

# EMOJIS AS PICTURES

EMAR MAIER

*University of Groningen*

I argue that emojis are essentially little pictures, rather than words, gestures, expressives, or diagrams. 🏠 means that the world looks like that, from some viewpoint. I flesh out a pictorial semantics in terms of geometric projection with abstraction and stylization. Since such a semantics delivers only very minimal contents I add an account of pragmatic enrichment, driven by coherence and non-literal interpretation. The apparent semantic distinction between emojis depicting entities (like 🚗) and those depicting facial expressions (like 😊) I analyze as a difference between truth-conditional and use-conditional pictorial content: 🚗 depicts what the world of evaluation looks like, while 😊 depicts what the utterance context looks like.

## 1. Introducing the Pictorial Account of Emojis

Wittgenstein is sometimes (half-)jokingly credited with the invention of emojis, on the basis of the following excerpt from his *Lectures on Aesthetics*:<sup>1</sup>

If I were a good draughtsman, I could convey an innumerable number of expressions by four strokes.



Such words as ‘pompous’ and ‘stately’ could be expressed by faces. Doing this, our descriptions would be much more flexible and various than they are expressed by adjectives. [...] I could instead use gesture or [...] dancing. In fact, if we want to be exact, we do use gesture or facial expression. (Wittgenstein 1966)

1. E.g., <https://qz.com/1261293/ludwig-wittgenstein-was-the-great-philosopher-of-the-20th-century-he-also-invented-the-emoji/>.

**Contact:** Emar Maier <e.maier@rug.nl>

The suggestion is that drawing stylized faces would be a useful addition to written language, as it would provide an efficient way to express certain meanings,<sup>2</sup> especially those kinds of meanings that are usually conveyed by gesture or facial expression, or else, less efficiently, or less precisely, by evaluative adjectives. In the context of other famous remarks like “the human body is the best picture of the human soul” (Wittgenstein 1958: 178), this could be taken as suggesting the view that such face drawings, like gestures and facial expressions, are *expressives*, that is, meaningful signs that do not directly contribute to truth conditions, but rather express something non-propositional, like the speaker’s emotional state.

There is no doubt that some modern emojis are used roughly as Wittgenstein envisages for his face sketches. Not surprisingly, some of the points Wittgenstein makes are echoed in recent linguistic analyses of emojis. In particular, we see suggestions that emojis are like gestures (Gawne & McCulloch 2019; Pasternak & Tieu 2022; Pierini 2021), that face emojis in particular are expressives (Grosz, Kaiser, & Pierini 2021; Grosz, Greenberg, De Leon, & Kaiser 2023). More generally, emoji are often considered to function somewhat like words (Barach, Feldman, & Sheridan 2021; King 2018; Scheffler, Brandt, de la Fuente, & Nenchev 2022; Tang, Chen, Zhao, & Zhao 2020).

What many of the modern linguistic approaches share is that they treat emojis, like regular words, as symbols—a mode of signification loosely characterized as conventional, non-natural, arbitrary, and/or learned. The obvious alternative to this symbolic view is one that treats emojis, like pictures, as icons—a mode of signification loosely characterized as based on resemblance between form and content (Peirce 1868). Such a view is apparently taken for granted by some semioticians (Cohn, Engelen, & Schilperoord 2019; Danesi 2016), but it is never explicitly argued for or made very precise. In this paper I propose, formalize, and defend such an iconic semantics for emojis. More specifically, I argue that emojis are simply little pictures. That is, like photographs and drawings, they are used to depict ‘what the world looks like’.<sup>3</sup>

A pictorial account of emojis promises several advantages over rival symbolic accounts. First, it sees the use of emojis as continuous with other, more obvious picture–text integrations, like stickers, gifs, and memes in modern internet communication, but also like illustrated books, instruction manuals, and comics—and of course Wittgenstein’s face drawings. This allows me for instance to borrow a fully general pragmatic model of picture–text composition (originally proposed for analyzing comics) and apply it to the use of emoji in Section 3.

---

2. Close reading suggests a ‘rebus’ account, where the face drawings would express words, rather than their meanings. We’ll disregard this arguably uncharitable reading.

3. In Section 2.3 I discuss, and reject, an alternative non-pictorial iconic analysis where emojis are diagrams.



Second, the pictorial account can explain creative, non-canonical emoji usage, like a use of the violin emoji to illustrate a cello performance, (1-a), or the use of the ‘persevering face’ emoji to express a host of seemingly unrelated emotions (frustration, sadness) and activities involving closing the eyes (praying, pretending to be asleep), united only by the fact that someone looks like that.


- (1) a. thanks to @anonymous for chatting to me about reaching a global audience online during lockdown with his stunning cello performances 🎻
- b. Oh man... that too? They stole that too? 😞
- c. Behind every quiet person there is sad untold story 😞
- d. I need a guy that’s ready for a serious relationship 🙏😞
- e. I always close my eyes and pretend to be sleeping 😞

To deal with such a range of conceptually distinct uses, a symbolic account would have to assume that the lexical item 🎻 or 😞 is multiply ambiguous, while the pictorial account readily predicts these creative usages from the generic pictorial meaning, that is, that ‘it looks roughly like this’. We’ll explore creative uses in Section 4.

Thirdly, and somewhat more speculatively, the pictorial account suggests a natural explanation of the rise of gender and skin-tone modifiers: on certain specific occasions, 🏃 or 👍 may simply more closely resemble what they depict (the runner I’m describing, or my use of the gesture) than the default 🏃, or 👍, respectively. A symbolic account that analyzes 👍 as a lexical item expressing the speaker’s approval would presumably assign the exact same meaning to the skin-tone variants 👍 or 👍 and thus have a harder time explaining their different distributions and felicity conditions. We’ll revisit this argument in Section 5.

Having made my sales pitch, I should add that there are also some limitations and caveats to the scope of my proposal. Emojis are not a wholly homogenous class, and my pictorial account will not treat all 3,521 emojis in the current Unicode standard uniformly. First, I’m assuming, with Grosz et al. (2021), a semantic distinction between emojis that depict facial expressions and hand gestures and those that depict other entities and eventualities. I propose to model that semantic distinction within my overall pictorial framework as follows: while the entity/event emojis depict what the world of evaluation looks like, the face/hand emojis tend to depict what the utterance context looks like. A face emoji thus essentially depicts what the actual speaker looks like while producing the utterance. The Wittgensteinian intuition that face emojis are expressive is then explained by the further assumption that the human facial expressions (or hand gestures) depicted are themselves meaningful signs, and that the pictorial meaning layer naturally composes with the expressive gesture, in a way to be made precise in Section 5.

I'm also leaving open the possibility that there is some subclass of emojis that are best analyzed as symbols. For instance, , the fifth most common emoji on Twitter according to emojitracker.com, is conventionally used to denote recycling or, more commonly, retweeting, but it's not obviously a *picture* of either of these activities—it doesn't resemble them. Similarly, the Belgian flag emoji doesn't resemble the country it stands for, but refers to it by an arbitrary convention, much like the English word 'Belgium', or an actual Belgian flag. Some emojis in the Emojipedia<sup>4</sup> category of 'Symbols' fall in the grey area between picture and symbol —is  just a conventional symbol of love, or is it perhaps in some sense a stylized (see Section 2.2) picture of a human heart, which by (fossilized) metonymic extension (see Section 4.3) is associated with love and positive emotions? For the purposes of this paper I'm happy to concede that there may be a subclass of symbolic emojis, with fuzzy borders. Still, there might be an alternative analysis that treats (some of) these symbolic emojis as pictures as well.<sup>5</sup> Following the two-stage meaning composition account I propose for face emojis in Section 5, briefly hinted at above, we might say that flag emojis are quite literally pictures of flags, which in turn are genuine symbols of countries. This would leave us with the task of explaining the apparent transparency of the pictorial layer here. While I do offer such an explanation for the apparent transparency of face emojis (Section 5), I will not pursue this route for flag and other symbol emojis here.


In addition to pictorial and symbolic emojis, and some in the grey area in between, there are also picture–symbol hybrids. For instance,  depicts a sleepy face with a giant snot bubble coming from the left nostril. The snot bubble here is a convention borrowed from Japanese manga and anime that symbolizes that a character is sleeping (Cohn 2013). We can analyze such mixtures by syntactically separating the symbolic elements from the pictorial elements, for instance as described for speech balloons and other picture–symbol hybrids in comics by Maier (2019). I will not discuss this matter further here.

Finally, for reasons of space I'll restrict attention to the use of emojis as separate discourse units, that is, typically inserted after a sentence, or as a stand-alone discourse move:

(2) Great idea  count me in  I'm on my way 

That is, I'm disregarding 'pro-speech emojis' (Pierini 2021)—emojis that are syntactically integrated into a sentence and 'replace' a specific word or concept, like

4. <https://emojipedia.org>

5. Many thanks to an anonymous referee for floating the idea that flag emojis—especially Apple's or Samsung's 3D looking ones —are pictures of flags.

'love' and 'present' in (3-a), 'happy' in (3-b), and sometimes even a specific (English) word sound or shape, in rebus-like fashion, like in (3-c).<sup>6</sup>

- (3) a. keep doing what you need to do, ❤️ u bro if I was in Detroit I'd give you a 📺.  
 b. Our project eventually succeeded, and I felt very 😊 (Tang et al. 2020)  
 c. In the 🍷 of her hand. (Scheffler et al. 2022)

With the above restrictions and caveats in place, we are left with the claim that a significant portion of emoji uses, the exact boundaries of which remain vague, but including many uses of emojis for animals, plants, objects, activities, hand gestures, people, facial expressions, are wholly or primarily pictorial. In this paper I will explicate in some detail what a pictorial semantics (and pragmatics) for emojis might look like.

The paper is structured as follows. In Section 2 I propose a formal semantic account of pictorial content in terms of geometric projection. In Section 3 I explain how the rather minimal projective semantics is enriched in the context of a discourse, building on insights from the study of linguistic discourse structure and coherence relations. In Section 3 I address what I call the pictorial overdetermination challenge: an emoji like 🚗 depicts a certain type of two-door red car, but can be used to denote cars of any color, make and model. The solution I propose invokes a pragmatic process of figurative interpretation, extending the basic projective contents via metaphoric and metonymic interpretation. In Section 5, finally, I turn to the second fundamental challenge facing a pictorial account: how to explain the apparent expressivity of face and hand emojis? I propose an extension of the Kaplanian account of use-conditional meaning to pictures and then show how the use-conditional pictorial content of an emoji naturally combines with the use-conditional content of the depicted facial expression or gesture to yield the observed expressive use of face and hand emojis.

