

Side-Splitter: A Visual Laboratory for Jokes with Verbal Humour and Physical Slapstick

Tony Veale

School of Computer Science,
University College Dublin,
Dublin D4, Ireland.
tony.veale@UCD.ie

Abstract

Humour encourages us to approach even the most serious subjects from a playful and non-serious perspective. It can soften our opposition to certain ideas and strengthen our rejection of others. It can bring us closer together, by dissolving the barriers that divide us, or it can use the same ridicule to drive us further apart. But a truly neutral comedian, or an automated joke generator, has no *in*-group loyalties to defend, nor any *out*-group opponents to resent. Comedy that lacks bite is asinine and bland, but a neutral comic can be funny and divisive by being equally cutting to all sides in a debate. This paper presents a demonstration of a system, *Side-Splitter*, that embodies this philosophy. The system uses the perspective-taking ability of LLMs to look at a given topic from each side, to see it as both a believer and a critic might, to generate divisive humour against either. *Side-Splitter* is a hybrid system that pairs a symbolic grammar with an LLM to produce a comic strip that is part visual slapstick and part stand-up routine. We describe how topical jokes are elicited from the LLM and selectively filtered according to how divisive they are, and show how these jokes are packaged in comic-strip form to lend them an extra dimension of visual humour.

Divided We Stand Up

It is often said that there are two sides to every story. Diversity of viewpoint is generally a good thing, but only if each side is willing to give the other an opportunity to speak. As impassioned believers become volatile partisans, digging in to further entrench their beliefs, discourse takes a back seat to difference, while echo chambers emerge in which each side only listens to the views of those that agree with them (Barberá et al. 2015). Rather than engage directly with those with whom they disagree, siloed partisans erect and knock down straw men that cannot fight back. Humour plays a part in this process, allowing each side to mock the other from the safety of their own assumptions and prejudices.

As noted in (Ford et al. 2017), humour can be used both as a social *lubricant* and as a social *abrasive*, to open groups to new members or to heighten inter-group tensions. In this dual capacity, humour can also bring both sides together, if only for them to laugh at each other, although it works best when it allows people to also laugh at themselves. Late night comedians and chat-show hosts now speak to, and for, in-

creasingly fragmented audiences. Whereas the hosts of old could reach a broad swathe of the population on a nightly basis, gently poking fun at both sides of an issue or a divide, the modern audience for these shows has fractured along political lines. Nonetheless, there is still an appetite for shows and for humour that sit in the middle and dish it out equally to both sides. Divisive humour is fun because it bites deeper than superficial wordplay. If used by each side it can deepen divides and lead to more, not less, entrenchment and polarization. But if used by a neutral figure in the middle, against both sides equally and unsparingly, it can expose each side to the views and assumptions of the other. The comedian who cries “A plague on both your houses!” can make each side see that theirs is not the only credible worldview.

It is hard to find such a truly neutral figure. Old-time American humorists such as Mark Twain, H.L. Mencken and Will Rogers expressed a general disdain for politicians of all stripes, but had clear political leanings of their own, from libertarian to Democrat to Republican. Their true loyalty, they claimed, was to the “common man”, and their claims to centrality were often grounded in a willingness to attack their own side. This is a tradition that is carried on by some modern political comics, such as Bill Maher. Maher’s show invites stakeholders from both sides of the political divide, taking no prisoners from either, and he burnishes his bona fides by rejecting neither out of hand. However, audience suspicions about the political leanings of a commentator now also extend to our large language models. Concerns that LLMs have been aligned with overly liberal viewpoints have led to the development of “anti-woke” models, such as Elon Musk’s *Grok* (Xiaoying Mei et al. 2026), that express more extreme views on politics and science. Culture wars are now fought over where we get our news, what TV shows we watch, and which LLMs we use. Still, when used in a balanced fashion, LLMs can serve as the neutral messenger that is accepted by both sides, or that both want to shoot.

