Multimodal simile
The “when” meme in social media discourse*

Adrian Lou
University of British Columbia

This paper analyzes the “when” meme, a popular internet meme, which prototypically juxtaposes a *when* clause with an ostensibly unrelated image. Despite the initial incongruity, I contend this image prompts selective mapping between verbal and visual elements to produce a multimodal simile. First, I attempt to define and more clearly distinguish simile from metaphor. Second, I show how this multimodal simile exhibits unique viewpoint mapping by prompting audiences to subsume viewpoints that are both unfamiliar and bizarre. Third, I connect the *like* construction in simile with the *like* reported speech marker to show how both concepts are intimately related. Ultimately, the paper seeks to contribute to studies of simile by bolstering its ties with multimodality, blending, metonymy, viewpoint, and embodiment.

**Keywords:** cognitive linguistics, internet memes, multimodality, simile, viewpoint

1. **Introduction**

The revolutionary effects social networking websites have had on communication are undeniable. In amplifying the speed at which we can now consume massive amounts of data and information, social networks have altered how media is produced and presented. Over the past few years, these global networking platforms, like Facebook, Twitter and Instagram, have become Janus-faced technological tools, often regarded as a democratic conduit that empowers political movements

---

* I would like to thank Barbara Dancygier for encouraging me to write this paper. It was Barbara who suggested I should be thinking about “when” memes in terms of multimodal simile (not metaphor). Without her help and insight, I would not have been able to finish this paper. Furthermore, Lieven Vandelanotte, who carefully reviewed every sentence of this paper, and the two anonymous reviewers, who provided invaluable comments, also deserve thanks. Any errors remaining are solely my own.
and humanitarian campaigns, and simultaneously as a lowbrow medium that sensationalizes trivial matters. Communication and discourse scholars have been intrigued by the potential these websites and apps have to truly effect sociopolitical change (see Tannen & Trester 2013). Yet, while many scholars are scrutinizing the broader implications of online viral trends and the impact they have on public discourse, few studies have probed deeper into the ways in which their generation and transmission shape our understanding and conceptualization of language and images. This paper targets these particular questions by analyzing the figurative language and cognitive properties of internet memes, one of the most popular forms of social media communication.

The term *internet meme* has been used rather loosely to label any digital artifact (e.g. text, video, photo, etc.) that is gradually modified and transformed by internet users when shared and transmitted online. Internet memes can come in a variety of forms, including hashtags (e.g. #throwbackthursday, #yesallwomen), video challenges (e.g. the ALS ice bucket challenge, mannequin challenge), and online characters (e.g. Grumpy Cat, Harambe the gorilla). A defining characteristic of all these internet memes is their template-like form, which gives users a pre-existing mold to express new thoughts and ideas in a familiar way. For instance, the Grumpy Cat meme, a photograph of a cat that has a very pronounced frown, is often used online by users to convey anger, resentment and misanthropy. Users usually superimpose short lines of text onto the image to better fit the meme into the discursive context in which it is used. This kind of verbal-visual meme, commonly referred to as the image macro, could be seen as the prototypical form of an internet meme. Used most frequently for humorous purposes, the image macro has cemented itself as a ubiquitous part of social media discourse with millions of social media users, ranging from the everyday teenager to celebrities and government officials, using the image macro form to communicate. Because adding text onto an image is becoming easier to do on digital devices, this memetic form has proliferated significantly online.

In adopting a cognitive linguistic approach to analyzing multimodal artifacts, this paper participates in an ongoing discussion on multimodal discourse by joining other cognitive linguists (Dancygier & Vandelanotte 2016, submitted) who view internet memes as a novel communicative method that offers new insight into the relationship between figurative language, viewpoint and embodiment.1 This

---

1. I have intentionally avoided viewing multimodal similes through the lens of construction grammar. This is one of the primary frameworks Dancygier & Vandelanotte (2016, submitted) applied in their analysis of internet memes; but to avoid complicating and convoluting the paper’s main objective, I decided to shift the focus away from the constructional characteristics of the “when” meme.
paper examines a very pervasive image macro, the “when” meme, which contains one verbal element, a *when* clause, and an ostensibly irrelevant visual image that is typically featured below the text. What makes this multimodal artifact unique is that the visual image, despite appearing unrelated, completes the *when* clause by prompting viewers to negotiate the incongruity of the visual element and to select meaningful aspects of the image to retain and make sense of.

As an initial example, consider Figure 1, which exhibits many of the generic components of a “when” meme:

![Figure 1. The “iMessage” meme](image)