## 2. A Pictorial Semantics for Emojis

### 2.1. Geometric Projection

Pictures are representations. They represent the world as being a certain way. Hence, just like utterances can be true or false with respect to a given world, we

---

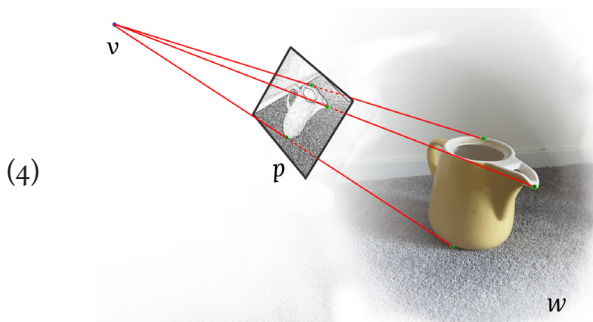
6. Most of these examples are in principle compatible with a pictorial account, though (3-c) is problematic for both symbolic and pictorial accounts.

could say that a picture is true or false with respect to a certain world—more colloquially, that a picture is an accurate or inaccurate picture of (a part of) that world. If we can capture a workable notion of pictorial truth we can define the semantic content of a picture as the set of worlds that the picture is true of—in the same way that we also define the content of an utterance in possible worlds semantics—in order to get a proper investigation of the semantics and pragmatics of pictures off the ground.


Intuitively, pictorial truth, unlike linguistic truth, is a matter of resemblance: a given picture is true of a world iff it resembles part of that world. On reflection, resemblance has turned out to be too vague, and arguably neither sufficient nor necessary for pictorial truth (Goodman 1976; Greenberg 2013). The geometrical notion of a projection function has been used with some success as a replacement for resemblance in pictorial semantics (Abusch 2020; Greenberg 2021).

More technically, a geometric projection is a recipe for turning a 3D scene into a 2D pictorial representation of that scene. It's a function,  $\Pi$ , mapping a possible world  $w$  and a viewpoint  $v$  (formally, a vector located at a certain spatiotemporal location, intuitively representing the gaze direction of some viewer/camera located somewhere in  $w$ ) onto a picture  $p$ :  $\Pi(w, v) = p$ .



There are many different such recipes that qualify as geometric projection functions. One well-known example  $\Pi$  takes the world and the viewpoint, and (i) puts a white picture plane perpendicular to the viewpoint direction vector, (ii) draws all 'projection lines' connecting some part of an edge of an object in the world towards the viewpoint origin, and (iii) marks in black wherever the projection line intersects the picture plane. This procedure will generate a linear perspective, black and white drawing.



More complicated projection functions might include rules for representing colors, distinguishing edges and surfaces in the world, or some additional distortion, abstraction, and stylization transformations to create depictions that deviate more or less from 'photorealistic' projection in different ways.

(5)  $\Pi(w, v) =$  

When we know how to turn some part of the world into a picture, using a  $\Pi$  and a  $v$ , we can define when a picture is true:

(6)  is true of  $w$  viewed from  $v$  iff  $\Pi(w, v) =$  

Here we're assuming the projection function  $\Pi$  to be fixed, that is, provided by the context, just as in the linguistic domain we assume the language to be given pre-semantically, that is, before computing the truth value of an utterance. In other words, we can think of the projection function as the pictorial analogue of a language (Giardino & Greenberg 2015; Greenberg 2013).

From the pictorial truth definition in (6) we can define various candidate notions or levels of pictorial content. A natural analogue of classic propositional content results from existential closure over the viewpoint: given a pictorial language  $\Pi$ , a picture expresses the proposition that there is a viewpoint from where the world projects onto that picture:

(7)  $\left[ \left[ \text{img alt="A black and white photograph of a white pitcher on a table, viewed from a high angle." data-bbox="188 533 354 591"} \right] \right]^\Pi = \left\{ w \mid \exists v. \Pi(w, v) = \text{img alt="A black and white photograph of a white pitcher on a table, viewed from a high angle." data-bbox="584 536 716 589"} \right\}$


Alternatively, for different purposes we may need other notions of content, for example, analogues of centered/diagonal propositions (sets of world–viewpoint pairs, Rooth and Abusch 2017) or horizontal contents (sets of worlds, i.e., assuming a fixed, contextually given viewpoint). I'll eventually introduce a dynamic semantic notion of pictorial content, in terms of information states.<sup>7</sup> To avoid potential difficulties delimiting and quantifying over the space of possible pictorial (or linguistic) languages, we will assume that a  $\Pi$  is pre-semantically given. How an interpreter arrives at this  $\Pi$  is then a matter of contextual pragmatic inference that we will only talk about informally.

7. It may be interesting to explore 'metaprojective' notions of pictorial content, such as 'metaprojective diagonals', i.e., sets of world-viewpoint-projection-triples, in order to semantically model reasoning and uncertainty about the exact projection parameter settings that gave rise to a given picture. The analogue move in the linguistic domain would lead to a level of content useful for describing interpretation by an interpreter who does not know what language she's interpreting.

Finally, it is worth noting that the purely projective pictorial content defined in (7) is a rather minimal notion of semantic content, what Kulvicki (2006) calls ‘barebones content’ — to be further enriched to what he calls ‘fleshed out content’ (see Greenberg 2021 a related view involving different levels of pictorial semantic content). In this paper I side with Abusch (2020) and stick with the barebones semantics in (7). I relegate the more fleshed out content derivation to the levels of discourse processing (see Section 3) and pragmatics (see Section 4).

## 2.2. Stylization

If emojis are pictures, they are not very ‘realistic’, but rather ‘abstract’ or ‘stylized’. In the geometric projection framework the differences between, say, a simple line drawing and a full color photograph can be thought of as corresponding to different parameter settings inside the projection function. A line drawing projection ignores colors, shadows, and other properties of surfaces, and instead focuses only on (clear, relatively sharp) edges of objects. Qualitatively very different scenes (solid blue cube on wooden table lighted from above, transparent glass cube on metal surface lighted from the left, etc.) could thus give rise to the same abstract line drawing.

Now, apart from ignoring surface textures, opacity, colors, and shadows — let’s call this ‘abstraction’ — a typical line drawing also simplifies the geometry of the edges that the basic linear projection algorithm would give us. Let’s call this ‘stylization’. By stylization, slightly crooked edges and small imperfections might be represented by perfectly straight lines on the picture plane. We could also stylize our depiction by approximating any shape projected on the plane with the closest simple polygon (with less than 37 sides, say).<sup>8</sup> We’ll consider such approximative geometric transformations part of the projection function (Abusch 2012; Greenberg 2021). With a properly abstract and stylized projection function, a simple wire cube drawing like  would be true of not just different geometric worlds where we’re looking at an actual floating Platonic cube (of arbitrary color and size), but also of worlds like ours where we’re looking at a shape that is roughly cube-like, like a sugar cube, a Rubik’s cube, or a dented cardboard box.

If emojis are pictures — as I maintain — they are clearly more like line drawings than like photos, with the relevant projection function involving multiple types of abstraction and approximative stylization transformations on top of a

---

8. We’ll have to define ‘shape’ and ‘closest’ more precisely to make this process deterministic, just as we need to define notions like ‘shadow’, ‘edge’, ‘background’. For some pointers about the literature on these topics, see for instance the recent overview of pictorial semantics by Abusch (2020).



basic linear projection algorithm. Take a common object emoji, like the ‘wrapped present’ emoji. Here are a few instantiations of this emoji in different emoji sets.



Apple appears to be using a more photorealistic type of projection and HTC a more stylized one. More specifically, in our projective semantics terminology, we would say that Apple’s projection function,  $\Pi_{Apple}$ , seems to involve a standard linear perspective;<sup>9</sup> uses a range of different colors to mimic a smooth, somewhat shiny lightbrown or gold surface; and marks shadows and shiny edge highlights as if light is falling on the object from top left.


The OpenMoji projection seems to involve more abstraction and stylization. The box and the bow for instance are entirely symmetrical, suggesting that  $\Pi_{OpenMoji}$  ignores the precise shape and location of the bow in the basic projection image in favor of an approximation. There are only a few colors, again suggesting approximation, and all edges are marked uniformly in thick black lines. On the other hand, the dropdown shadow from the lid is still preserved. The type of perspective in OpenMoji’s projection remains unclear because a full frontal viewpoint seems to have been chosen.

HTC, finally, uses the same canonical viewpoint as OpenMoji and builds similar color, texture, and symmetry abstractions into its projection function, though with a slightly different edge processing, and now even ignoring all shadows.

Note that these informal (abductive) inferences about the nature of the three  $\Pi$ ’s, drawn on the basis of just one image each, are all defeasible: Apple might in principle have intended to depict a crooked, multicolored box through a parallel method of projection, without shading; and HTC’s projection function might be sensitive to shadows and shading but we don’t notice because the scene had a light source near the viewpoint, etc. In fact, all three pictures might in principle be high resolution photographs of more or less abstract drawings of objects (Greenberg 2013; Kulvicki 2013).

As discussed also at the end of Section 2.1 we’ll ignore this general projection uncertainty and assume that context, common sense and experience with pictures of common objects pragmatically (or in any case, pre-semantically) narrow

---

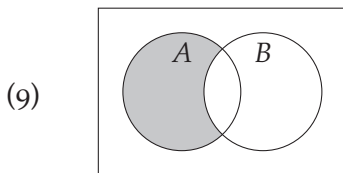
9. A referee suggests that the projection appears parallel rather than linear, just like the wire-cube above . Zooming in on the emoji picture, it seems to me that the edges in the back are slightly shorter than the parallel ones in front, which are moreover not quite parallel, suggesting a vanishing point. But ultimately, it doesn’t matter, as parallel projections are just as ‘pictorial’ and ‘iconic’ as linear projection (Giardino & Greenberg 2015).

down the space of possible parameter settings to something like the assumptions laid out above.

### 2.3. Icon, Picture, or Diagram

At the HTC level of stylization and abstraction we might be tempted to consider an alternative account, as suggested by an anonymous referee, viz., that emojis are diagrams. That is, 📺 denotes a wrapped gift box in roughly the abstract but still iconic way that overlapping, partly greyed out circles in a Venn diagram might denote that all men are mortal. So how exactly do diagrams differ from pictures, icons, and symbols?

Giardino and Greenberg (2015) define iconicity as representation by virtue of “a kind of ‘direct’ or ‘natural’ correspondence between the spatial structure of the sign and the internal structure of the thing it represents.” Pictures fall under this definition, with the relevant correspondence provided by (the inverse of) the projection function. Importantly, pictorial correspondence is essentially viewpoint-dependent: a given linear perspective photograph may be a true depiction of a world from some specific viewpoint, but false from another viewpoint. According to Giardino and Greenberg, such viewpoint-dependence is what sets pictures apart from other icons, most saliently diagrams: a Venn diagram like (9) may convey, in virtue of a natural correspondence between circle overlap and set intersection, the proposition that all humans are mortal, regardless of any specific viewpoint or perspective.