This paper presents a demonstration of a humour system named *Side-Splitter* that is designed to be this neutral comic in the middle. It is bipartisan not because it never says anything critical of one side or the other but because it is equally vociferous in its biting jokes about each side of a divisive topic. Its jokes are directed, against one side or the other, but this direction is changeable and under the control of the system, which uses the perspective-taking capacity of an LLM

(Góes et al. 2023) to see a topic from the perspective of fans and critics alike. It is the measurable gulf between these perspectives that drives the LLM’s ability to filter good jokes, or at least provocative ones, from the weak tea that is the default humour output of most LLMs. To add to the humour and package it in an unserious and playful way, thus making it less threatening to those with opposing views, we define a visual grammar for slapstick that turns a flat set of jokes into a comical stand-up routine, that is, a stand-up routine presented as a comic strip. By generating a whole routine rather than a single joke on a topic, Side-Splitter increases its chances of making the user laugh at least one of its gags, while giving different users something different to laugh at.

Taking Sides in Joke Generation

LLMs are fluent and facile generators of many different kinds of content, but a biting sense of humour is rarely one of them. There are many factors that influence the success of a joke, from word choice and phrasing to timing and delivery. A good joke is a potent mix of opposites: of relevance and irrelevance, of sense and nonsense, expectation and surprise, and aptness and unfairness. Humour is not a unitary concept: it can take many forms and be produced by a wide variety of strategies. We built Side-Splitter as a laboratory in which to explore different forms and strategies, beginning with the classic short joke format of *setup + punchline*. This suits our current goal of rendering a comedy routine as a comic strip, since short setups and punchlines fit more naturally into text balloons. Moreover, this bipartite structure allows a setup to be introduced in one panel and its punchline in the next, creating a sense of flow through the comic.

We explore two joke-generation strategies here, for which the Side-Splitter architecture defines distinct prompt chains. These alternate strategies are interleaved: a batch of jokes is generated with the first strategy, and then selectively filtered, before a new batch is then generated with the next strategy and filtered. Generation proceeds in an alternating *generate-and-test* fashion until the required number of filtered jokes for a routine (10 by default) is obtained. The current strategies are *frame collapse*, a blend of superiority theory (Attardo 1994) and incongruity resolution theory (Veale 2004; Mihalcea, Strapparava, and Pulman 2010; Veale 2021), and *trumping*. We now describe each of these in turn.

We see the world as it is framed by language, by belief and by societal expectations. These conceptual frames are actively supported by some and tacitly accepted by others. For instance, president Donald Trump actively frames himself as a winner, a self-made billionaire, an outsider, and a saviour. These frames flatter him, and so are actively maintained by his followers and enablers. A critic aiming to poke fun at Mr. Trump can actively undermine these frames and cause them to collapse, perhaps by pointing out an internal inconsistency or a mismatch between belief and reality. Take the frame of self-made billionaire: we can collapse this frame by noting that inherited wealth gave Trump a considerable head-start in life. This is the specific strategy for our joke; we must now express the original frame as a setup which promises a compliment, and express the collapse as a punchline that delivers a put-down instead. Guided in the

right way by exemplar-laden prompts that describe the concept of frame collapse, an LLM (GPT-4.1) generates the following joke: “*How did Donald Trump climb his way to the top? He started at the top and climbed his way sideways.*”

Our second strategy, *trumping*, is named not for the person but for the idea of a “trump card.” (Veale, Feysaerts, and Brône 2006) define trumping as a form of adversarial humour in which a respondent punishes the lazy and unthinking use of a tired metaphor by a complacent speaker. For instance, an employee might claim to “*do the work of two people*” for a company, and his disgruntled boss might agree, albeit with the trump “*Yes, Laurel and Hardy.*” Likewise, the platitude “*think of it as a blessing in disguise*” is trumped with the comeback “*Well, it’s a bloody good disguise.*” Trumping takes the proffered metaphor and dismantles it from within, by identifying where it breaks down, or where it ambiguously admits a crushing counter-example. For our purposes, just one speaker will suffice. In the setup, the comedian echoes a common figure of speech that flatters the topic, before trumping this figure in the punchline.