In Figure 1, we see the text at the top, the *when* clause, which reads, “When your iMessage isn’t working and you have to send texts”, and a picture of an Amish family on a horse-drawn carriage placed below it. Despite the verbal-visual incongruity (that is, what is said has little to do with what is visually shown), the “when” meme can still be easily processed. Making our interpretation of the multimodal artifact possible is the image’s vivid evocation of certain frames. A frame (see Fillmore 1982; Sullivan 2013) is a set of ideas that can be elicited by presenting one related unit within that set of ideas. For example, uttering the term *waiter* in a conversation evokes a restaurant frame and makes other terms like *food* and *table* mentally accessible and relevant. Images, of course, can evoke frames too. Figure 1’s photograph conjures up an Amish life frame, one that is associated with primitive technology and a slow-paced lifestyle. Reconciling the verbal and visual incongruity requires audiences to map the characteristics of this frame onto what is being said in the verbal element of the meme. Thus, we can perhaps make a connection between the slowness of a horse-drawn carriage and the slowness of sending SMS texts compared to sending iMessages, a messaging application that relies on an internet connection and not a cellular signal.
It is possible to initially conceive of the “when” meme as a type of multimodal metaphor. I define multimodal metaphors as creatives blends that integrate components from two different domains (source and target) to form a final product that is presented through multiple modalities. This theory of conceptual blending comes from Gilles Fauconnier & Mark Turner (2002), whose framework has been applied to unpack the cognitive ways human beings make, and make sense of, creative expressions. In their elucidation of what they call the single-scope blend, Fauconnier & Turner explain that the single-scope blend mirrors the operations of “conventional source-target metaphors” (2002: 127). They give the example of the Boxing CEOs, where two competing businesses are metaphorically described as CEOs exchanging punches until the stronger one knocks the other one out. Here, the source domain, boxing, supplies the terms to construe the target domain, business. While this metaphorical scenario is not multimodal, we can easily imagine it rendered as a visual-verbal comic: the illustration depicts the CEOs in a boxing ring with the logo of their company printed on their trunks to indicate their affiliation.

What differentiates Figure 1 and other “when” memes from single-scope blends is that “when” memes do not fully integrate the source and target, providing only cues as to how the artifact should be figuratively interpreted. A metaphorical version of Figure 1, and therefore a more conventionally-looking visual metaphor, would incorporate elements of, say, iMessage or SMS in the picture. Because there is no final image with horse-drawn carriage and SMS blended, I argue that the meme exhibits the qualities of a simile, and not a metaphor, since the meme suggests that sending SMS texts feels like riding a horse-drawn carriage in a world filled with mechanical vehicles. As we will see throughout this paper, the visual inputs (that is, the main image) of “when” memes function in the same way as like or as, used conventionally to form similes.

Though a widely recognized trope throughout the history of rhetoric, simile has been relatively understudied by cognitive linguists and often simplified and relegated as a subtype of metaphor. My exploration of the “when” meme attempts to bolster recent studies of simile (Israel et al. 2004; Moder 2008, 2010; Dancygier & Sweetser 2014) by not only highlighting its pervasiveness in online discourse, but also explicating the defining cognitive qualities of simile through examining multimodal variations of this typically verbal trope. Engaging with and expanding on research on multimodal metaphors (Forceville & Urios-Aparisi 2009), the paper posits that the “when” meme operates, most commonly, as a type of multimodal simile with the visual image supplying context and content that audiences must selectively integrate with the textual parts of the meme. The goal of this paper, moreover, is to investigate multiple types of the “when” meme, to categorize the visual inputs used, and to document and explain how these multimodal similes
rely on – and in turn illuminate – frame metonymy, conceptual blending, viewpoint, embodiment and reported speech.

The paper begins by briefly surveying the history of this meme and suggesting that it originated first from a purely verbal construction known as the “awkward moment”. In tracking the evolutionary history of the meme, I suggest that the “when” meme is an exemplary case in which the advancement of digital communicative tools has successfully enabled the supplanting of verbal forms with visual ones to form ideal instances of multimodal communication. I then proceed to build a taxonomy of “when” memes, contending that each one depends on a range of figurative tropes and cognitive processes. This taxonomy includes four different variations of the “when” meme, distinguished according to the image’s level of accessibility. I draw inspiration from Ariel’s (1988) model, which gauges accessibility of discourse referents both grammatically and pragmatically, to delineate the accessibility of the meme’s visual inputs. Beginning from one end of this spectrum, the first “when” meme is the least accessible of the four, requiring audiences to be aware of very specific cultural allusions that are made accessible by a very specific visual frame. The second type, slightly more accessible than the first, has a visual input that contains images requiring less cultural knowledge. These visuals, instead, highlight a prominent perceptual pattern. Building on type two, the third type of “when” meme uses facial features and physical gestures as the primary visual input. I connect the image of facial expressions to the mimetic performances seen commonly in reported speech following the like discourse marker. Forging this connection, I intend to delineate the relationship between like as a reported speech marker and as a signifier of simile to further highlight simile as a trope not only distinctly different from metaphor but also as a trope fundamentally tied to gestures and embodiment. Lastly, playing with the conventions of the form, the fourth “when” meme demonstrates awareness of the form and consequently exaggerates the ridiculousness of the simile’s selective mapping.

2. A short history of the “when” meme

The phenomenal proliferation of internet memes has started to generate serious academic interest in the process by which cultural ideas propagate and take new forms (see Shifman 2014; Wiggins & Bowers 2015). Many existing studies, unsurprisingly, have probed the internet meme from the theoretical angle that shares its name: memetics. The memetic approach to internet memes is informed by its original definition first proposed by Richard Dawkins, who coined the term in his 1976 book The Selfish Gene. For Dawkins, the meme was meant to be the cultural equivalent of a gene – a tiny slice of information or data that spreads itself from
person to person. This gene analogy was intended to illustrate the process of biological evolution. Like phenotypical characteristics of organisms, which would be modified by the conditions of the natural environment, the meme is shaped, refined, or even removed, at the discretion of people, their society, and their culture.