Back to emojis. In the description of the projective content of the HTC wrapped present emoji in 2.2 I’ve assumed that it’s a stylized depiction from a canonical, frontal viewpoint. One might argue that if we incorporate such a fixed full frontal viewpoint into  $\Pi_{HTC}$  we’d technically end up with a viewpoint-independent projection, which by Giardino and Greenberg’s (2015) definition might already technically put us outside the pictorial domain. However, viewpoint-independence doesn’t seem to suffice to make an icon a diagram, for then interactive VR games or 3D marble sculptures would be diagrams as well.<sup>10</sup>

10. As would various non-Western art styles that rely on fixed viewpoints. See Hagen (1986) for discussion of canonical viewpoints in ancient Egyptian.

Following Hagen (1986) I will extend the label ‘pictorial’ to include forms of projection that stipulate a canonical viewpoint. This leaves it an open question whether 🎁, while clearly iconic, is best thought of as a picture or a diagram.

Perhaps the diagrammarians’ case is strongest for face emojis. In our terminology, 😊 and 😞 are iconic in the sense that they denote facial expressions and/or emotions—we return to this matter in Section 5—by virtue of a ‘natural’ correspondence between shapes in the sign (specifically the shapes of mouth and eyes) and properties of the speaker’s face and/or emotional state. Although this paper is primarily concerned with defending this iconic account against symbolic accounts, let me briefly explain my reasons for going one step further and pinning down the correspondence in question as pictorial, as cashed out in terms of the fairly well-understood geometric notion of a projection function.

First, to the extent that it makes intuitive sense to consider Apple’s colorful and detailed 🎁 a picture, it makes sense to try and extend that approach, if possible, to more abstract emojis like 😊 and other emoji sets like HTC’s, and perhaps even more abstract representations like emoticons :). On the account presented in 2.2 above, 🎁 and 🎁 simply exemplify different pictorial dialects, characterized in terms of projection functions with different types of stylization and abstraction built in. Going in the other direction on the abstraction scale, we already noted in Section 1 that the pictorial account also provides an intuitive link between emojis and, say, Whatsapp stickers or gifs that often include drawings, photos, or videos that are uncontroversially projective and viewpoint-dependent in nature.

In Section 1, while comparing our pictorial account with the symbolic alternative, we also mentioned two other potential advantages of the pictorial view: (i) it correctly predicts the flexibility and creativity of emojis (e.g., 🎻 denoting a cello performance, or 😞 denoting a wide variety of conceptually unrelated emotions and activities associated with a face that looks like that); and (ii) it can easily make sense of the rise of skin-tone and gender modifiers. It is not obvious whether a diagrammatic account would be similarly well positioned to explain these two phenomena—the explanations sketched earlier at least require a much more vision-like type of form–meaning correspondence than we find in, say, Venn diagrams (for which, indeed, creative interpretations and skin-tone/gender modifiers seem rather unlikely).

It all comes down to how exactly the diagrammarians spells out the ‘natural’ correspondence between emoji and denotation in a way that is not projective or pictorial but instead distinctly diagram-like. The deeper problem is that it’s not clear what counts as ‘diagram-like’. Beyond detailed accounts of some specific logical and mathematical diagram systems (Shin, Lemon, & Mumma 2018), there is, as far as I’m aware, no general, formally precise, positive characterization of diagrammatic representation. Hence, the view that (some) emojis are diagrams is less informative than describing geometric projection functions with styliza-

tion, abstraction, and canonical viewpoints. In this paper I'll henceforth restrict attention to the pictorial view.

### 3. Emojis in Discourse

The pictorial semantics I have proposed for emojis is incredibly minimal:

$$(10) \quad \llbracket \text{📺} \rrbracket^{\Pi_{Apple}} = \left\{ w \mid \exists v . \Pi_{Apple}(w, v) = \text{📺} \right\}$$

$\approx$  the proposition that there's a viewpoint somewhere in space and time from where the world looks like this: 📺.

As we saw in 2.2, already some defeasible presemantic reasoning about the underlying  $\Pi_{Apple}$  is required to get even this much. A lot more pragmatic reasoning is needed to turn this basic pictorial content into something worth adding to an actual tweet or text. I follow Grosz et al. (2021) and Kaiser and Grosz (2021) in appealing to coherence and discourse structure as a crucial factor in the pragmatics of emojis, but my reliance on pragmatic enrichment will be somewhat more radical, in part due to my much more minimal, pictorial semantics.

#### 3.1. Coherence in Verbal and Visual Language

Hobbs (1979) famously proposed a systematic theory of discourse interpretation where maximizing coherence is a driving force behind various pragmatic inferences in communication and textual interpretation. Consider the simple discourse in (11):

(11) I missed another Zoom meeting this morning. My internet was out.

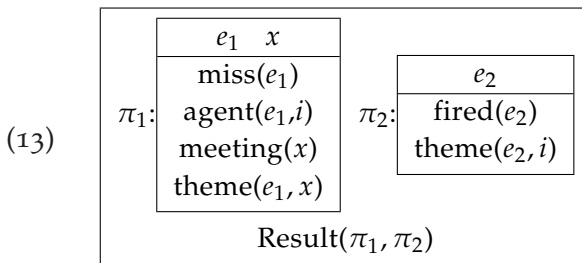
We don't merely interpret this as a conjunction of two eventualities occurring (missing a meeting and the internet being down), but almost automatically infer some kind of causal link between the two: I missed a meeting *because* my internet was out. Depending on the nature of the eventualities described we can infer different relations between them. While in (11) we inferred a relation commonly known as Explanation, in (12) we'll likely infer a different one called Result.

(12) I missed another Zoom meeting this morning. They fired me.

There are a number of more or less formalized theoretical frameworks describing the inference of these so-called coherence relations (Asher & Lascarides 2003; Kehler 2002; Mann & Thompson 1988). In all of them it is assumed that there is

a certain finite number of such relations, ultimately grounded in “more general principles of coherence that we apply in attempting to make sense out of the world we find ourselves in” (Hobbs 1990). Following Pagin (2014) and Cohen and Kehler (2021) I refer to coherence-driven inferences as a form of ‘pragmatic enrichment’ of the more minimal underlying semantic content.

The most comprehensive coherence theory, that is also immediately compatible with the formal semantic machinery we’ve introduced thus far, is called Segmented Discourse Representation Theory (SDRT, Asher and Lascarides 2003). In SDRT, discourse relations like Result, Explanation, Contrast, Background, and Narration are represented at a level of discourse representation that extends a given dynamic semantic framework, typically Discourse Representation Theory (DRT, Kamp 1981). The relata are elementary discourse units, typically corresponding to sentences or clauses that express propositions, typically describing the existence of certain events or states (Davidson 1967). Using special propositional discourse referents ( $\pi_1, \pi_2, \dots$ ) to label these elementary units and using DRT to represent their semantic contents, we get Segmented Discourse Representation Structures (SDRS) like (13)



In this traditional box notation, the outer box is an SDRS proper. It describes two discourse units,  $\pi_1$  and  $\pi_2$ , as related by the Result relation:  $\pi_2$  is the result of  $\pi_1$ . The smaller boxes are DRS’s, they represent the contents of the individual discourse units. The first discourse unit,  $\pi_1$ , corresponding to the first sentence of the discourse in (12), is characterized by this DRS box as (i) contributing two discourse referents, viz., an event  $e_1$  and an individual  $x$ , and (ii) ascribing a number of properties and relations to these discourse referents, viz., that  $e_1$  is an event of missing, that the agent of  $e_1$  (the person who is missing something) is  $i$  (a special indexical discourse referent picking out the actual speaker), etc.

The model-theoretic interpretation of coherence conditions in SDRT can be formalized as an extension of the semantics of DRT, which is a dynamic extension of first-order logic. For instance, Narration holds between two units iff the information carried by both units is true (or, more dynamically: if both units update the common ground consecutively) and the main event described by the first unit ( $e_{\pi_1}$ ) immediately precedes (or ‘occasions’, notation:  $\prec$ ) the main

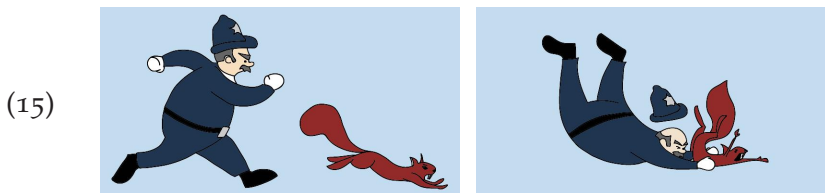
event described by the second. Note that this semantics presupposes that both units introduce a main event. With standard (S)DRT notation, that is,  $K_\pi$  is the DRS box associated with unit  $\pi$ ; and  $\oplus$  stands for DRS merge,<sup>11</sup> the DRT way of dynamic information updating (spelled out at the representational level of the DRS).

$$(14) \quad \llbracket \text{Narration}(\pi_1, \pi_2) \rrbracket = \llbracket K_{\pi_1} \oplus K_{\pi_2} \oplus \boxed{e_{\pi_1} \prec e_{\pi_2}} \rrbracket$$

In words, (14) says that interpreting two units  $(\pi_1, \pi_2)$  joined by Narration means roughly that we add the information of both units together (i.e., we join the sets of discourse referents they introduce, and we join the sets of conditions they impose on them), and then add a new condition that says that the main event described in the first unit immediately precedes that described in the second.

The introduction of coherence relations in the discourse interpretation process (i.e., the step by step construction of an SDRS from a sequence of utterances) is guided by a global constraint that seeks to maximize overall discourse coherence (i.e., add as many coherence relations as possible) and a number of defeasible pragmatic inference rules. For instance, if one unit  $\pi_1$  introduces a state and a subsequent (structurally accessible) unit  $\pi_2$  introduces an event, then all else being equal we can add ‘Background  $(\pi_1, \pi_2)$ ’ to the SDRS under construction. We’ll skip over all details of context change composition in the model-theoretic semantics, accessibility, complex discourse units, etc.—see Geurts, Beaver, and Maier (2020) for a gentle introduction to DRT and SDRT, and Asher and Lascarides (2003) for details.

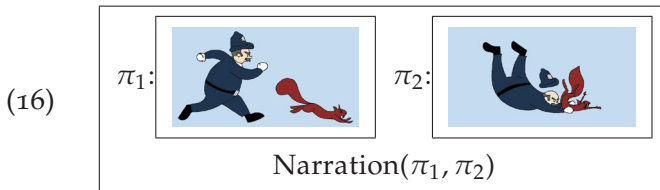
We’re interested in the application of coherence theory, and SDRT in particular, as a way of modeling pragmatic enrichment with partly or wholly pictorial discourses. First let’s rephrase Maier & Bimpikou’s (2019) DRT style analyses of purely pictorial narratives like (15) into the SDRT framework, by viewing the panels as elementary discourse units.<sup>12</sup>



11. Formally, a DRS is a pair consisting of a set of discourse referents ( $e_i, x_3, y$ , etc.) and a set of conditions (‘woman ( $x_2$ )’, ‘walk ( $e_2$ )’, etc.). DRS merge is defined as the pairwise union of these two sets, i.e., to merge two DRSs you just take the union of the two sets of discourse referents, and the union of the two sets of conditions.