The trumping prompt chain first explains the approach to the LLM with exemplars, and asks the LLM to suggest benign but trumpable metaphors that are often used for the topic. It is then tasked with expressing the benign view as a setup before delivering its trump card in the punchline. Suppose the topic is ex-Fox News host and now podcaster Tucker Carlson. A benign setup that frames how an ardent fan might see him is as a “*national treasure*”, but the LLM finds two ways of trumping this jaded idiom. In the first it replies with the punchline “*Just like those confederate statues everyone loves.*” In the second it notes, with some sharpness, “*There is a reason why some treasures are buried.*”

Humour is hard (Winters 2021; Jentzsch and Kersting 2023), and LLMs often deliver weak, sub-standard efforts. As a filter for the *generate-and-test* loop, we use a measure of subjective offense. The LLM first rates the offensiveness of a joke from the perspective of a believer in the topic, returning a number from 0 to 100. The LLM is then asked in a new context to rate the offensiveness of the same joke from the viewpoint of a critic. The difference between these two numbers is its *parallax rating*. To favour jokes with bite, the generator discards any with a parallax less than 25, ensuring that the jokes it accepts will create a rupture in world-view between the setup and the punchline. For instance, this joke about Elon Musk is accepted with a parallax rating of 55: “*How did Elon Musk become a self-made billionaire? He started from the bottom of his family’s emerald mine.*”

See You in The Funny Pages

Using cycles of generate-and-test with parallax filtering over its interleaved strategies, the generator ultimately produces a full routine on the chosen topic. The jokes that comprise this routine are then given a comic-strip form, with the setup introduced in one panel and the punchline delivered in the next. A common backdrop, the stage of a comedy club, is used throughout, and a single character, the comedian, delivers all of the jokes. The LLM has been instructed to generate jokes with short setups and snappy punchlines, so that each fits well into a single speech balloon. The LLM also gener-

ates an introductory remark on the topic, which the comedian delivers as he walks on stage, and a concluding remark, which the comedian delivers as he walks off stage at the end.

We use the *Excelsior* comics framework of (Veale 2022; 2023; 2024; 2025) to generate comics. This provides a range of pre-defined panel backdrops and character poses, and defines a comic via an XML specification that indicates the poses and backdrop to be used in each panel, as well as the content of any speech or thought bubbles. Because Side-Splitter is a laboratory for humour, we want to visually convey the parallax score of each joke in the panel that delivers its punchline, so as to record the LLM’s overall assessment of the joke within the comic. This record takes the visual form of audience laughter: for every 5 points of parallax in a joke, we add a floating “Ha” emanating from the bottom of the panel, where the audience is expected to reside. A weak joke, by the LLM’s own reckoning, will produce few titters, but a strong joke will produce many. Like the canned laughter in a sitcom, this also serves to nudge the reader toward laughter for the LLM’s best efforts. A complete example of a comic strip, with added laugh track, is presented in Fig. 1.

Excelsior pre-defines a wide range of visual assets: about 500 backdrops, of which we use just a few, and about 1000 character poses, of which we use a very broad palette. These poses offer heightened visualizations of emotions and lend themselves well to slapstick. They allow characters to laugh, cry, yell, sputter, rage and swoon, and to use props like guns, clocks, phones, cigars, crystal balls, wands and doves. Most props can be thrown or used comically as a weapon. Coherence across panels is important, so that a prop is typically introduced in one panel before it is used comically in the next. To maximize coherence, we define a visual grammar, named *Slapstick*, to relate poses to each other in this way, so that a pose used in one panel suggests a natural follow-on in the next. Because *Slapstick* emphasizes comedic effect, its pose connections are designed to punctuate a comic beat: an upbeat pose for the setup which introduces a joke suggests optimism or pride; the downbeat pose for the punchline that follows suggests disdain, surprise, or dramatic failure. Pose pairings that exploit the same prop are linked, so that a prop may be used in multiple ways over 2 or 3 successive jokes.