There is room for a memetic approach to studying the “when” meme in that such analysis might yield interesting findings about how social media users and platforms collectively facilitate the development of a mutually intelligible lexicon of phrases and jokes. In this section I attempt to uncover a possible evolutionary trajectory of this “when” meme, arguing that it has, only recently, fully congealed to form a new simile expression functionally comparable to using the conventional simile signifiers, *like* or *as*. There are some prime candidates that can be labeled as the progenitor of the “when” meme, the clearest one being the “awkward moment” meme that rose to prominence in 2009. While not an academically reliable source, the database www.knowyourmeme.com traces the origins and use of memes. According to this site, which investigates the popularity of the meme as a search term on Google, the “awkward moment” meme usage peaked in 2011.

The “awkward moment”, as the name suggests, can be read as a linguistic construction that allows social media users to speak directly about awkward encounters and experiences of daily life. These are jokes akin to ones made by observational comedians. Here are some examples:

1. That awkward moment when you walk past someone you met once and you don't know if you should say hi.
2. That awkward moment when you’re trying to get over someone you never even dated.

The survival and propagation of a meme, based on the core principles of memetics, is dependent on the meme’s simplicity, functionality and the extent to which it can be easily transmitted. Memetically speaking then, the “awkward moment”, similar to a knock-knock joke (a cultural meme), fulfills these criteria as it has a distinct form that lucidly signifies what type of humorous utterance is about to be made and presents a formulaic linguistic pattern where new information can be entered. The subject matter that the meme deals with, moreover, fittingly complements the quotidian topics expected on social media.

From the perspective of Mental Space Theory (Fauconnier 1994), the initial part of the meme, “That awkward moment when”, can be understood as a mental space builder that sets up the primary mental space where subsequent ideas are situated. The adjective *awkward* specifies the way the mental space is meant to be construed. Because the meme typically uses the pronoun *you*, the viewpoint is projected towards the listeners. Crucially, the listeners are positioned in the primary
mental space, which means that they are supposed to be the ones experiencing the awkward moment and the ones coming to the realization that what is being depicted in these lines is indeed awkward. Though the viewpoint manipulation of the “awkward moment” meme is quite a complex cognitive feat, it is not new in any way or unique to this construction. As suggested earlier, many observational comedians would refer to such experiences by using you as a means of prompting audiences to enter these mental representations of social situations.

However, the main limitation of the “awkward moment” meme is that it exclusively focuses on bizarre, though relatable, life experiences. Naturally, as the meme spread online, it was modified to describe other social situations. The novel constructions that formed saw the removal of the awkward, spawning a new variant: the “that moment when” meme. Without awkward, the direct and unambiguous indication of the original meme’s tone disappeared with it. Without the adjective, no longer could audiences identify immediately what type of feeling the textual voice is attempting to articulate.

Figure 2. The “moment when” meme

The deletion of the word awkward, however, did not restrict what can be said; rather the new memes lent themselves to be supplemented with visuals that work to illustrate the expressed feelings set up by the textual voice. The images of the “moment when” meme, which have supplanted the awkward, can be fairly iconic
pictorial representations of emotions and feelings. Figure 2 has a cartoon image of a man with a facial expression that is normally associated with sadness and disappointment: the man is looking down and has a visible frown on his face. Even without explicit signals or references, viewers of the meme are able to recognize that the face here belongs to the textual voice and not to the interlocutors to whom the voice is telling the story. We process the image as a visual depiction of the voice’s inner feelings and the way he/she might look in that instance when the story being told fails to incite laughter. What’s more, viewers might be able to see themselves in the meme too and can align their own viewpoints with the meme. The incorporation of visuals to complete the text substantially complicates the meme and paves the way for it to become less accessible and thus more creative. Figure 2 is, to all intents and purposes, literal because the face presented resembles a facial and physiological display of sadness. Later, when I explore different variations of the “when” meme, we will see (in Sections 4 and 6 respectively) how the accessibility of the visual inputs diverges, with some becoming less iconic and more obscure, and others retaining iconicity but simultaneously displaying a strong sense of unfamiliarity and strangeness.

3. The “when” meme as a multimodal simile

In spite of its ubiquity in discourse, simile has not attracted the same scholarly attention as metaphor. In the history of rhetoric, simile, very early on, was likened to metaphor by Aristotle, perhaps marginalizing its uniqueness and its importance (1960: 192). More nuanced rhetorical definitions that arose much later detailed the selective mapping process of simile, though not in those terms. The seminal 20th century rhetorical treatise The New Rhetoric notes that similes “seek to transfer to the person to whom they are applied some of the characteristic quality of the chosen illustration” (Perelman & Olbrechts-Tyteca 1969: 362). Cognitive linguists have recently made greater strides in differentiating simile from metaphor. Dancygier & Sweetser argue that metaphors and similes have “different patterns of mapping” (2014: 138). As suggested earlier, multiple aspects of a metaphor’s source domain (e.g. BOXING) are mapped onto the target domain (e.g. BUSINESS), forming a blended structure where these projected elements interact (e.g. the CEOs, who are the physical embodiment of their company, attempt to physically defeat their opponent). Critically, simile only “maps specific attributes” onto the target domain (Dancygier & Sweetser 2014: 138). Consider this verbal simile:

(3) The car sped through the intersection like a bullet.
While the mapping between target and source, at first, resembles a single-scope blend, we will notice that only a few selected qualities of the bullet are projected onto the final blended space. The larger shooting/gun frame is not wholly present, and so the simile does not intuitively evoke related concepts within a larger organizing frame. That is, we are not prompted to conceive of the driver as a shooter or the destination as a bullseye or a target. Rather, only the speed of the bullet and its straight and precise trajectory are blended with the car. As Gentner & Bowdle (2001: 232) put it, “similes [are] more likely than metaphors to elicit potential similarities between unrelated targets and base”. By presenting two dissimilar domains, similes then induce us to discover potentially similar qualities (Israel et al. 2004). As such, we can begin to see the key mechanisms that separate metaphor from simile, and the ways in which a simile’s selective mapping lends itself to be structured in the image-macro form.