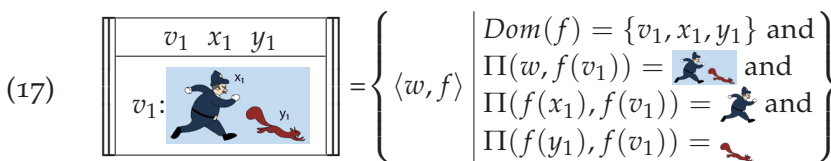
12. Drawings by Sofia Bimpikou, taken from Maier & Bimpikou (2019).

The basic assumption behind Maier & Bimpikou’s (2019) PicDRT is that pictures are like elementary discourse units, that is, they express information about what the world looks like. As outlined above, to interpret a sequence of propositional units—pictorial or linguistic—as a coherent narrative means that we infer coherence relations. In this case, and in many panel transitions in many comics, the inferred coherence relation defaults to Narration: the policeman is chasing a squirrel *and then* he catches it.



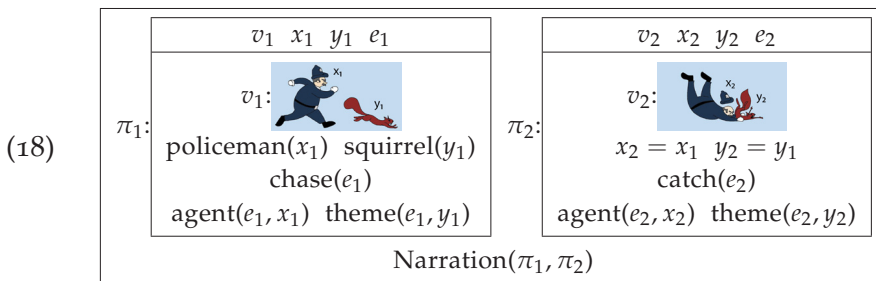
The first thing to note is that, unlike in competing dynamic semantic accounts of pictorial discourse (Abusch & Rooth 2017; Wildfeuer 2019), the semantic representation in (16) literally contains pictures as constituents of its DRS boxes. This is in line with the original motivation of (S)DRT as a model of human cognitive discourse processing, with the (S)DRS approximating a dynamically changing structured mental representation ‘in the hearer’s head’ (Kamp 1981). The idea behind PicDRT representations like (16) is that human mental representations can be partly symbolic (as modeled by discourse referents and first-order logical formulas like ‘agent ( $e_v, x$ )’), but also partly iconic and even pictorial (as modeled by the inclusion of picture conditions).

Zooming in on the pictorial DRS components in (16), Maier & Bimpikou (2019), inspired by Abusch (2012), add a ‘syntactic’ level of processing where pictures are labeled with viewpoint referents ( $v_v, v_2$ ), and what they call salient regions of interest in the picture are labeled with individual discourse referents ( $x_v, y_2$ ). A preliminary DRS representation of the first panel, with 2 salient regions introducing discourse referents, would be model-theoretically interpreted as in (17). Note that  $f$  is a partial assignment function, mapping discourse referents onto elements in the model’s domain (i.e.,  $f$  maps  $x_1$  to an individual,  $v_1$  to a viewpoint vector, and  $e_1$  to an event). The (dynamic) semantic content of a DRS is an ‘information state’, that is, the set of world–assignment pairs that verify the DRS (Nouwen, Brasoveanu, van Eijck, & Visser 2022).



Paraphrasing informally, the DRS in (17) contributes the information that (i) there is a certain viewpoint from which the world looks like the whole picture (i.e.,  $\Pi(w, f(v_1)) = \text{👮🐿}$ ); (ii) there’s an individual that, when projected from that same viewpoint, looks like the bluish region (i.e.,  $\Pi(f(x_1), f(v_1)) = \text{👮}$ ); and (iii) there’s another individual that looks like the smaller brownish region (i.e.,  $\Pi(f(y_1), f(v_1)) = \text{🐿}$ ).

At a post-semantic level, based on general world-knowledge, genre, and background information about what things look like under common projections, properties and relations may be freely predicated of these discourse referents (e.g., ‘policeman ( $x_1$ )’), as a form of ‘free pragmatic enrichment’ (Recanati 2010). Moreover, different discourse referents from different pictures can be equated (e.g.,  $x_2 = x_1$ —a free pragmatic pictorial analogue of anaphora resolution, Abusch 2012). In addition to the discourse unit labels ( $\pi_1, \pi_2$ ) and coherence relations (‘Narration( $\pi_1, \pi_2$ )’) sketched in (17), we further add to the post-semantic enrichment stage the introduction of event discourse referents ( $e_1, e_2, \dots$ ). Note that this last enrichment is crucially driven by the semantics of Narration, which, as defined in (14), presupposes that both units introduce an event discourse referent.



The model-theoretic interpretation of (18), the post-semantically enriched version of the SDRS in (17), is a straightforward enrichment of the information state in (17) involving only standard DRT and SDRT semantics (but I will skip over the formalities here).

In sum: the semantics proper of a single picture is very minimal, basically, ‘the world looks like this at some point in space and time’. When presented with a few pictures in a seemingly deliberate sequence we go beyond mere conjunction of those minimal propositions (the world looks like this at some point and like that at some point), just like we do when presented with a series of utterances.<sup>13</sup> The sequencing thus triggers a cognitive enrichment process that crucially

13. As a vivid illustration of the human tendency to infer coherence relations between utterances consider a discourse like: ‘John took a train from Paris to Istanbul. He likes spinach’ (Hobbs 1979). Although the reader may have never connected Istanbul and spinach, and the actual words don’t specify one either, they’ll naturally infer some causal connection between the two pieces of information.



involves the inference of various coherence relations in order to satisfy a global desire for maximal coherence.

Since the coherence-driven enrichment mechanism (here formalized in SDRT) thus applies uniformly to verbal and visual discourse, it should be well suited to modeling multimodal mixtures, like comics with textual elements (Wildfeuer 2019), illustrations with captions (Rooth & Abusch 2019) or tags (Greenberg 2019), and film (Cumming, Greenberg, & Kelly 2017; Wildfeuer 2014). If emojis are pictures, semantically, this same machinery should help us enrich the communicative content of emojis in relation to the surrounding text and/or other emojis.

### 3.2. *Emojis and Coherence*

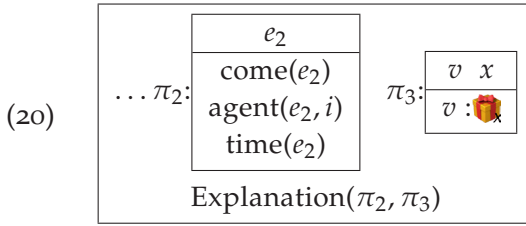
We've assigned a minimal, pictorial content to object emojis like 📦 that can be roughly paraphrased as 'there is a viewpoint near where there is some object that looks like that.' This semantic content is a proposition, or, if we follow the DRT approach sketched in the previous section, the dynamic equivalent of a proposition, viz. an information state.<sup>14</sup> Following Lascarides and Stone's (2009) original analysis of speech–gesture integration, but more directly following Grosz et al.'s (2021) analysis of activity emojis, these propositional emoji uses can be analyzed as discourse units in their own right, alongside the textual units.

(19)  $\pi_1$ : Happy Birthday!  $\pi_2$ : I'm coming over this afternoon  $\pi_3$ : 📦

Maximizing coherence means inferring coherence relations between these discourse units.  $\pi_2$  contributes the existence of an event of the speaker coming over in the afternoon of the utterance day, while  $\pi_3$  contributes, roughly, the existence of a gold/brown box with a red bow at some point in space and time. The conjunction of those two pieces of information as such is not a coherent discourse, so we infer a relation, probably a causal relation (the box is the reason for the visit), which in SDRT would be called 'Explanation':

---

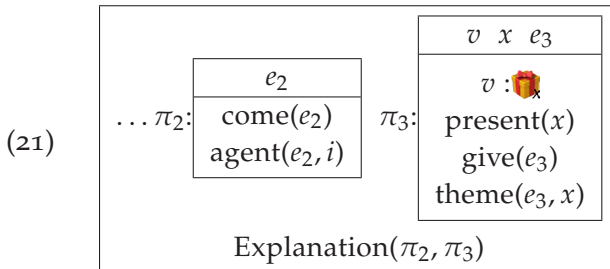
14. Arguably, this is fundamentally the wrong semantic type for analyzing 'pro-speech emojis', i.e., emojis replacing a specific word or phrase, which is part of the reason why I proposed in Section 1 to disregard these uses in this paper. Still, the occurrence of pro-speech emojis does not count against the current endeavor. In principle, the formal machinery I'm developing in this paper, in particular the pictorial analysis of emojis, could well be extended to build semantic constituents of more complex, subsentential types that could be integrated in the grammatical structure of a sentence.



The semantics of Explanation, like Narration, (14), demands two events, and says that the second unit’s main event causes the first’s. Thus, the inference of an Explanation relation (to increase coherence), triggers the further inference that the picture depicts not just what the world looks like, but contributes an event. But how does a picture of a box depict an event?<sup>15</sup>

At this point I defer to what cognitive scientists call general cognitive schemas, scripts, or frames (Fillmore 2008): things that look like that, that is, nicely wrapped boxes with bows, typically contain gifts, and gifts are typically quite saliently involved in events of giving and receiving. Note that this is the same reasoning as what gave rise the inference that the entity depicted by the mostly blue shape in the comic in (15) is (probably) a police officer and that the inferred position of his arms legs are (probably) snapshots of him running (rather than assuming a weird pose and floating in the air).

With all the defeasible pragmatic enrichments above, the coherent interpretation of the tweet in (19) now looks something like this:



Paraphrasing the interpretation of the SDRS in (21): there’s an event of the speaker coming over and an event of giving a present that looks like this, 🎁, and the latter event explains the former.

15. Altshuler and Schlöder (2021) maintain what they call Abusch’s Constraint: pictures can only contribute states (viz., in the current informal terminology, what the world is/looks like), not events. On my view, Abusch’s Constraint makes sense at the level of the purely semantic, pictorial content, where perhaps we could take it one step further by maintaining that those minimal contents are devoid of eventualities altogether. In any case, Abusch’s Constraint does not hold if we consider the total, pragmatically enriched contribution of the picture in context.

But does the gift really have to look just like that, that is, in a gold-colored box with a red bow? Of course, we're assuming that Apple's projection function includes some abstraction and stylization, leading to some indeterminacy about the actual size, shape, color, lighting, and background of the box that is depicted, but what if it's a blue box with a yellow bow, or a ball wrapped in newspaper without a bow, or even just an electronic gift certificate? As it stands, the pictorial account would make the speaker a liar<sup>16</sup> if she meant to give such a gift, which is obviously absurd. The emoji can be used for almost any kind of gift, it doesn't really have to look like this 📦. I consider this the fundamental challenge for a truly pictorial account of emojis, and I address it in the next section.

