look in the pattern of these present-time counterfactuals:

- If it wasn't raining we'd be outside.
- There's no milk in the fridge.
- Where is the missing chair?
- He has a gap in his knowledge.
- She will not listen.

The last set are interesting because they are all comments about a no-thing: no rain, no milk, no chair, no knowledge, not listening. In order for these statements to make sense we need to imagine a situation that doesn't exist: dry weather, milk in the fridge, a chair, more knowledge, listening. Otherwise, how would we know they weren't there? Once you realise this, you'll see just how often negation sets up counterfactual thinking.

And Gregory Bateson had things to say about the effect of nothings, "In the world of mind, nothing – that which is not – can be a cause. The letter which you do not write can get an angry reply."⁴ Presumably the letter-writer's counterfactual thinking prompted their anger.

Future counterfactuals

Future-orientated counterfactuals are the most difficult to spot because what constitutes 'factual' is much more limited in the future than in the past. Many more things *could or might* happen in the future which means there are fewer things that we can be sure are 'facts'.

To maintain a clean stance, we need to know what are counter-to-facts *in the other person's worldview*. For example:

- I'm leaving tomorrow otherwise I could have come to the party.
- I don't care if it's harmless, I'll die if that spider comes anywhere near me.
- If only my father were alive, he'd be proud of what I am going to do.
- I would retire next year if I had more savings.
- Nothing can go wrong.

The Solution-Focused Therapy 'miracle question' relies on counterfactual thinking (some versions more than others):

Imagine that after you have gone to bed tonight, a miracle happens. But since you are asleep, you will not know that the miracle has happened and has completely solved your problem. When you wake up tomorrow morning what will be different that will let you know that a miracle has happened?

Conditionals and desired Outcomes

Counterfactuals are a particular subset of 'conditionals' which include a range of "possible/hypothetical situations and their possible/hypothetical consequences/outcomes".⁵ Conditionals are essentially what-if thinking.

A person who says "If I win the lottery, I will buy a Rolls-Royce" is not using a counterfactual since it is a hypothetical possibility. Whereas a person who says "I never buy lottery tickets but if I won the lottery, I would buy a Rolls-Royce" is speaking counterfactually since it is not possible to win if you don't buy a ticket.

All desired Outcomes are imaginative since they have yet to happen. Although most desired Outcomes and Remedies do not involve counterfactual thinking, some are very likely to, e.g.:

- I want our marriage to go back the way it was [before I had the affair].
- I want to always be 100% truthful.
- I only want to receive positive feedback.
- I want my husband to accept my mother exactly as she is.
- I don't want my children to feel the pain I did.
- I don't want to be the person I've become.

We're not saying people shouldn't have these outcomes, we're saying that if counterfactual thinking is involved, it could be useful for you to notice as part of supporting the client to self-model.

And, we have to be very careful about assuming what is and is not possible. Until very recently if a 70 year-old person said "I want to be an astronaut" you might think they were living in a counterfactual world – until Sir Richard Branson came along.

Experimental generalities

According to Byrne & McEleney (2000)⁶, when people think about how an outcome could have been different, they mentally 'undo' their representation of aspects of the factual situation in regular ways. For example, experiments show that *in general* people:

- Mentally undo the first cause in a causal sequence rather than subsequent causes
- Mentally undo the most recent event in a non-causal, temporal sequence rather than earlier events
- Mentally undo events under a person's voluntary control rather than events outside their control
- Judge that people will feel worse about their actions rather than their failures to act.

A neurological study using fMRI concluded that counterfactuals imply the relevance of keeping in mind both factual and supposed information whereas hypothetical conditionals imply that real world information is irrelevant to the processing and is omitted. The need to sustain representations of factual and suppositional events during counterfactual sentence processing requires increased mental imagery and integration efforts.⁷

Counterfactual metaphors or metaphorical counterfactuals?

Metaphors don't need to be counterfactuals and counterfactuals don't need to be metaphorical – but they can be.

Metaphors and counterfactuals both share a two-ness and can be described in similar terms of 'mapping across', or 'blending'. However they differ in that a counterfactual needs to have some kind of 'truth' or 'actuality' associated with it – metaphors don't.

The rest of this section is taken from Mark Lee (2006).⁸

Metaphors ask us to imagine the world in a new way,
while conditionals may ask us to imagine a whole new world.
Cohen (1998)

One-sided metaphorical counterfactuals are where either the antecedent or the consequence is metaphorical, but not both. For example:

- If John had put that idea to one side, then he would have been happy.
- If Bill had been late again, Cheryl would have blown her top.

Double-sided metaphorical counterfactuals feature a metaphor in both the antecedent and the consequence. There are weird examples like:

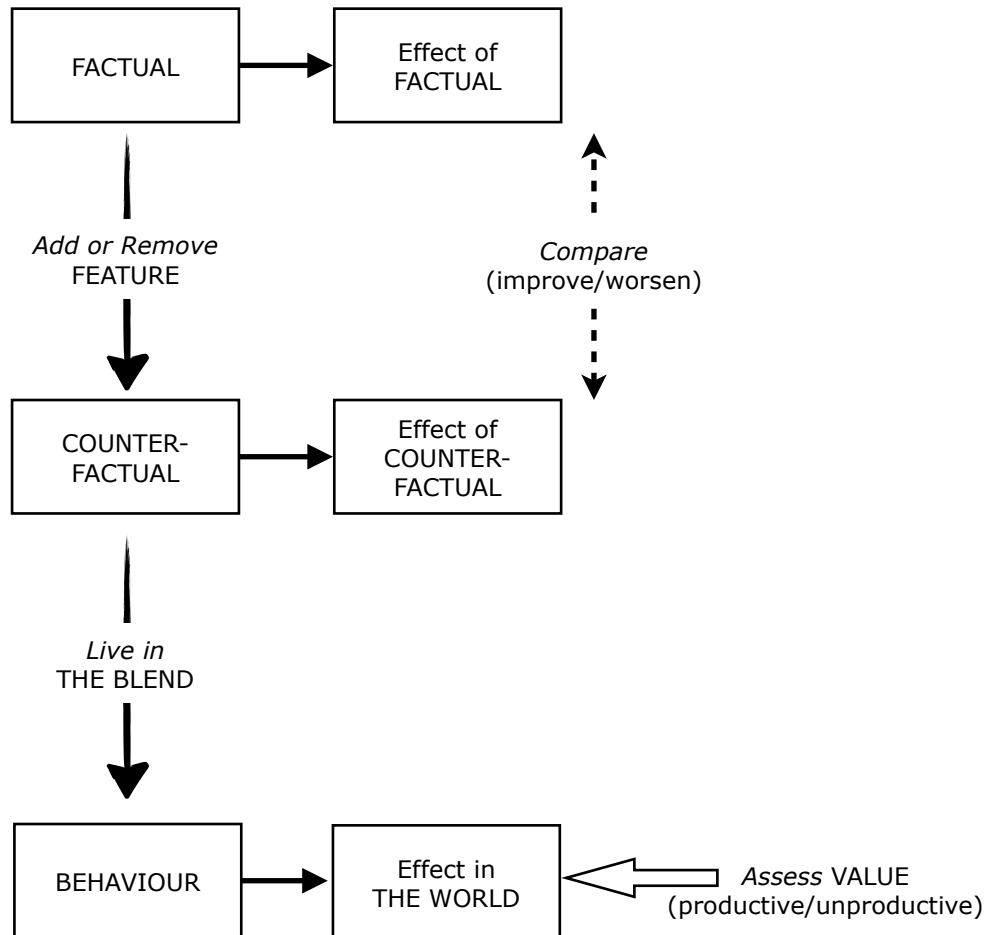
- If time was money, then the poor would be rich.
- If wishes were horses then beggars would ride.
- If war was a marketplace, then soldiers would be coins.

We can't recall coming across any like this in a session and so for practical purposes we can probably ignore them. Still rare but more likely would be something like:

- If John had put that idea to one side, he'd have egg on his face.
- If Bill had played loose with time again, Cheryl would have blown her top.

Our model

The following diagram summarises our model of the components and the mental actions (in *italics*) involved in creating a counterfactual, living in the blend for a while, and all of this having some effect on the person.



How and when an *individual* does any of this is where modelling comes in.

Concluding remarks

We hope the above will have prompted you to think in new ways. On the workshop we'll spend day-1 modelling counterfactual thinking and on day-2 we'll all look at how we can work cleanly with counterfactuals.

We don't know whether our investigation will turn out to be significant or an interesting side-shoot. We're trusting the group can help us find out.

Whatever happens we're sure that getting to grips with how people do this fascinating mental gymnastic will improve all of our modelling skills.

Notes

¹ Requires imagining:

5. The footballer *did* score.
6. I had 'said that' and it still would not have improved things.
7. If we hadn't 'taken the bet' we would still have our money.
8. She didn't have the operation and she died.

² It is important to distinguish this comparison from the assessment of whether the CF thinking is productive/unproductive, i.e. has real-world benefits/downsides (see the diagram on p.9)

³ Austin Frakt <http://theincidentaleconomist.com/wordpress/counterfactuals/>

When you want to know the causal effect of an intervention (policy change, medical treatment, whatever) on something, you need to compare two states of the world: the world in which the intervention occurred and the world in which it did not. The latter is the counterfactual world. Since most of us only get to live in one world (most of the time), observing the counterfactual is a rather tricky thing to do.

What we really want to know is how the world is different due to the intervention and only the intervention. A randomized controlled (or experimental) design is the gold standard approach. The assumption in that case is that, statistically, the two parts of the world (the treatment group and the control group — the counterfactual) are similar enough in all respects other than the intervention that comparing them gives you the true effect of the intervention. You've constructed a plausible counterfactual world. Good trick!

Next time you're involved in a policy debate, ask your opponent, what (s)he is taking as the counterfactual?

⁴ Bateson, G. (1972) Form, substance and difference. In *Steps to an Ecology of Mind*, pp.455-471. The essay is taken from Gregory Bateson's 1970 presentation at the Alfred Korzybski Memorial Lecture. <https://faculty.washington.edu/jernel/521/Form.htm>

⁵ This is a nice and clear site about conditionals and other vagaries of the English language <https://www.usingenglish.com/articles/conditional-sentences-in-english-1.html>

⁶ Byrne, R. M. J., & McEleney, A. (2000). Counterfactual thinking about actions and failures to act. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(5), 1318–1331. <https://doi.org/10.1037/0278-7393.26.5.1318>

⁷ Eugenia Kulakova, Markus Aichhorn, Matthias Schurz, Martin Kronbichler, Josef Perner, (2013) Processing counterfactual and hypothetical conditionals: An fMRI investigation, *NeuroImage*, 72:265-271, <https://www.sciencedirect.com/science/article/pii/S1053811913001079>

⁸ Lee, M.G. (2006). On Counterfactual Metaphors and Paradox. *The Baltic International Yearbook of Cognition, Logic and Communication - Vol. 1*, pp.95-109.