Dancygier & Sweetser (2014) put forth the term “limited-scope blend” to underscore the distinct ways similes select foregrounded elements of one domain. Unpacking the simile “The classroom was buzzing like a beehive”, they clarify the selectivity of the limited-scope blend:

[The simile’s] organizing frame is established as the simile is processed, and is a focused aspect of the source frame. In the target frame, an element of the frame structure is selected (noise is an aspect of the Classroom frame, though it is typically not the most salient one) and enhanced by being mapped onto its counterpart element in the source frame (the noisy Beehive).

(Dancygier & Sweetser 2014: 145)

In addition, this paper also borrows Moder’s (2008, 2010) conception of the narrow-scope and broad-scope blend. To be clear, the paper views the narrow-scope and broad-scope blends as subtypes of limited-scope blends. First, a narrow-scope simile spotlights one specific aspect of the source domain, generally “evoking a vivid or extreme example of a perceptual pattern” (Dancygier & Sweetser 2014: 143). The simile in (3) is an instance of a narrow-scope blend. Broad-scope similes, in contrast, are not as intuitive since they usually require an additional sentence that contextualizes what is being stated. Consider this broad-scope simile:

(4) Mia is like a bottle of wine. They both get finer with age.

Broad-scope similes, like (4), are frequently used as jokes, with the humour stemming from the juxtaposition of absurd pairs. Listeners are induced to select a reasonable link between the two domains, and then to deselect all the other elements that are irrelevant to the comparison. In doing this, humorous broad-scope similes reflect the frame-shifting in on-line jokes, discussed extensively by Coulson (2001). Coulson’s argument is that humorous utterances initially establish
a particular frame which is then shifted to reveal that the initial construal is incorrect. Resolving the sudden frame-shift allows listeners to identify humorous similarities or connections that were originally unnoticed. The following example is from Coulson (2001: 55):

(5) Everyone had so much fun diving from the tree into the swimming pool we decided to put in a little water.

Coulson’s concept of frame-shifting here is useful in homing in on the relationship between humour and the manipulation of frames. Broad-scope similes work along similar lines by expanding on the ostensibly dissimilar connection at the end of the statement. The simile comparing Mia to wine, for instance, ultimately guides listeners to see Mia as a person who gets more attractive as she gets older, just like wine tastes better with age. Other possible frames related to wine that might have been generated in the first sentence (i.e. its source, colour, texture, etc.) are deselected. The adjective *finer* moreover acts as a pun tying the concepts together as it is used typically to characterize wine and is a colloquial qualifier to describe an attractive person.

Based on these recent insights on (verbal) simile, I propose that a multimodal simile, like its verbal counterpart, triggers mappings without relying on the presence of a final blended space where all inputs are integrated. This is the crucial difference that I wish to substantiate throughout the paper. My investigation of multimodal simile will present discursive examples to provide further evidence to support these claims. The analysis will differentiate metaphor from simile by accentuating simile’s selective mapping and its limited scope blending, which foregrounds a particular feeling that is being used as a basis for comparison. In the next section, I will analyze the first type of the “when” meme, whose simile depends on metonymic evocation derived from sociocultural knowledge.

4. **Multimodal broad-scope simile: Frame metonymy and reversibility**

As an example of the first type of “when” meme, consider Figure 3, which features comic-like panels depicting the crowning of Miss Philippines during the 2015 Miss Universe pageant. For those not aware, the crowning of the winner became a trending story on social media because the announcer, Steve Harvey, incorrectly declared the winner to be Miss Colombia who was crowned but then had to have her crown taken away on stage and given to Miss Philippines. The embarrassing

---

2. Many of the memes I look at are from the UNILAD and 9GAG Facebook page. Both pages, for those interested in finding other examples themselves, regularly post “when” memes.
situation became fodder for netizens across the world who ridiculed Harvey for the mix-up. Memes that appeared immediately afterwards tend to focus on either Harvey’s clumsiness, Miss Philippines’ incredulity or Miss Colombia’s premature celebration of her victory, instead of on the entirety of the event. The fragmenting of this event into more comprehensible sub-scenes aligns with memes’ tendency to only transmit small strands of information, making them also evocative visual metonymies. In assuming that audiences have knowledge of what transpired, these memes depict these individuals, often in the form of video screenshots, as metonymic representations of their respective actions and feelings. Screenshots of Harvey, for instance, were taken up by the internet community and used in memes in which he accidentally gives an Oscar to Leonardo DiCaprio, whose failure to win an Oscar (before his first win a few months after this blunder) had been a long-running joke online. These memes poked fun at the fact that Harvey might be the only person who can give DiCaprio a chance at winning an Oscar. Jokes, like this one, are good instantiations of how viral news stories and the individuals in them are stripped down to a fundamental core and function as frame metonymies and as input spaces for more complicated blends.

Figure 3. The “Miss Philippines” meme
Miss Philippines received similar treatment in that the video footage and photographs of her surprising win and the expression of disbelief on her face were compressed into visual metonymies, which were subsequently deployed in various other memes that reduced the whole experience to emblematic representation of those feelings – feelings we can now relate to.