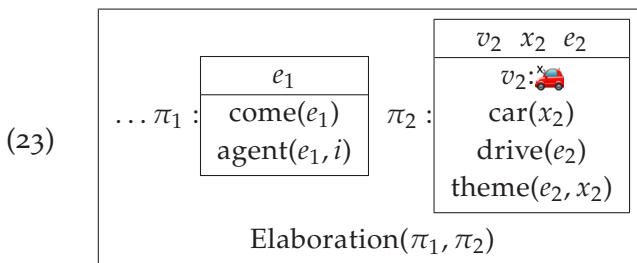
## 4. Emoji Pragmatics: Figurative Depiction

### 4.1. The Pictorial Overdetermination Challenge

To illustrate our discourse semantics framework once more, consider another example tweet:

(22) I'm coming over this afternoon 🚗

In this variation of (19) the emoji still clearly counts as a separate discourse unit, but now the connection is likely one of Elaboration (the event of my coming over involves a car) rather than Explanation.<sup>17</sup>



In words: there's an event of the speaker coming over and that event includes the event of driving a car that looks like this. The phrase 'looks like' as always has to be understood in terms of projection.

Since Apple's emojis tend to have various different colors that in many ways seem to reflect the actual colors of the depicted objects with some degree

16. I'm assuming with Viebahn (2019) that one can lie with pictures.

17. Though in context it's of course possible that the speaker is explaining that she's coming over in order to return the speaker's car, or gift them one, etc.

of faithfulness,<sup>18</sup> we might reasonably assume that Apple’s projection function approximates actual color (within a finite set of fixed color codes). But then, counterintuitively, (22) would be false if the speaker drives a silver car.<sup>19</sup>

A first attempt at addressing this color mismatch problem in particular is to assume that apparently color is not in fact preserved in the relevant projections. Instead, just like we’re already abstracting away from details of shape, texture, shadow, etc., the Apple projection apparently also ignores most colors and just maps certain real-world surfaces to a default red. It’s tricky to define exactly what colors get mapped to red, and which to black, white, and to the various shades of grey and or blue that occur in this particular car emoji, and how these color marking rules should be adjusted for different categories of objects and their emoji depictions. These may just be technicalities, in which case it’s worth noting that the resulting projection would intuitively count as a pictorial mapping, in keeping with our starting point that the car emoji is a car picture.

However, this leaves us with a parallel problem of shape mismatches. What if the intended present is a round object wrapped in newspaper? Or the car is a big black 4 door BMW that looks nothing like 🚗? Extending the color solution outlined above to shapes would mean that Apple’s projection specifies a fixed shape and then effectively maps every car to that shape. Now note that such a projection would be essentially concept-based in the sense that anything that falls under the concept of ‘car’, no matter what it looks like, gets mapped onto the car emoji. This would take us away from pictorial representation and well into symbolic territory. In fact, this projection function is literally just the inverse of the lexical semantic meaning of the English word *car*, so we’d end up with a symbolic rather than an iconic account.

Before I present a solution to the overdetermination challenge that avoids going symbolic like this, let’s first explore this alternative

#### 4.2. Creativity and Symbolic Overdetermination

As alluded to in Section 1, the apparent default view of emoji semantics treats emojis as essentially an extension of the lexicon of a certain genre of written language. In (Grosz et al. 2021: 348), for instance, activity emojis are said to “serve as free-standing event descriptions, whose core argument is anaphoric”. Although they try to remain agnostic on the details of the lexical entry of an activity emoji

---

18. To see this, think of the emojis for entities like footballs, trees, sheep, but also people and hand emojis with various skin tones, see Section 5.4. Also, when I talk about ‘actual colors’ of objects I’m assuming canonical lighting conditions.

19. What’s worse, if it’s a text message sent from an Android device, the speaker themselves might see Google’s silver car emoji while the recipient might get Apple’s red one.

like 🎻 (because it has to cover both playing the violin and being a violinist, among other things), their notion of an event *description* that contains an *argument* that is moreover *anaphoric*, points in the direction of a language-like, that is, symbolic, lexicon extension, rather than a strictly pictorial semantics like the one I'm defending here.<sup>20</sup>

Interestingly, when we look at the variety of uses of our limited set of emojis, it becomes apparent that the symbolic approach also runs into an overdetermination problem. While it avoids overdetermining what the object looks like, it overdetermines its conceptual classification. In some cases, 📦 really *does* denote something that looks like that without being a gift. For instance, in the tweet in (24) the advertiser is most likely creatively using the 'wrapped present' emoji to elaborate on the event of collecting parcels, relying on the fact that parcels are often boxes that kind of look like that.

- (24) We deliver in South Africa via pep store for R59.95 or you can collect your parcels at Ferndale, Randburg 📦

A symbolic account that treats 📦 as a word that's a synonym of 'gift' or 'wrapped present' or that otherwise ties the meaning symbolically to the concept of a gift,<sup>21</sup> would predict that (24) would be infelicitous if the author intended to include people simply ordering stuff for themselves rather than as a gift.

In this particular case, the symbolist might object that the use in (24) *is* infelicitous—the author should have chosen 📦 to illustrate the general concept of parcels or delivery.<sup>22</sup> But the same kind of creative usage of course happens when there is no better alternative emoji. For instance, on a symbolic account, 🎻 probably denotes something involving a violin. But that would exclude uses where it denotes playing a viola, or a cello (neither of which have their own dedicated emoji).

- (25) Double-cello action in #Arensky's beautiful quartet has inspired our two cellists to treat the audience to some bonus duos... 🎻 🎻

20. See for instance Abusch's (2012) for the claim that there is no real anaphoricity in pictures. Purely semantically, in the police/squirrel comic in (15), both panels semantically depict a police officer, so it is only by defeasible pragmatic inference (based on world-knowledge and coherence) that we may equate the discourse referents introduced by the bluish shapes in consecutive panels. Unlike in the linguistic re-telling (*A policeman chased a squirrel. He caught it*), where we have first indefinites and then pronouns, nothing in the second panel itself tells us to look for an antecedent to bind the new discourse referents to.

21. Extrapolating Grosz et al.'s (2021) remarks, they would have a lexical entry of the form:

[[📦]] =  $\lambda e \exists y. (\text{present}(y) \wedge \text{agent}(e, x) \wedge \text{theme}(e, y))$ .

22. An anonymous referee in fact objects thus. Interestingly, this same referee notes earlier that they often use 📦 to refer to celebrations quite generally (i.e., metonymically, see 4.4 below), which suggests individual differences in the limited active emoji repertoires of speakers can have consequences for the range of creativity they allow in emoji interpretation.

In light of (25) the symbolist might propose ‘bleaching’ their lexical entry to accommodate cellos (e.g., to ‘bowed classical instrument’), but there will always be new, unforeseen use cases that don’t quite fit any proposed lexical definition (e.g., *bring your ukulele!* 🎸). We’ll discuss a more extreme case in the next subsection and suggest an appeal to metaphoric and other figurative meaning extensions to deal with such emoji usages. This appeal is in principle open to both symbolic and pictorial accounts, but I’ll argue that it works best with the pictorial account.

### 4.3. 🍆 as Image Metaphor

A good illustration of the flexibility of emoji meaning involves the well-known use of 🍆 and 🍑 to refer to somewhat taboo body parts and events involving them. On a symbolic account, we might theoretically give a lexical semantic interpretation of 🍆 that includes both eggplants and male genitalia. But probably a more intuitive approach would have eggplants as the literal meaning and derive the other use as a secondary,<sup>23</sup> non-literal meaning somehow. But what kind of non-literal meaning is this?

Lakoff (1993) uses the term ‘image metaphor’ to describe a metaphorical interpretation based on visual resemblance between the literal meaning (‘source’) and the metaphorical interpretation (‘target’). He illustrates the phenomenon with linguistic examples like (26):

- (26) a. My wife... whose waist is an hourglass. (André Breton, cited by Lakoff 1993)  
 b. His toes were like the keyboard of a spinet. (Rabelais, cited by Lakoff 1993)  
 c. The road snaked through the desert (Barnden 2010)

Note for instance that the waist in (26-a) is not in any way conceptually related to an hourglass—it doesn’t help keep time, for instance—it just looks like one.

Examples like (26) show the need for a general account of image metaphor, wholly independently of emoji or pictorial semantics. Without getting into the details here, we’ll assume such an independent account, say Lakoff’s, or, more conveniently integrated within the kind of SDRT framework we’re already using, Agerri, Barnden, Lee, and Wallington’s (2007). Now, proponents of both the pictorial and the symbolic account could appeal to that account to explain

---

23. Of course, in this case ‘secondary’ does not mean less frequent. Also, like linguistic metaphor and metonymy, non-literal picture uses may start as genuinely pragmatic inferences and then gradually become stale and entrenched until they eventually can be said to become part of the lexicon, perhaps even replacing the original literal meaning.

the common non-literal interpretation of the eggplant emoji. Interestingly, since the symbolist makes the eggplant emoji roughly synonymous with a linguistic utterance of ‘(there’s an) eggplant’ we would expect the eggplant–penis metaphor to occur linguistically as well—and occasionally it does:

- (27) The Warri pikin took to his IG account this morning to flaunt his eggplant in wet white underwear.<sup>24</sup>

Although examples like (27) are not hard to google, they really are not that common either. In fact, it’s quite possible that many of the linguistic instantiations of this particular image metaphor are derivative on the widespread emoji usage.

On my pictorial account of 🍆, we can readily explain why the emoji instantiation is so much more prominent than its linguistic cousin. The the emoji literally tells us what the world looks like, and hence, when we get to the level of pragmatic enrichment, this pictorial content is readily extended with further image-based inferences. In cognitive processing terms, the basic pictorial semantics would predict engagement of the visual system in semantic processing already,<sup>25</sup> so we can expect image-metaphoric pragmatic processing to be a natural follow-up, that is, exploiting more cognitive processing terminology, the image metaphoric pragmatics is primed by the pictorial semantics.<sup>26</sup>

#### 4.4. Varieties of Figurative Meaning in Emoji Interpretation

Generalizing beyond 🍆, the interpretation process I propose is as follows. The literal semantics of the emoji is projective: from some viewpoint, the world projects onto this, 🚗, which means it contains a red car shaped object. With this minimal meaning in hand (e.g., (mentally) represented in the form of a basic pictorial DRS condition), we enter into the realm of pragmatics, which includes various kinds of pragmatic enrichments, including the inference of coherence relations,

24. <https://gossipnaija.ng/2019/12/tuoyo-is-at-again-as-he-flaunts-his-eggplant-in-wet-white-speedo/>

25. Note that DRT is often viewed as providing a level of semantic representation relevant for describing human linguistic processing (Brasoveanu & Dotlacil 2015). Our current use of pictorial conditions in (S)DRSs very strongly suggests an account where pictorial processing (including then the interpretation of emojis) must quite literally engage the visual system (see also the earlier remark about our choice of including pictures in DRS conditions in Section 3.1).