Slapstick generates the visual script for a comedy routine first, independent of any jokes that the comedian will eventually tell. This visual script leans heavily on physical humour and is designed to be funny in itself, regardless of the jokes that accompany it. This may still give rise to emergent connections between the visuals and the jokes in the final comic, since readers will naturally seek out the relevance of one to the other. We don’t wish to make one dependent on the other, but we do want the LLM to be inspired (that is, contextually primed) by the visuals as it generates its jokes. To this end, a visual script is generated first, for a fixed number of jokes (10 by default). The visible actions of the comedian in each beat are then verbalized as a text that is packed into the context given to the LLM when it is prompted to produce jokes.

Look Left, Look Right: Laughing at both sides

We use offensiveness as the basis for our parallax measurements because heavily-aligned LLMs are rather good at de-

tecting potential offense. It might seem more apt to base our parallax readings on the LLM’s ratings of the *funniness* of a joke, but LLMs are still poor judges of joke quality, and especially of their own efforts (Morain and Ventura 2025). While they can explain jokes well enough, they tend to over-rate the weakest and under-rate the strongest (Veale 2024).

Parallax should, in principle, cut both ways, against believers and critics alike, but in practice it favours the latter. The jokes in Fig. 1, for instance, all poke fun at Tesla and Elon Musk, and none make fun of their detractors. So, while parallax scores can be positive *and* negative, they are mostly positive: that is, a supporter of a topic is judged to be more offended by a joke than a critic. Of the 5000 jokes that Side-Splitter has so far generated and accepted, less than 5% have a negative parallax score. When generating a comic routine about *Guns*, for instance, most jokes are anti-gun, as in “*How do guns show you love your country? By proving you don’t trust anyone in it*” (parallax= 57). Only one joke has a negative parallax score (of -17): “*Why do people say guns keep you safe? Because bullets are faster than 9-1-1.*”

This imbalance is hardly surprising, given that the generation strategies *frame collapse* and *trumping* work best when toppling a powerful figure or undermining a common view. It is thus much more difficult to generate a truly balanced set of jokes for *and* against a given topic. But we find it can be more effective to change the topic than the strategy. For instance, to elicit a set of pro-Musk jokes, a generator can shift its focus to *NASA*, *General Motors* or *Sam Altman*. By asking the LLM to suggest mirror topics that invert the dynamic between critics and supporters of a topic, we can elicit new jokes with the same strategies that elevate, not diminish, the initial topic. If we appeal equally to the supporters of a topic, they may be more willing to engage with the criticism too.

Conclusions: A Visual Laboratory for Jokes

Side-Splitter is not designed to be a creator of jokes for their own sake, to be used e.g., by comedians (Toplyn 2014), writers, or public speakers. This kind of assistive application, as explored in (Toplyn 2021; 2022), has a rather clear use-case. Rather, we see it as a test-bed for humour generation that lets us test new generative strategies (beyond frame collapse and trumping) and other forms of audience parallax (beyond the merely offensive), and to explore new multi-modal forms for machine-crafted jokes (beyond the purely textual). A multi-modal packaging gives the generator more levers and pulleys to operate, and more dimensions in which to be funny, while directing the humour equally at supporters *and* critics alike.

The comics medium gives us a grounded (yet easily exaggerated) visual semantics that heightens the physicality and drama of a joke without requiring real physical action (Eisner 1985; McCloud 1993; Cohn 2013). The visual cohesion across comic panels also gives jokes a forward momentum that they lack in isolation, or in a simple joke list. The global structure of a comic, and of a comedy routine, now becomes a target of research in its own right, allowing us to study how comedy works at the level above that of individual jokes.

Acknowledgments

This paper and resource are wholly the work of the author.