The form of the “when” meme is to use a photo, regardless of its complexity, to convey usually one specific feeling, action or emotion. A defining characteristic of the multimodal broad-scope simile is the initial semblance of incongruity between the verbal element and the visual one. The ideas being evoked by the text do not immediately link with what many of us would associate with a beauty pageant. Of course, knowing the context of the pictures allows audiences to establish mappings between the exam and the pageant, and between you and Miss Philippines. Moreover, the when clause builds the first mental space, setting up the exam frame. The main viewpoint is a student’s, specifically a student who might not be a competent test-taker. The visual input pictorially sequencing Miss Philippines receiving her crown metonymically evokes this specific frame of an unexpected victory.

Though the frame metonymy supplies the meme with a wealth of content, all of this additional information is deselected, leaving only one feeling being spotlighted. The limited-scope nature of similes means that “the cross-mapping is more focused and is not expected to yield rich inferences” which metaphors, not similes, are good at doing (Dancygier & Sweetser 2014: 148). Emphasizing one specific feeling or attribute is one of the rhetorical strengths of the simile, and we see that this is the same strength being augmented in its multimodal form. But are there differences between the verbal and multimodal form? To gauge this, we can convert the multimodal simile into words, and the simile might sound something like this:

(6) The realization that you passed an exam that you thought you failed is like when Miss Philippines realized she won Miss Universe when she thought she lost.

(7) You are like Miss Philippines. You both thought you failed, but you actually didn’t.

The problem with a verbal version of this meme is that it fails to capture the humour of the original, probably because it needs to tediously verbalize all the information in order to forge a link between the two domains. Another failure of this monomodal simile is that it prioritizes the exam frame as the topic and target of the simile and relegates the Miss Universe blunder as merely the source domain. In other words, the multimodal simile can arguably be interpreted in a reversed way:
(8) Miss Philippines winning Miss Universe is like when you passed an exam you thought you failed.

This reversed reading is made possible by both the generalized nature of the *you* and the multimodal format itself. That is, we can conceptualize Miss Philippines viewing her own victory as a moment reminiscent of this fairly relatable experience of passing an exam that one thought one failed. We will see later in Section 7 on source-focused similes that some “when” memes can only be interpreted through the viewpoint of the person or character pictured in the visual input and not through the more general viewpoint of the audience despite the presence of *you*.

Given the aesthetic design of “when” memes, audiences are likely to see the image first before processing the text. Once the text is read, audiences can then discover and map similarities. The multimodal form thus permits audiences to break free from the rigid syntactic structure of a verbal simile.\(^3\) For comparison, though we can reverse the *Mia is like a bottle of wine* simile in a similar fashion, this would require us to completely alter the sentence and move *wine* into the subject position. This, however, completely changes the focus of the statement. Verbal similes do not allow for this reversed reading primarily because their syntactic construction firmly secures the position of the source and target domains; multimodal similes, on the other hand, have no conventional locations for source and target, making such reversed interpretations possible. Additionally, this reversibility of the multimodal simile makes sense given the nature of social media – a platform that facilitates the spreading and modification of visual artifacts. Because social media is a place where individuals share and comment on unusual and humorous photos and videos, we can view the simile of this “when” meme not as a way for us to better understand the feeling of passing a test, but as a way to better understand the blunder of the Miss Universe pageant with the terms and language most familiar to everyday audiences. Users can write a *when* clause to caption a photo to make it into a “when” meme; or conversely, users can upload a photo as a response to another user’s textual expression, thereby enabling the simile’s reversibility. The underlying contrast between the verbal and multimodal broad-scope simile, then, is that the contextualizing component for the multimodal form is a powerfully evocative frame metonymy that manages to supply information in the same way as a verbal elaboration expression.

\(^3\) Further potential for this might lie in the fact that the visual input of “when” memes is usually constituted by real photographs, meaning that they depict a real-life event, even though they might be highly improbable ones. That is, it is easier to see through these perspectives because of the photograph’s inherent realism. The final products of multimodal metaphors, in contrast, would produce more hypothetical situations and unreal entities because the images tend to be designed or digitally manipulated, and therefore cannot be easily interpreted in a reversed way.
5. Multimodal narrow-scope simile: A constellation of viewpoints

Photographs of animals are used frequently in “when” memes to illustrate everyday human experiences and behaviour. While verbal similes frequently compare animals and human beings (e.g. *they ate like pigs; he floats like a butterfly*), multimodal similes can go beyond mapping select features of the animals being highlighted by additionally aligning the audience’s viewpoint directly with that of the animal in the visual. The richness and vividness of the visual provides a constellation of viewpoints which cannot be easily conveyed through verbal similes.

Figure 4. The “crocodile Uber driver” meme

The meme in Figure 4 exemplifies the qualities of our second type of the “when” meme. Different from the “Miss Philippines” meme, this artifact demands less specific cultural context, relying instead on more intuitive embodiment through the physical juxtaposition of a predator and a prey. However, that is not to say no context is required. This meme contains a photograph of a pelican standing on top of a crocodile whose body is partially above the water. The text reads, “When the uber driver looks like a serial killer but you get in anyway because you got places to be”

---

4. Other common visual inputs include famous individuals like celebrities, athletes and characters from movies, television shows and cartoons. There is a less popular, but perhaps more interesting version that uses images of inanimate objects, like cars, that are in positions that make them resemble the human body.
be”. The “when” meme jokes about the physical appearance of Uber drivers, private citizens who use their own cars to offer taxi services to clients via a smartphone application called Uber. The Uber driver reference, of course, can be replaced with a taxi driver and the meme would essentially convey the same figurative meaning. In my analysis, this type of “when” meme is a multimodal simile that resembles the narrow-scope simile, where understanding of the artifact hinges on inherent viewpoints and vivid perceptual patterns.