26. Although image metaphor and picture semantics are thus closely related and perhaps even continuous with each other, we should be careful not to treat them as a single phenomenon. If we were to reduce Lakoff’s notion of visual resemblance to geometric projection, we would be back where we started, that is, facing the pictorial overdetermination challenge of Section 4.1: no geometric projection function will map both a small red Honda coupe and a big black BMW sedan unto this picture 🚗 (unless that ‘projection’ is itself concept-based and hence not really pictorial).

properties, events, (as described in Section 3) but also, typically, finding a non-literal meaning whenever that fits the context better.<sup>27,28</sup>

In the eggplant case, deriving this figurative meaning involved a rather pure image metaphor, for example, mapping the depicted vegetable to a body part on the basis of a visual resemblance. In other cases, the metaphor may be partly based on conceptual similarity. This is unproblematic for both symbolic and pictorial accounts, because since Lakoff and Johnson (2003) it is commonly accepted that metaphors involve similarities or analogies at the level of semantic content (i.e., mental concepts, in their cognitive semantic framework), rather than at a strictly linguistic level. This means that even if image metaphors are the most natural complements to our proposed pictorial semantics, any other kind of figurative interpretation we find in text or speech can in principle be applied to emojis as well—after all, on the projective account, pictures and linguistic utterances express similar semantic contents (viz., possible worlds propositions, or their dynamic equivalents, information states).

Let's apply this to the examples illustrating the pictorial overdetermination challenge in Section 4.1. In the case of 🚗 referring to a big black BMW, the interpreter may map the depicted red car shaped object to various makes and models of cars on the basis of a mixture of resemblance and conceptual similarity (viz., all being cars). In the case of 📦 referring to an electronic gift certificate, we have to move beyond image metaphor altogether and assume a purely conceptual mapping (from box with bow to gift card, on the basis of both being gifts).

Finally, in addition to metaphor, there are many metonymic uses of emojis that likewise require non-resemblance-based mappings. For instance, 📷 (literally) depicts an old-fashioned camera, but can be used metonymically to denote photos, and 🥑 literally depicts a half avocado, but can be used metonymically to denote avocados generally, or healthy vegan food more generally (28-a), or even the typical consumers of said food (28-b).

---

27. We can think of the various non-literal meaning extensions uniformly as driven by coherence maximization, just like the more basic enrichments described in Section 3. One might eventually exploit the formal machinery of SDRT to actually develop a unified framework for coherence-driven pragmatic enrichments, including metaphor and metonymy, but as far as I'm aware there is no such framework yet, and building one here would take us too far afield (since this framework would transcend emoji and pictorial discourse anyway). See Agerri et al. (2007) for a step in this direction, viz. a formal, computational account of metaphoric mapping between SDRT representation.

28. As mentioned above, any figurative interpretation can turn stale over time and thereby drift from pragmatics into semantics. This conventionalization process may eventually make the intermediate pictorial semantics obsolete, at which point we might admit that the emoji has become a symbol. Perhaps this is what happened with ❤️ (long before the advent of emojis proper), or perhaps, as one referee suggests, even with 🍆. I leave further investigation of the proper diagnostics of metaphor lexicalization (in both the linguistic and the pictorial domain) for future research.



- (28) a. Forever mad at myself for taking so long to go vegan 🥑  
 b. Proud to be #Hipster 🥑

To sum up, the pictorial account of emojis suggests a continuity between emoji semantics and image-metaphoric pragmatics, which correctly predicts the widespread use of image metaphors in emoji usage (see 🍆). In addition, pragmatic emoji interpretation on my pictorial account—as on any symbolic alternative—may also invoke (partly) conceptual metaphor, or various forms of metonymy. All kinds of non-literal meanings can be associated with any concept, whether it's introduced symbolically by a word, or pictorially by a painting, animated gif, sticker, or emoji.

A proper description and formalization of ' (vague) resemblance' (either projectively or non-projectively), of 'conceptual similarity', of hybrid image-based/conceptual metaphors, of the metonymy–metaphor distinction, of the integration of metaphoric meaning extensions in the coherence-driven SDRT account of pragmatic enrichment, and of the conventional entrenchment and eventual lexicalization of stale metaphors/metonyms over time, is all well beyond the scope of the current paper. To defend these substantial omissions I can only point out that accounts of all these phenomena are already independently needed for the proper analysis of any figurative meaning in any kind of language, and hence in no way tied to the interpretation of emojis or pictures specifically.

## 5. Face Emojis and Expressives

### 5.1. *The Special Status of Face Emojis*

Up until this point we have been focusing almost entirely on a specific subclass of emojis, viz., those depicting familiar, concrete objects. In actual usage, object emojis however are decidedly less common than emojis depicting expressive parts of the body, especially faces and hands. According to emoji tracker, the top 20 emojis include 14 face and 2 hand emojis, and 0 object or event emojis (there is also a recycling symbol and 3 types of heart emojis, which I already put aside as potentially symbolic rather than pictorial in Section 1).

Apart from some 'symbolic modifiers' (like the heart-shaped eyes in 😍, for which in Section 1 I deferred to Maier 2019), on my account these face emojis are just as pictorial as the object emojis discussed above, or as animated gifs, cartoons, or manga panels. Nonetheless, they are known to interact somewhat differently with the surrounding text. According to Grosz et al. (2021; in press), Kaiser and Grosz (2021), face emojis are expressives, meaning that they are used to express the speaker's emotional state, roughly the same way verbal expres-

sives do. Thus, the two utterances in (29) mean roughly the same: Kate said that Sue sent the report and I have a negative emotional attitude about that.

- (29) a. kate said that sue sent the report to ann 🙄 (Grosz et al. 2021)  
 b. kate said that sue sent the f cking report to ann (Grosz et al. 2021)

Potts (2007) lists the defining characteristics of expressives: their contribution is hard to paraphrase precisely in non-expressive terms; they are speaker-oriented ('indexical') and (hence) unaffected by embeddings (but in some special cases may be subject to pragmatic 'perspective shift' in the sense of Amaral, Roberts, and Smith 2007; Harris and Potts 2010) and they are infinitely gradable (e.g., by varying intonation or repetition).

Emojis satisfy Potts's characteristics. Regarding (i), the 🙄 in (29-a) indicates a negative attitude, but the linguistic paraphrase I gave above is just a rough approximation, not by any means semantically or pragmatically equivalent. Regarding (ii), Grosz et al. (2021) show that face emojis—unlike activity emojis—tend to express the emotional state of the producer of the utterance, while activity emojis can be anchored to any salient entity, depending on what connection creates the most coherent output.

- (30) a. Sue's on her way now 😊  
 ~ ... and {I'm/\*she's} happy about that  
 b. Sue's on her way now 🚗  
 ~ ... and {\*I'm/she's} traveling by car

What's more, face emojis tend to project out of embeddings, while activity emojis can also be interpreted under negation:

- (31) a. If I had gone, I'd have missed Ada 😊  
 ~ I'm happy (that I didn't go, because now I could hang out with my friend Ada)  
 ✗ If I'd gone, I'd have been happy (because then I'd have missed that annoying Ada)  
 b. By now, Sue hasn't trained for months 🏄 (Grosz et al. 2021)  
 ~ surfing is part of the training<sup>29</sup>

Furthermore, Kaiser and Grosz (2021) show experimentally that face emojis are not always anchored to the actual speaker, but like linguistic expressives

29. Pierini (2021) follows Schlenker's (2018) analysis of co-speech gestures as contributing co-suppositions, i.e., (31-b) licences the inference that if Sue had trained it would have involved surfing. Grosz et al. (2021) leave open this possibility as an alternative to their coherence-based analysis that should be able to derive the embedded interpretation by construing the negation scoping over a complex discourse unit consisting of the training and the surfing connected via Elaboration.

may indeed be subject to a perspective shift, for instance in constructions with a salient third-person experiencer argument the face emoji may be interpreted as conveying either the speaker's or the experiencer's attitude:

- (32) Anna admired Betty 😊  
 ~ I'm glad about that  
 ~ Anna has a positive attitude

Regarding (iii), while face emojis themselves are not as flexible as Wittgenstein's suggestion of drawing expressive faces by hand, their emotive content can be scaled indefinitely by creating sequences of similar or the same emojis (McCulloch & Gawne 2018):

- (33) Omggggggg he's so cute 🥰🥰🥰🥰🥰🥰

The linguistic data reviewed above strongly suggest that face emojis are first of all semantically different from the object and event emojis that we discussed in the previous section, and second of all that they seem to be expressives. In this section I reconcile these observations with my primary claim that emojis—face, object, and event emojis alike—are pictures. This requires that we first get clear on what expressives are and how to analyze them semantically (*viz.*, in terms of use conditions). I then argue that many facial expressions and hand gestures are really expressives, and that face and hand emojis are 'use-conditional pictures' of such expressive gestures.

## 5.2. Expressivism, Expressives, and Use-Conditional Content

Expressivism is the view that some linguistic constructions can express meaningful semantic content that does not contribute to the derivation of truth-conditional content. Philosophers and linguists, more or less independently of each other, have provided expressivist accounts of ethical and esthetic vocabulary, knowledge ascriptions, mental state self-ascriptions, exclamatives, epithets, slurs, etc. What exactly is expressed by these constructions or statements containing them is a matter of debate, ranging from the emotional state of the speaker (Ayer 1936; Potts 2007; Stevenson 1944) to a more abstract semantic notion like use-conditional content (Charlow 2015; Gutzmann 2015; Kaplan 1999; Predelli 2013). While Grosz et al. (2021) opt for a more Pottsian (2007) analysis (defining emotive content in terms of real intervals signifying emotional valence), I'll introduce and adopt the latter, more minimalistic approach to expressive content, which has the benefit of not forcing the semanticist to make any assumptions about the underlying cognitive architecture of emotions.

The use-conditional analysis of expressives can be traced back, again, to Wittgenstein:

We ask, ‘What does “I am frightened” really mean, what am I referring to when I say it?’ And, of course, we find no answer, or one that is inadequate. The question is: ‘In what sort of context does it occur?’ (Wittgenstein 1958)

In other words, expressive utterances are not amenable to a standard compositional semantic treatment in terms of reference and truth. “I am frightened” is not so much a (truth-evaluable) assertion about what the world is like, but rather an expression of the speaker’s emotional state. Instead of trying to capture the propositional content, that is, the set of worlds where the sentence is true, we should look for the ‘contexts of use’. While Wittgenstein himself takes this idea much further, turning ‘meaning is use’ into a general characterization of linguistic meaning across the board, Kaplan (1999) offers a way to isolate this insight about the meaning of expressive vocabulary and integrate it into an otherwise traditional formal semantic framework.

There are words that have a meaning, or at least for which we can give their meaning, words like ‘fortnight’ and ‘feral’. There are also words that don’t seem to have a meaning, words like [‘ouch’ and ‘oops’]. If the latter have a meaning, they’re at least hard to define. Still, they have a use, and those who know English know how to use them. (Kaplan 1999)

Gutzmann (2015) works out the details of semantic composition, adding significant extensions to Kaplan’s program. I’ll adopt some of Gutzmann’s implementation and notation below. The general idea is that in semantics we encounter two types of content: descriptive (or truth-conditional) and expressive (or use-conditional). Some expressions carry only descriptive content (‘flower’, ‘walk’, ‘every’) and combining them into a sentence will give us its truth conditions, in linguistics typically captured as a possible worlds proposition. In the following we’ll use  $\llbracket \alpha \rrbracket$  to denote the descriptive content of a term  $\alpha$ .