References

- Attardo, S. 1994. *Linguistic theories of humor*. New York/Berlin: Mouton de Gruyter.
- Barberá, P.; Jost, J.; Nagler, J.; and Tucker, J. 2015. Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological Science* 26(10):1531–1542.
- Cohn, N. 2013. *The Visual Language of Comics: Introduction to the structure and cognition of sequential images*. London, UK: Bloomsbury.
- Eisner, W. 1985. *Comics & Sequential Art*. Tamarac, Florida: Poorhouse Press.
- Ford, T. E.; Breeden, C. J.; O’Connor, E. C.; and Banos, N. C. 2017. Jokes and humor in intergroup relations. In Powers, M., ed., *Oxford Research Encyclopedia of Communication*. Oxford University Press.
- Góes, L.; Sawicki, P.; Grzes, M.; Brown, D.; and Volpe, M. 2023. Is GPT-4 good enough to evaluate jokes? In *Proceedings of the 14th International Conference on Computational Creativity*.
- Jentzsch, S., and Kersting, K. 2023. ChatGPT is fun, but it is not funny! Humor is still challenging Large Language Models. *arXiv* 2306.04563.
- McCloud, S. 1993. *Understanding Comics: The Invisible Art*. New York, NY: Harper Collins.
- Mihalcea, R.; Strapparava, C.; and Pulman, S. 2010. Computational models for incongruity detection in humour. In *Proceedings of CICLing, Computational Linguistics and Intelligent Text Processing*, 364–374.
- Morain, R., and Ventura, D. 2025. Optimizing prompt engineering for creative performance. In *Proceedings of the 16th International Conference on Computational Creativity*.
- Toplyn, J. 2014. *Comedy Writing for Late-Night TV*. New York, NY: Twenty Lane Media.
- Toplyn, J. 2021. Witscript: A system for generating improvised jokes in a conversation. In *Proceedings of the ICC-21, the 12th International Conference on Computational Creativity*, 22–31.
- Toplyn, J. 2022. Witscript 2: A system for generating improvised jokes without wordplay. In *Proceedings of the ICC-22, the 13th International Conference on Computational Creativity*.
- Veale, T.; Feyaerts, K.; and Brône, G. 2006. The cognitive mechanisms of adversarial humor. *Humor: International Journal of Humor Research* 19(3):305–339.
- Veale, T. 2004. Incongruity in humor: Root-cause or epiphenomenon? *Humor: International Journal of Humor Research* 17(4):419–428.
- Veale, T. 2021. *Your Wit Is My Command: Building AIs with a Sense of Humor*. Cambridge, MA: MIT Press.
- Veale, T. 2022. Two-fisted comics generation: Comics as a medium and as a representation for creative meanings. In *Proceedings of ICC-22, the 13th International Conference on Computational Creativity*. 59–66.
- Veale, T. 2023. Have I Got Views For You! Generating “Fair and Balanced” Interventions into Online Debates. In *Proceedings of ICC-23, the 14th International Conference on Computational Creativity*.
- Veale, T. 2024. From symbolic caterpillars to stochastic butterflies: Case studies in re-implementing creative systems with LLMs. In *Proceedings of the 15th International Conference on Computational Creativity (ICCC 2024)*.
- Veale, T. 2025. Flipping the script: Complementary modes of story generation in comics. In *Proceedings of Text2Story, the 8th Workshop on Narrative Extraction From Texts*, volume 3964 of *CEUR Workshop Proceedings*, 039–048.
- Winters, T. 2021. Computers learning humor is no joke. *Harvard Data Science Review* 3(2).
- Xiaoying Mei, K.; Wolfe, R.; Weber, N.; and Saveski, M. 2026. Grok in the wild: Characterizing the roles and uses of Large Language Models on social media. *arXiv* 2602.11286.

