

Jokes and the Mechanics of Beliefs

Steffen Huck*

WZB Berlin Social Science Center

University College London

July 27, 2026

Abstract

I look at jokes as a transaction between two parties, a teller and a listener, in which the punchline performs an operation on the listener's beliefs. Leaning lightly on game theory, I argue that the majority of jokes fall into three different categories: (i) *collapse of beliefs*: a belief the listener has been led to hold shifts dramatically with the punchline; (ii) *salience of beliefs*: beliefs about different readings the listener held from the beginning become suddenly apparent; (iii) *common knowledge of beliefs*: privately held beliefs are revealed as shared. Some jokes are hybrids, drawing on more than one of these channels at once.

1 Introduction

This paper looks at jokes from a game theory inspired perspective and shows how many jokes play on beliefs and their manipulation. I offer a taxonomy distinguishing three categories. The first category plays on Bayesian manipulation: as the joke unfolds weights on different narratives shift and the final reveal shows the misguidedness of that updating. The second category leaves beliefs untouched but shifts attention from one part of its support to another. The third category leaves privately held beliefs also untouched but has a reveal that shows that one is not alone.

My taxonomy suggests that while jokes may appear on their surface infinitely varied, many of them operate on identical margins. Punchlines can overturn beliefs the listener has been nudged toward *collapse*. They can shift the listener's attention between different mass points

*I am grateful for comments from Rani Spiegler.

in beliefs already held: *salience*. Or they can take a belief the listener privately holds and make it *common knowledge*.

The next section takes the three in turn, with examples, asking in each case why the joke is funny and what its signature reveals about the belief operation beneath it, and also considers hybrids that draw on more than one channel at once. I then ask a second question that cuts across the taxonomy: beliefs about *what*?

2 The Taxonomy

A. Collapse of Beliefs

No other joke made me laugh more when I was a teenager. Here is its skeleton:

Two men buy a pair of horses and keep them together in one field. They fall to worrying that come next morning they will no longer know whose horse is whose. One of them proposes a solution: they will notch a horse's ear. By the next morning, both horses are notched—the second horse, it seems, has admired the first's new look and adopted it. Undeterred, the men dock a tail; come next morning both tails are docked alike, the horses having kept pace with one another's fashions. So it continues, the two men deliberating in good faith over each fresh measure, until, one morning, one of them turns to the other and says: "Why don't you take the black one, and I'll take the white?"

When a policeman colleague of my father told this joke to a bigger group I was part of, he took probably almost five minutes to tell it. And it became clearer and clearer that the two new horse owners faced a genuinely non-trivial problem. He embellished the joke quite a bit: the notching was actually done by biting but served, of course, just as distraction. The embellishment and the long-windedness (that not everyone can pull off, as I learned the hard way trying to re-tell the joke) simply aided the Bayesian updating where the state of the world in which the horses might have different colours, clearly present at the outset of a how-can-we-distinguish-two-horses problem, became ever less likely in our minds. Surely, already after the first turn, the first ear notched, nobody thought about it anymore, and the joke might work with one turn only. Its more elaborate version makes sure, however, that even the slightest traces of a prior that puts weight on different colours get eradicated.¹

¹For a masterclass in this technique of elongation, see Norm Macdonald's moth joke, as told to Conan O'Brien: minutes of Dostoevskian despair, delivered by a moth in a podiatrist's office, whose sole function is to bury the prior that this is a joke about a moth and a light.

When the punchline comes, the built-up belief collapses entirely. And that is very funny.

This type of joke, playing on the collapse of built-up belief, is one of the plainest things a punchline can do, and one of the most common: it makes the listener update, hard, away from a belief he had been led to hold.