To deal with indexicals, Kaplan (1989) had already introduced a second semantic parameter,  $c$ , to the semantics. That way we can model the truth-conditional proposition expressed by an utterance of an expression in a context, (34-b), as well as the more abstract ‘character’ modelling the descriptive linguistic meaning of the sentence, (34-c). Notation:  $sp_c$  and  $ad_c$  denote the speaker/writer of context  $c$  and hearer/reader/interpreter/addressee of  $c$ , respectively.

- (34) a. truth condition:  $\llbracket \text{I see you} \rrbracket_w^c = 1$  iff  $\langle \llbracket \text{I} \rrbracket_w^c, \llbracket \text{you} \rrbracket_w^c \rangle \in \llbracket \text{see} \rrbracket_w^c$  iff  $\langle sp_c, ad_c \rangle \in \llbracket \text{see} \rrbracket_w^c$  iff  $sp_c$  sees  $ad_c$  in  $w$ .
- b. truth-conditional content:  $\llbracket \text{I see you} \rrbracket^c = \left\{ w \mid sp_c \text{ sees } ad_c \text{ in } w \right\}$
- c. truth-conditional content:  $\llbracket \text{I see you} \rrbracket^c = \left\{ w \mid sp_c \text{ sees } ad_c \text{ in } w \right\}$

As we saw in the quotation above, the starting assumption of Kaplan's (1999) expressivism was that there are some expressions that do not contribute to this truth-conditional content (or character), but instead express content we can model in terms of use conditions. Take Kaplan's central example: 'Oops'. While it's weird to judge 'oops' as either true or false, we can judge whether a particular 'oops' was uttered felicitously on a given occasion by a given speaker. For instance, in the context where someone just saw a car run over and kill their family's beloved pet, an 'oops' would be infelicitous, because the word 'oops' seems reserved for what Kaplan calls 'minor mishaps', as captured in the use condition in (35-a). We then define use-conditional content as a set of contexts — those where the expression is felicitously used, (35-b).

- (35) a. use condition: a use of 'oops' is felicitously uttered in  $c$  iff the speaker just observed a minor mishap in  $c$
- b. use-conditional content:

$$\llbracket \text{oops} \rrbracket = \left\{ c \mid sp_c \text{ observed a minor mishap in } w_c \right\}$$

Gutzmann goes on to set up a type system to model the compositional contributions of hybrid and subsentential expressives, but first he introduces a nice fracture notation for Logical Forms (LF) that puts expressives (and their use-conditional interpretations) on top, and descriptions (and their truth-conditional interpretations) below.<sup>30</sup>

$$(36) \quad \llbracket \text{Oops, I did it again} \rrbracket = \left[ \left[ \frac{\text{Oops}}{\text{I did it again}} \right] \right] = \frac{\llbracket \text{Oops} \rrbracket}{\llbracket \text{I did it again} \rrbracket}$$

$$= \frac{\left\{ c \mid sp_c \text{ observed a minor mishap in } w_c \right\}}{\left\{ \langle c, w \rangle \mid sp_c \text{ did it again in } w \right\}}$$

30. I could present all formulas below using the (S)DRT box notation. The fractured LF would then give rise to a fractured (S)DRS, which would in turn be model-theoretically interpretable as a pair consisting of a use-conditional content and a truth-conditional content. Neither the intermediate representational DRS level (between LF and model-theoretic interpretation), nor the dynamic semantics of information states and updates would add any deeper insight into the phenomena or theory here. I choose therefore to stick with familiar, straightforward semantic interpretations in possible worlds semantics.

The full linguistic meaning of a sentence with some expressives is thus a pair consisting of a use-conditional content and a truth-conditional character.<sup>31</sup>

### 5.3. Expressive Gestures

Emblematic gestures like the middle finger or waving goodbye are sometimes said to be non-verbal expressives (Ebert 2014). Indeed, we can easily verify this by checking off Potts's (2007) criteria, the way we already did for emojis above. For instance, the meaning of the middle finger gesture concerns the speaker's attitude (towards their addressee); it is surely negative but hard to pin down with a purely descriptive paraphrase; and it can be graded continuously by exaggerating or repeating the gesture (or combining it with facial expressions or verbal expressives).<sup>32</sup> Since these emblematic gestures are as much conventional, intentional, symbolic, and hence as 'language-like' as Kaplan's verbal examples 'ouch' and 'oops', it makes sense to analyze them semantically on a par, that is, as contributing use-conditional content.

- (37) a. use condition: use of middle finger gesture is felicitous iff the speaker is very annoyed at the addressee  
 b. use-conditional content:

$$\llbracket \langle \text{middlefinger} \rangle \rrbracket = \left\{ c \mid sp_c \text{ is very annoyed at } ad_c \text{ in } w_c \right\}$$

Crucially, as with verbal expressives, there are both felicitous and infelicitous uses of the middle finger. Someone who gives their neighbor the finger to greet her is doing something wrong, or at least breaking an established convention, as is someone who gives someone the finger while she is angry at someone else.

31. It is worth noting that both Gutzmann and Kaplan suggest different ways to define a more traditional, unidimensional 'informative content' of the sentence on the basis of these two levels. Different ways of 'collapsing' will give rise to different notions of informative content, useful for validating different intuitive inference schemas. For instance, as it stands, without collapsing the two meaning dimensions, we can't really account for the evident (Moorean) paradoxicality of statements like (i), where the levels of meaning conflict:

(i) I really like you, you fucking asshole.

I refer the reader to Kaplan and Gutzmann for some discussion of collapsing options, most of which would indeed predict infelicity for (i), and hence help explain why examples like these give rise to irony or otherwise non-literal re-interpretations of either the expressive or the descriptive part of the message.

32. Potts's 'perspective shift' criterion may be the hardest. It's hard to come up with a situation — outside of pretending, acting, or quoting — where you give someone the finger but intend it to illustrate the annoyance of a third person. However, one might say that, if anything, a lack of perspective shifting makes these gestures even more expressive than their less rigid verbal counterparts.

Hence, the use-conditional content provided by the definition in (37) is non-trivial and arguably approximates the gesture's core linguistic meaning.

Now, the same considerations apply to (some) facial expressions. Though a smile, unlike the middle finger gesture, is to some extent more natural and perhaps even culturally universal (Darwin 1872), and not always intentional or conscious, we can still say that it is felicitous if the speaker has a friendly disposition towards the addressee, and infelicitous otherwise.<sup>33</sup> Notation: I'm using an 'overline' notation borrowed from sign language studies to denote co-speech gestures.

$$(38) \quad \left[ \overline{\text{I'm coming over}}^{\langle \text{smile} \rangle} \right] = \frac{[\langle \text{smile} \rangle]}{[\overline{\text{I'm coming over}}]} = \\ = \frac{\{c \mid sp_c \text{ has a friendly disposition towards } ad_c \text{ in } c\}}{\{\langle c, w \rangle \mid sp_c \text{ is coming over in } w\}}$$

#### 5.4. Face Emojis as Use-Conditional Pictures

Facial expressions and face emojis both behave like expressives, as we verified by checking off Potts's criteria in 5.3 and 5.1, respectively. But only face emojis are at the same time pictures. I've proposed viewing face emojis as pictures of facial expressions, which are in turn expressives. But this does not immediately explain why the emojis themselves behave like expressives.

To close the gap between 'picture of expressive' and 'expressive behavior' we could first try to appeal to some kind of pictorial transparency. A representational system is called transparent iff in that system a representation of a representation of X is itself (interpreted as) a representation of X (Kulvicki 2003; 2006). Some forms of pictorial representation are indeed sometimes viewed as transparent: a drawing of a drawing of a mountain is, arguably, in some cases, also a drawing of a mountain.<sup>34</sup> Linguistic description, by contrast, is not usually transparent: a linguistic description of a linguistic description of a mountain is a description of a sequence of words, not of a mountain. What we really need for our current purposes is a cross-medial transparency principle that allows us

33. As Sarah Zobel and Thomas Weskott (p.c.) suggest there may actually be two semantically distinct types of smiles, one purely emotive, expressing happiness/contentedness, and one more communicative, explicitly directed at an addressee, showing friendliness. It remains to be seen if this distinction somehow correlates with that between so-called Duchenne and non-Duchenne smiles, where the former is the more natural indicator of happiness and the latter is the more controllable, deliberate gesture (Ekman, Davidson, & Friesen 1990). See also Ginzburg, Mazzocconi, and Tian (2020) for a detailed semantic analysis of smiles and laughter.

34. I mean here 'drawing of a mountain' not in the causal, *de re* sense, but in what Kulvicki calls the 'syntactic' sense paraphrased as 'a mountain-drawing'.

to infer that a picture of a sentence or gesture expresses what that sentence or gesture expresses.

One complication in arguing for such a principle involves indexicality: a painting of an inscription that reads ‘I love you’, if indeed it expresses anything about love, does not necessarily express the painter’s love; likewise, a photo of Ada giving Stella the finger expresses not the photographer’s negative emotion, but (at best) Ada’s. Yet, as argued in 5.1, we need to account for the observation that a use of a face emoji, a picture of a facial expression, expresses the negative emotions of the current speaker. To get this right the transparency-theorist should then stipulate that face emojis depict the speaker, along with stipulating the relevant cross-medial transparency.

I prefer a slightly different route. I propose to extend Kaplan’s and Gutzmann’s distinction between descriptive words (with truth-conditional content) and expressive words (with use-conditional content), to the pictorial domain. While 🌍 depicts what the *world* looks like, 🗺️ depicts what the *context* looks like. More precisely, let’s capture the meaning of an ‘expressive picture’ like 😊 in use-conditional terms:

(39) use condition: a use of 😊 is felicitous in *c* iff *c* looks like this: 😊

Instead of saying that the picture is true of a world (and then letting pragmatic enrichment determine where, when, and how in the world things look that way), (39) defines when a picture is felicitously used in an utterance context. Saying that the context looks a certain way should be understood as saying that the world of the utterance context, seen from a canonical viewpoint associated with the utterance context, projects onto the given picture. I’ll assume that each utterance context determines a default, canonical viewpoint,  $v_c$ , which is the viewpoint that corresponds to someone looking straight at the current utterer.

(40) use-conditional content:  $\llbracket \text{😊} \rrbracket = \left\{ c \mid \Pi(w_c, v_c) = \text{😊} \right\}$

Paraphrasing (40): the use-conditional content of the face emoji 😊 is the set of utterance contexts in which the speaker looks like this: 😊.