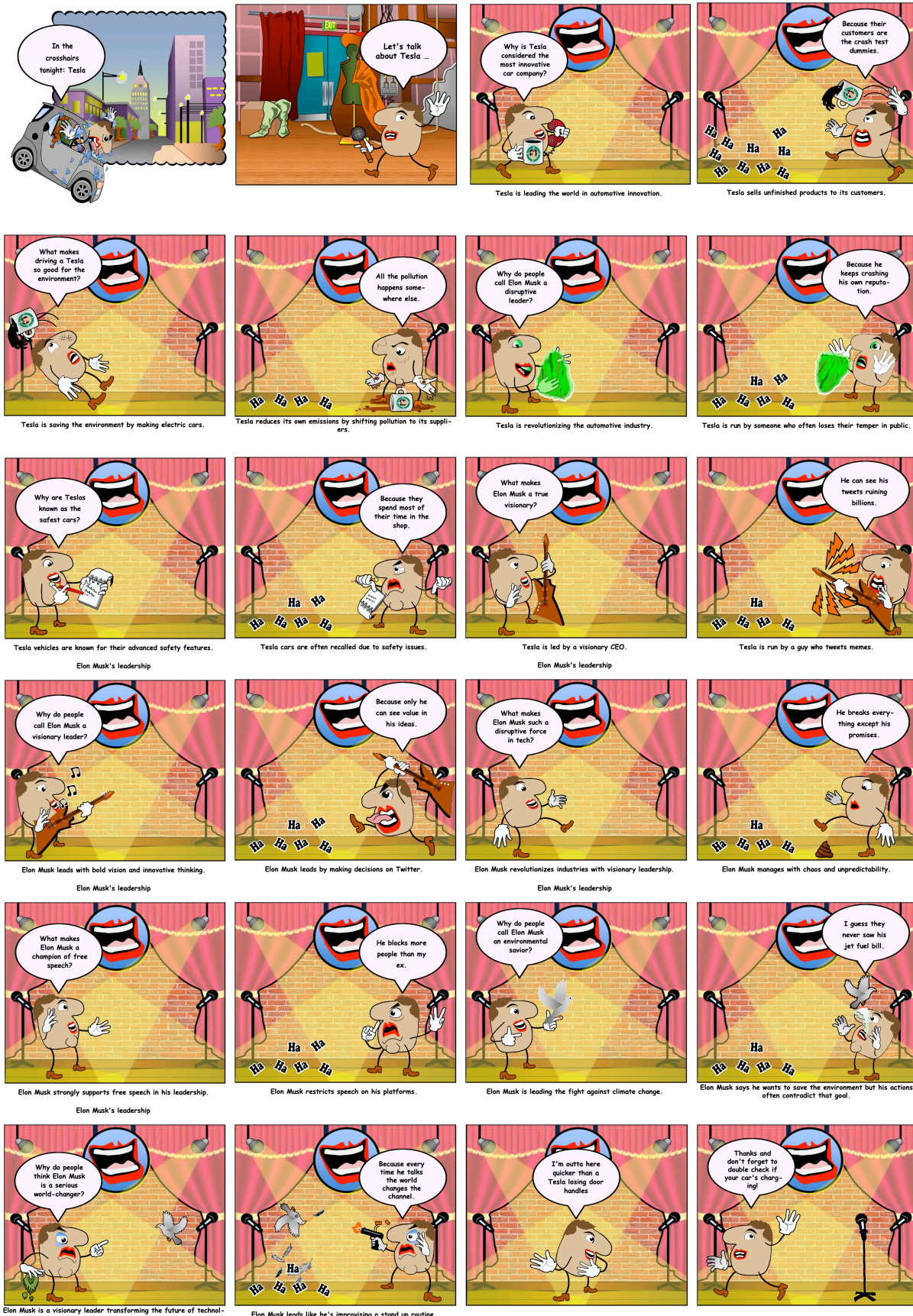


Figure 1: A Side-Splitter comedy routine about *Tesla*, rendered as a comic strip using Slapstick.