Belief dynamics as a source of pleasure are, of course, not unknown to economics. Ely, Frankel, and Kamenica² model the demand for sports and fiction as preferences over the dynamics of beliefs: suspense is expected movement of the posterior, surprise its realised movement. The collapse jokes of this section clearly live in that neighbourhood—and yet a lurching posterior alone is not funny. Nobody laughs at a surprising stock price, and the fifth-set upset delights without a single chuckle. What the punchline adds is retrospective coherence: it does not merely move the posterior, it recodes the entire history of the joke, revealing that every detail was consistent all along with the reading the listener had discarded. The horses were black and white from the first sentence. Tennis offers no such recode: the surprise winner does not reveal that the earlier sets meant something else. Surprise is necessary for this category of jokes; it is the recoding that turns it into one.

Jokes do not have to be long, though, to work through this channel. The manipulation of beliefs and their subsequent collapse can come fast. Consider this one.

What is red and smells like blue paint?

Red paint.

The question places weight on a narrative—*this is a riddle, the colour and the smell are clues pointing to some third thing*—and the answer pulls that weight away. There is no third thing. The fact that it is paint was the answer; the colour was a mere distraction. It is the reverse image of the horse joke (where colour is the reveal) but operates through the same channel.

I posit that this is perhaps the largest category of jokes and add a few more examples here without explaining them further. Go figure!

One morning I shot an elephant in my pyjamas. How he got into my pyjamas I'll never know.

*“Doctor, will I be able to play the piano after the operation?” “Yes, of course.”
“Great. I never could before.”*

²J. Ely, A. Frankel, and E. Kamenica (2015), “Suspense and Surprise,” *Journal of Political Economy*, 123(1), 215–260.

B. Salience of Beliefs

While my first category of jokes involved a lot of Bayesian updating—first in a false direction, then with a dramatic reveal—my second category involves no updating at all. Rather, jokes in this category shift the salience of mass points in prior beliefs. What the punchline changes is not the distribution but which point in its support the listener is attending to. Consider this:

I told my wife she was drawing her eyebrows too high. She looked surprised.

The listener never revises anything. The eyebrows are high, and raised eyebrows equal surprise. The only thing that shifts is the attention, away from a purely physical reading to a reading that carries an emotional expression. The joke simply swings the listener's attention from one reading to another.

Two further examples make the same point.

A man walks into a library and asks for books about paranoia. The librarian whispers, "They're right behind you."

"They" is read first as the books and then as the persecutors the paranoiac imagines.

Parallel lines have so much in common. It's a shame they'll never meet.

"Have so much in common" is heard first in its social sense and then in its geometric one. Again no weight moves: both meanings were available throughout, and the punchline only redistributes the listener's attention across them.

It is worth stating the discriminant between this category and the previous one explicitly, for ambiguity alone does not settle it. Many ambiguity jokes are collapse jokes: in the piano joke, the recovery reading is killed by the punchline; it was simply wrong. What marks salience is that no reading dies. The eyebrows are drawn too high *and* the wife looks surprised; parallel lines have much in common in the social *and* the geometric sense. After the punchline all readings stand, co-valid and permanent. The test is thus not whether a sentence is ambiguous—ambiguity is a device available to both categories—but whether the punchline falsifies a reading or leaves them all intact. Lots of updating there, none here.

Jokes in the salience category can afford to be much shorter, as there is no belief manipulation required.

Hybrids of A and B

Some jokes do not sit in a single category alone but are really hybrids. Consider this one.

The doctor says, “I’m afraid I have bad news and worse news. The bad news is you have twenty-four hours to live.” The patient says, “That’s terrible! What could be worse than that?” The doctor says, “I’ve been trying to reach you since yesterday.”

There is a genuine collapse: the solemn frame of twenty-four hours to live collapses into you-will-die-now. The listener updates hard. But the abandoned reading is not simply discarded. It is retained as the term of a comparison: the punchline is funny only if the listener holds the counterfactual—the message received yesterday, a day of life intact—against the actual, you will die now. Two states are kept in mind at once, but not as two co-valid readings of one signal, the way salience requires; they are held in different statuses, one actual and one counterfactual. The joke is, thus, neither pure collapse—the overturned reading is not let go entirely, it lingers—nor is it pure salience, since real updating occurs.