The *when* clause establishes the roles and values of the subjects and the relationship between the two, whereby “[a] role is a linguistic description describing a category; a value is an individual that can be described by that category” (Croft & Cruse 2004: 34). Take for instance this common role/value substitution we might see in an article of an entertainment magazine: “Harrison Ford is piloting the Millennium Falcon again in the new *Star Wars* movie.” Here, Harrison Ford, the value, is being used to stand in for the role, Han Solo. The role/value distinction helps explicate the figure of simile because in making a comparison between two entities, similes induce the mapping of frames embedded in roles which are consequently transferred to its corresponding value. In the *when* clause of Figure 4 specifically, the roles and values are mapped in the textual element: the passenger and driver are the roles, whereas the audience and the serial killer are the respective values.

With the visual input however, the role and value relationship is complicated. Taking the textual information, audiences have to reconstrue the image of the pelican and the crocodile by appropriately reconciling the incongruities through mapping textual roles and values onto the visual roles and values of the image. The use of the *you* pivots the audiences’ viewpoint, further positioning them to see both as the pelican and to observe the surroundings from the pelican’s perspective. Such an intricate and vivid expression of viewpoint is not easily achievable in verbal similes. That is, verbal similes cannot normally situate hearers in these bizarre spaces since the effect will derive mostly from common knowledge we might have of pelicans and crocodiles. As a result of processing the visual with the verbal, the pelican and crocodile are now construed as the passenger and driver respectively. This figure is thus more characteristic of narrow-scope simile because the contextualizing information is embedded in the extreme visual juxtaposition of predator and prey and not in a highly evocative frame metonymy.

To further clarify the unique viewpoint expression of the narrow-scope “when” meme, Figure 5 takes up, in simpler ways, what we have already shown in the previous example. Here, the text sets up a party frame and suggests that the *you* is an introverted person who stays close to his/her friend during a party filled with strangers. The visual input depicts a baby koala bear anxiously clinging onto the
When you’re at a party full of people you don’t know so you u stay with ur friend the whole time

![Image](image.png)

**Figure 5.** The “koala” meme

ankle of what looks to be a zookeeper. In order for the image to supply additional content to the text, the audience needs to map the roles and values accordingly. Naturally, the shy partygoer is the koala bear and the friend whom he/she stays next to is the leg of the person. With the roles and values mapped, the image then offers more details about the you: Not only is this person unsocial, but this person needs to physically hold onto a friend, who feels safe in this setting.

The conspicuous perceptual pattern of the photo is the act of clinging onto someone, which itself is not a characteristic of any known baby koala frame just like riding a crocodile is not part of a pelican frame. This is precisely what makes a multimodal narrow-scope simile similar to and different from its verbal counterpart. The similarity is that Figure 4 and Figure 5 depend less on contextualization and more on vivid imagery. However, the chief difference is that the multimodal narrow-scope simile generates context not through pre-established cultural frames but from the image itself, thereby producing a myriad of different viewpoints that can be taken up. It is the very essence of the verbal modality that restricts verbal similes from creating such vividness. When a “when” meme like Figure 5 is shared online, users tend to tag their friends as if to say that their friends, or even they themselves, are like, or look like, this koala bear whenever they go out to parties. This communicative and rhetorical function of the “when” meme, which is to make a humorous comparison between an everyday activity and a strange photo, is perhaps what has facilitated its spread.
6. Mimetic speech and performance

To recap, my first two categories of the “when” meme were distinguished by the visual input’s level of accessibility. The “Miss Philippines” meme was significantly harder to make sense of than the “Crocodile” or the “Koala” ones because the broad-scope simile refers to a highly specific event. Yet, regardless of accessibility, all of these memes hinged on the same fundamental mechanisms: the incongruity of the verbal and visual components is reconciled by the selective mappings of frames, viewpoints and roles. Continuing with this spectrum of accessibility, my third type of “when” meme is arguably the most accessible because the visual input relies on the audience’s intuitive grasp of facial expressions and physical gestures. These are expressions that are more natural and innate, and less cultural or arbitrary. Figure 6 again features an animal performing human activities. Here, an iguana is standing on top of a laptop keyboard with its eyes glaring back at the camera. The meme’s text tries to evoke from the iguana’s facial expression the same physical appearance of someone giving another person a suspicious glare. The text instills these human attributes in the iguana, which happens to look like a person slightly rotating his/her head back to stare someone down. Moving beyond just facial expressions, Figure 7 features a photograph of an employee of an office supply store posing awkwardly at work. The text compares the man’s bodily posture to a similar pose one might instinctively make when coming into contact with hot water.

When you typin in your password and you realize your friend is behind you

Figure 6. The “iguana” meme
I submit that this type of “when” meme parallels the reported speech and thought marker *be like* (Romaine & Lange 1991). In fact, many of the *be like* memes, which I have not yet mentioned, use the *when* clause as well. We could reconfigure Figure 6 and 7 into a “be like” meme (in which the verb *be* is left typically unconjugated) to show how its inclusion does not significantly alter the artifacts’ meaning:

(9) When you [sic] typing in your password and you realize your friend is behind you, you be like

(10) When your shower is too hot and it hits your back, you be like

The *be like* construction acts in ways that match the colloquial *like* for reporting speech (e.g. *She told me to do that, and I was like no!*). In our multimodal form then, the *like* builds a mental space to store, in this instance, the reported or mimetic facial expression or performance. *Like* has been the topic of much linguistic research (Siegel 2002; Tagliamonte & D’Arcy 2005; Buchstaller 2014). Though relatively few have extensively commented on the relationship between simile and

---

5. There is another common variant that explicitly signals the mimetic performance: “The face you make when you [sic] typing in a password and you realize your friend is behind you.”
the *like* discourse marker (see Vandelanotte & Davidse 2009; Vandelanotte 2012),
there are conspicuous parallels, however. Romaine & Lange (1991: 266) remark
explicitly that *like* as a word that also means ‘similar to’ has paved the way for *like*
to introduce a “mimetic performance”.

I posit that a person’s cognitive processing of a mimetic performance of report-
ed speech is motivated by the same mapping process we perform when interpreting
figurative language. Echoing the notion of unattributed speech (Pascual 2014), I
suggest that when a person mimics the facial expression or speech of an individual,
audiences selectively retrieve foregrounded, and sometimes hyperbolic, attributes
and map them onto the generated mental space of the person being described.
Even though this subject is not present and even though the person doing the
mimicking does not resemble that person, audiences are able to reconcile the dis-
parities as trivial differences. It would make sense to say that the “when” meme,
as well as the “that moment when” meme before it, began initially as some sort of
signifier ushering in a mimetic performance (like Figures 2, 6 and 7). Only later
did the visual inputs evolve into more complex entities (like Figure 3).

What I intend to reaffirm here is an intimate connection between simile and
*like* as a marker of reported speech and thought. Each “when” meme presented
in this paper has demonstrated, to varying degrees, the mechanics of simile with
the more inaccessible ones (i.e. “Miss Philippines”) exhibiting characteristics less
associated with reported speech and thought, and *vice versa*. It is, nevertheless,
still possible to see the visual input of the “Miss Philippines” meme as a mimetic
performance, albeit as an exaggerated and dramatic one. If this is the case, we can
outline an accessibility scale with visual inputs that are mimetic performances
placed at the accessible end of the scale, and visual inputs composed of complex
visual metonymies placed at the inaccessible end. Multimodal similes, and per-
haps all similes in general, are therefore inextricably tied to embodiment and to
some form of mimetic performance. The purpose of proposing an accessibility
spectrum is to avoid a binary distinction of literal and figurative language and
also to acknowledge the shared functions of *like* as a reported speech and thought
marker and as a conventional marker of the simile construction. Generally speak-
ing, the literal and figurative division of language assumes that literal language
does not overlap with figurative language. However, as we can see in this section,
multimodal mimetic similes derive their figurativeness from the photograph’s
peripheral content (e.g. the stationery store setting of Figure 7). However, there
exists some core structure (e.g. the man’s arched back pose) within the visual input

6. Other similar memes have phrases like “the face you make when,” “when you suddenly
realize,” and “that moment when you realize.”
that directly resembles what the target attempts to compare the source to (e.g. the pose one makes when hot water hits his back).

7. The source-focused “when” meme

The last “when” meme that I will touch upon briefly here is the source-focused “when” meme. I call this type “source-focused” because the source domain (the visual input) directly dictates the content and expression of the target domain. Our first example, Figure 8, is taken from a Facebook page called Classical Art Memes; classical art “when” memes marshal much of their humour from their use of modern commentary, superimposed as a textual voice that reframes the conceptualization of medieval and renaissance artwork. The meme contains Johan Zoffany’s Self-portrait as David with the Head of Goliath, and though the text is highly bizarre and makes little sense without some idea of what cover photos of

![Image](image_url)

**Figure 8.** The “David and Goliath” meme
Vogue, a lifestyle and fashion magazine, look like, the interaction between David and Goliath in the text does not describe the visual input in radically different terms. Unlike many of the other memes we have looked at where the text inserts new relationships between roles and values, the text maintains core aspects of a David and Goliath frame.

In short, the source-focused “when” memes do not establish a stark contrast between itself and the image, nor does the text modify its viewpoint to accommodate the you in ways that would make it a viewpoint we can relate to personally. The same goes for Figure 9 and its photograph of a dog sitting in a chair using a computer. Audiences are not meant to see themselves through David’s or the dog’s viewpoint, even though we have seen previously that “when” memes can induce us to take up the viewpoints of animals in spite of jarring incompatibilities. The idiosyncratic quality of this meme is that the text tends to illustrate the image in detail, resulting in a multimodal artifact that appears to poke fun at the “when” meme’s form. Memes that begin to reference themselves to form meta-jokes are not an uncommon occurrence. Dancygier & Vandelanotte (2016) observed this phenomenon with the “said no one ever” meme, whose spread led to social media users repurposing that very construction to mock the popularity of the meme. As one reviewer commented, the source-focused meme perverts the typical “when” meme form by disabling our ability to connect and sympathize with the simile.

![Image of a dog using a computer](image)

**Figure 9.** The “hacker dog” meme
Figures 8 and 9 resemble the “crocodile Uber driver” meme insofar as all three attempt to humorously verbalize the unusual and distinct qualities of the picture. But unlike the “crocodile Uber driver” meme, these two make a comparison without setting up logical roles for the you. Though hacking is an activity performed exclusively by humans, human beings do not hack to find out if they are “a good boy”, a clichéd utterance that pet owners might say to their dogs when they perform a trick. Figure 9 makes fun of this frame in which the dog goes above and beyond to impress its owner. For Figure 8 too, audiences cannot possibly relate to the experience of killing Goliath, and are highly unlikely to have ever been called to model for Vogue. Instead of making witty comparisons, source-focused “when” memes attempt to describe the content of the picture and retain as much of the picture’s absurdity as possible. Translating Figure 8 into a verbal simile might help make sense of what is being articulated:

(11) This painting of David looks like a Vogue photo shoot.