A second, more famous case sits at the same boundary:

Why did the chicken cross the road? To get to the other side.

This can be read as a pure salience joke—keeping all options open, the other side in the literal and in the figurative sense. But, of course, deep down we know that the chicken didn’t quite make it to the other side of the road, so it also entails a dark collapse.

C. Common Knowledge of Beliefs

The third category leaves the listener’s private beliefs also untouched, but the punchline reveals that they are not purely private. It changes what the listener knows about *other* minds. This is the staple of much of observational humour. Here is an example:

Why do we press harder on the remote control when we already know the batteries are dead?

The listener learns nothing about himself; he knew he did this. What the joke does is make a privately held, slightly embarrassing belief *common knowledge*: he now knows that the teller does it too, and that everyone in the room does, and that everyone now knows that everyone knows. The laugh is the arrival at this shared knowing. Its content was never in doubt—only its publicity.

This is a kind of joke that benefits from the size of the audience. It does work with an audience of one, but, just as common knowledge becomes more powerful when there are more players, its achievement becomes funnier when the audience grows.

A similar trope is this:

Do you also speak louder when somebody doesn't understand your language?

In general, the common knowledge category is mostly associated with observational humour and stand-up comedy, which appears natural given the role of audience size.

3 Beliefs about What?

A second question cuts across the taxonomy: beliefs about *what*? In almost all of the examples above the beliefs concern an exogenous world—horses, paint, eyebrows, parallel lines—with teller and listener standing outside it as mere observers. But beliefs can also be about the endogenous world that includes the two of them. The common knowledge jokes of the previous section live there by construction: higher-order beliefs are beliefs about the minds in the room.

The other two operations can cross the line as well. Consider the cruelty routines of stand-up comedians such as Anthony Jeselnik or Ricky Gervais. The updating their punchlines trigger concerns, first, the teller himself: each reveal shows that the character delivering the joke is crueler than any prior admitted. But there is a second, subtler operation. The listener who laughs has produced evidence—about himself. His own laugh is data on his own penchant for cruelty, and he must update accordingly. The joke runs an experiment on its audience: the punchline elicits the reaction, and the reaction reveals the type. And in a full room the collective laughter makes the shared penchant common knowledge—the mechanism of the previous section, now with much darker content. This is collapse crossing the line: beliefs about the sender, about oneself, and about the room, all at once.

Salience can make the crossing, too. The mid-century American comedian Henny Youngman was the acknowledged master of the one-liner, and his signature line—so canonical that generations of self-deprecating jokes descend from it—is a case in point:

Take my wife... please.

Nothing is falsified. “Take” as in “take, for example” and “take” as in “take off my hands” both stand after the punchline. The shift is pure salience—but the reading that suddenly becomes salient is about the teller, his marriage, his despair. Much of self-deprecating humour lives

in this cell.

The line's origin adds a twist that is profound in its own right. Youngman, by his own account, was not joking when he first said it: he had brought his wife to a radio show and asked a stagehand to take her to a seat with that very line. It was not intended as a joke but next came a laugh from the stagehand and the salience shifted.

That shift happened entirely in the listener's head; the teller had supplied an ambiguity innocently and learned of it only from the laugh. This raises a small flag: most operations that this note studies are, at bottom, operations of reception—a joke can occur without anyone intending to make one. That this happened to a professional comedian, earnest for once, and handed him his signature line, is itself a joke of a hybrid variety, played by the world on him.

4 Conclusion

Jokes are widely various on their surface and, I have argued, considerably less various underneath. A punchline that lands does one, or a combination, of a small number of things to beliefs. It can force the listener to update, abandoning a narrative he had been led to build—*collapse*. It can move his attention across readings he held all along, without touching their weight—*salience*. Or it can take a belief he privately holds and make it shared—*common knowledge*. And each of these operations can be aimed at the exogenous world or at the very people in the room, the teller and the listener themselves.

I do not claim the three exhaust the field. I was just unable to find a good joke that would sit outside the proposed taxonomy.

Tell me one.