The fact that the meme allows users to phrase this more conventional simile in such an unconventional way is a testament to rapid linguistic change that occurs on social media, wherein users collectively take up, use and manipulate existing constructions to engender new and mutually intelligible expressions. In other words, one key trait of the multimodal simile worth noting is how its verbal-visual configuration has given rise to a novel version of the simile – one that is able to replicate the figurative effects of a verbal simile without having to use its generic like or as form.

8. Conclusion

This paper has several objectives. In a broader sense, the paper aimed to underscore the wealth of artifacts on social media that linguists of all stripes can mine and analyze to discover the creation and evolution of novel linguistic expressions.

---

7. One can even suggest that the image here is anthropomorphic (although the image is a photograph and not a cartoon rendering) because the dog is dressed and posed as a human being.

8. Because it is much easier to make a witty description of a photograph than to find a photograph to accurately describe a textual description, the multimodal format of the “when” meme, in many ways, controls how these similes are created. Both source and target domains are equally easy to construct for a verbal simile. The hurdles of making multimodal similes, in turn, reinforce the fact that the simile is reversible. The text is superimposed to make sense of the photograph, rather than the other way around. The text can thus be viewed as the source domain despite appearing first in the artifact.
More specifically, the paper, in positioning itself alongside cognitive linguists interested in multimodality, views the “when” meme as a type of multimodal simile. From this cognitive semantic angle, the paper has attempted to contribute to the growing study of multimodal tropes by demarcating and elucidating the defining properties of a multimodal simile, a figure that has been vastly understudied verbally – let alone multimodally.

This investigation of the multimodal simile has certainly been fruitful in that it has illuminated the relationship between simile and a plethora of other well-established tropes and cognitive operations, including frame metonymy, conceptual blending, role and value mapping, mimetic performances and embodiment. Each of these operations connects to one or more of the four unique “when” memes that I have classified: (1) The multimodal broad-scope simile features a visual input that is jarringly incongruous with the text. Processing the simile requires the unpacking of its frame metonymies, which evoke compressed ideas of particular events and people which the simile tries to compare the target domain to. (2) The multimodal narrow-scope simile exhibits more accessibility; even though the image and text are initially incompatible, both the image and text highlight a similar perceptual pattern that facilitates mapping across modalities. (3) The multimodal mimetic simile contains an image that visually depicts the actions being expressed or implied in the meme’s text. My analysis of this mimetic simile posits an underlying connection between the *like* used in simile and the *like* used in reported speech. (4) Lastly, the source-focused simile presents a comparison that constructionally resembles the other “when” memes. However, the simile’s textual element retains the absurdity of the image, rather than likening it to something intuitively comparable. The simile thus subverts conventionality by restricting audiences from directly relating to what the simile tries to convey.

As always, there is a lot more work to be done. Throughout my research, I stumbled across countless versions of the “when” meme; and by being subscribed to meme pages on social media, I encountered dozens of them on a daily basis. One type that is difficult to analyze in a paper is the “when” memes with visual inputs that are video clips taken from television programs and movies. Its aesthetic appearance is identical to the ones I have shown. The text stays static, and video footage takes the place of the image. These video artifacts correlate with the four types of “when” memes I have discussed, with some that would clearly focus on facial expressions, while others feature decontextualized clips of scenes from movies. Right now, at the time of writing, the “when” meme is still very much alive, and continues to evolve and undergo linguistic change right before our eyes.

It is not uncommon, these days, for people to caption their online photos, or photos uploaded by other people using the “when” construction. In one instance, a picture of a cat when he was still a kitten elicited the response “When you knew
you would become an evil genius”, which plays with the common conception that cats are manipulative pets. American socialite Kim Kardashian made headlines recently for posting a nude selfie of herself on Instagram with a caption that reads, “When you are like I have nothing to wear LOL.” The when clause’s function as a construction for captioning is certainly worth elaborating on in future studies that would interest cognitive linguists and communication, rhetorical and new media scholars alike. During this research, I was exposed to so many that it became increasingly difficult to categorize them. Yet, the reality of working with online artifacts is their ephemeral nature. The pace at which these artifacts change is what makes them fascinating objects of study. Future research can also probe further into the connection between reported speech and simile in order to recognize the nuanced gradients of figurative language and to better distinguish what separates the literal and figurative. The ultimate goal is that we can finally say with some certainty what makes simile different from metaphor and to draw out what these essential cognitive processes are. Perhaps, with these memes, which have revitalized the ancient trope of simile, we can finally find the answers to these questions.

References


Dancygier, Barbara & Lieven Vandelanotte. Submitted. *Internet memes as inter-modal constructions*.


Moder, Carol. 2010. Two puzzle pieces: Fitting discourse context and constructions into cognitive metaphor theory. *English Text Construction* 3 (2): 294–320. doi:10.1075/etc.3.2.09mod


Author’s address

Adrian Lou
Department of English
University of British Columbia
397 – 1873 East Mall
Vancouver, BC, V6T 1Z1
Canada
adrian.lou@ubc.ca