To be sure, an actual tweeter doesn’t always actually wear such a big grin on her face while typing a 😊. What (40) says is that she conveys (in a use-conditional way) to her addressee that she has such a facial expression. In other words, by using 😊 the speaker presents herself as looking that way, but as with any form of linguistic communication that presentation may involve some pretense, exaggeration, or even insincerity or deception.<sup>35</sup>

35. A referee suggests this amounts to a kind of ‘error theory’ of emoji usage. As far as I can tell, the insincerity involved is no different from when we say ‘nice weather’ or ‘I’m good, how are you?’ without actually believing the weather is nice or we’re good.



We can now analyze descriptive text adorned with expressive face emojis as follows:

$$(41) \quad \llbracket \text{I'm coming over} \text{ 😊} \rrbracket = \frac{\llbracket \text{😊} \rrbracket}{\llbracket \text{I'm coming over} \rrbracket} =$$

$$= \frac{\left\{ c \mid \Pi(w_c, v_c) = \text{😊} \right\}}{\left\{ \langle c, w \rangle \mid sp_c \text{ is coming over in } w \right\}}$$

The fact that the picture depicts the context from its canonical viewpoint  $v_c$  now gives us the observed speaker orientation of face emojis. But we don't yet see any of the expected emotional content in (41).

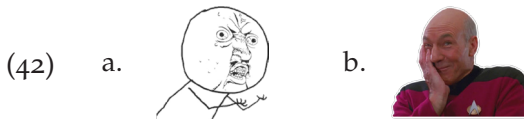
Instead of appealing to a general stipulation of cross-medial transparency, we can actually derive the emotional content of the emoji pragmatically by simply combining the use-conditional pictorial content in (41) and the use-conditional content of the smile. Let's consider, step by step, what the use of the emoji in (41) is communicating to the receiver of the text message. Assuming, in Gricean fashion, that the speaker is cooperative, their use of the picture in  $c$  must have been felicitous, which entails that the context, or more specifically the speaker, looks like 😊 (where 'looks like' is understood projectively, relative to Apple's stylized projection function). The speaker's looking like that smiley face plausibly entails that the speaker of the context was smiling.<sup>36</sup> Finally, by the use-conditional semantics of smiling (a smile is felicitous in  $c$  iff  $sp_c$  has a positive disposition towards  $ad_c$  in  $c$ , = (38)) we can infer that a (cooperative) speaker that is smiling has a positive disposition towards their addressee. By this chain of rational reasoning, we have effectively composed the use-conditional pictorial semantics of face emojis with that of facial expressions and thereby pragmatically derived exactly the kind of transparency we need.

The two-stage pictorial account of face emojis presented above, though consisting of a few more moving parts than an expressive symbol account like Grosz et al.'s (2023), retains the general benefits of a pictorial account already listed in Section 1. Let's briefly revisit these benefits, applied specifically to expressive (face and hand gesture) emojis.

---

36. One can perhaps think of far-fetched scenarios where the context, seen from the canonical speaker-directed viewpoint, looks like the smiley face emoji (e.g., the speaker might be wearing a smiley face mask) but is arguably not smiling, and not friendly or happy. In such cases the current account predicts that a smiley face emoji might be used felicitously, while the pure expressivist account would predict it would be infelicitous. This situation is reminiscent of the creative uses of object emojis to denote things unrelated to their supposed dictionary definition, discussed below and in Section 4.2.

First, the pictorial account treats face emojis as continuous with other pictures of expressive facial expressions, like in (42), likewise typically used to express the message author’s emotional state.<sup>37</sup>



Second, it naturally predicts the kind of creative usage of emojis like 😞, which can be used to indicate a host of psychologically unrelated but outwardly similar looking states, from physical exertion to helplessness, as illustrated in (1) (Section 1). Instead of a multiply ambiguous term, the pictorial account treats it simply as a picture of what the speaker looks like, leaving the exact nature of the underlying physical or mental state unspecified, only to be filled in pragmatically in context.

Third, on lexical expressive accounts like Grosz et al. (2021; 2023) the recent rise of gender-and skin-tone specific emoji variants is quite puzzling. Let’s look at a hand gesture emoji, 👍, since face emojis don’t have skin tone specific versions (yet). For the symbolist, 👍 literally—if use-conditionally or emotively—expresses that the speaker is giving approval (see Grosz et al. 2023 for a very detailed account along these lines). But then why would we want a 👍 and 👏 to evidently express the exact same thing? On the pictorial account the use of a skin tone matched emoji variant is only natural, given the pictorial use condition: 👍 is used felicitously by me now if my hand now projects onto 👍. My hand projects more straightforwardly (i.e., with less abstraction and stylization transformations) onto a skin tone matching version than onto a default yellow version, and it arguably doesn’t project at all to a clearly mismatched skin tone one. Exactly the features of the pictorial analysis that gave rise to the pictorial overdetermination challenge (Section 4.1) now explain the use of skin tone matching emojis.<sup>38</sup>

## 6. Conclusions

I have argued that emojis can be analyzed quite literally as ‘little pictures’. Not lexical expressives, typographic gestures, anaphoric event descriptions, or dia-

37. There may be interesting semantic differences to explore between the use of gifs or memes like (i) and face emojis—in future research.

38. I’m aware that the lack of support for skin tone specific face emojis in the current Unicode Standard could now be used as an argument against my pictorial analysis. It is worth keeping in mind that the emoji inventory is centrally controlled by a single committee and thus not quite like a developing language. I find it hard to explain, semantically, why we currently have skin tone specific hand gestures but not facial expressions (especially, since the two often occur together and are both used expressively). I leave this for future research.

grams, but pictures that, like photos or drawings, inform us what the world looks like. I thus proposed a formal semantics of emojis in terms of geometric projection, as used also to model the semantic interpretation of pictures and visual narratives (Abusch 2012; Greenberg 2013).

A lot of the communicative work that emojis do in computer mediated communication, for example, elaborating on eventualities described in the text or expressing speaker emotions, relies on various kinds of pragmatic inferences on the basis of the rather minimal semantic content provided by the geometric semantics. I've discussed pragmatic enrichment through the inference of coherence relations and their presupposed events, and through metaphor and metonymy. When it comes to face and hand emojis I've discussed how to pragmatically link a use-conditional picture semantics with a use-conditional semantics of facial expressions and other gestures.

On the account developed here, emojis and text can combine to form genuinely multimodal discourse. The text–picture integration analysis I've proposed here immediately extends to the use of other arguably pictorial elements commonly inserted in text messages, like emoticons or ASCII drawings, but also more obviously pictorial elements like stickers and animated gifs. The framework is also partly continuous with—and indeed inspired by—semantic accounts of multimodal text–image combinations in more static print media, like comics or instruction manuals. In Section 5 we went beyond those static types of multimodality by looking at the expressive usage of the subclass emojis depicting faces and hands, which is mainly useful in interactive communication like chat, text, or Twitter. Here my account incorporates insights from semantic accounts of expressivity in (spoken) language and gesture.

Many issues in the semantics and pragmatics of emojis remain wide open. On the semantic side I'd like to gain a better understanding of the large grey area between arguably pictorial emojis (🚗, 📺, 😊) and arguably symbolic emojis (♻️, ❤️), and about the integration of symbolic and pictorial elements inside a single emoji (👉). On the pragmatic side I'd like first and foremost to gain a better understanding of the different types of figurative interpretations that we ultimately have to appeal to to extend the use of emojis beyond the entities, people, or events they literally depict. Since these are big issues that have been deserving of formal philosophical and linguistic scrutiny already independently of emojis I will leave it at this for now.

## Acknowledgments

This work is supported by NWO Vidi grant 276-80-004 (FICTION). Thanks to Tatjana Scheffler, Dorit Abusch, Patrick Grosz, Sarah Zobel, Dolf Rami and online

audiences at Sinn und Bedeutung 25, the Bochum Language Colloquium, and the VICOM workshop at DGfS 2022 for discussion. Huge thanks to two anonymous journal referees that provided extensive and very constructive comments.

## References

- Abusch, Dorit (2012). Applying Discourse Semantics and Pragmatics to Co-reference in Picture Sequences. *Proceedings of Sinn und Bedeutung*, 17, 9–25.
- Abusch, Dorit (2020). Possible-Worlds Semantics for Pictures. In Daniel Gutzmann, Lisa Matthewson, Cécile Meier, Hotze Rullmann, and Thomas Ede Zimmerman (Eds.), *The Wiley Blackwell Companion to Semantics* (1–31). John Wiley & Sons.
- Abusch, Dorit and Mats Rooth (2017). The Formal Semantics of Free Perception in Pictorial Narratives. *Proceeding of the Amsterdam Colloquium*, 21, 85–95.
- Agerri, Rodrigo, John Barnden, Mark Lee, and Alan Wallington (2007). On the Formalization of Invariant Mappings for Metaphor Interpretation. *Proceedings of ACL*, 45, 109–112.
- Altshuler, Daniel and Julian Schlöder (2021). If Pictures Are Stative, What Does This Mean for Discourse Interpretation? *Proceedings of Sinn und Bedeutung*, 25, 19–36.
- Amaral, Patricia, Craige Roberts, and E. Allyn Smith (2007). Review of the Logic of Conventional Implicatures by Chris Potts. *Linguistics and Philosophy*, 30 (6), 707–49. <https://doi.org/10.1007/s10988-008-9025-2>
- Asher, Nicholas and Alex Lascarides (2003). *Logics of Conversation*. Cambridge University Press.
- Ayer, Alfred (1936). *Language, Truth and Logic*. Victor Gollancz.
- Barach, Eliza, Laurie Feldman, and Heather Sheridan (2021). Are Emojis Processed like Words?: Eye Movements Reveal the Time Course of Semantic Processing for Emojified Text. *Psychonomic Bulletin & Review*, 28(1), 978–91. <https://doi.org/10.3758/s13423-020-01864-y>
- Barnden, John (2010). Metaphor and Metonymy: Making Their Connections More Slippery. *Cognitive Linguistics*, 21(1), 1–34.
- Brasoveanu, Adrian and Jakub Dotlacil (2015). Incremental and Predictive Interpretation. *Semantics and Linguistic Theory (SALT)*, 25(1), 57–81.
- Charlow, Nate (2015). Prospects for an Expressivist Theory of Meaning. *Philosophers' Imprint*, 15(23), 1–43.
- Cohen, Jonathan and Andrew Kehler (2021). Conversational Eliciture. *Philosophers' Imprint*, 21(12), 1–26.
- Cohn, Neil (2013). Beyond Speech Balloons and Thought Bubbles: The Integration of Text and Image. *Semiotica*, 2013 (197), 35–63. <https://doi.org/10.1515/sem-2013-0079>
- Cohn, Neil, Jan Engelen, and Joost Schilperoord (2019). The Grammar of Emoji? Constraints on Communicative Pictorial Sequencing. *Cognitive Research: Principles and Implications*, 4, Article 33. <https://doi.org/10.1186/s41235-019-0177-0>
- Cumming, Samuel, Gabriel Greenberg, and Rory Kelly (2017). Conventions of Viewpoint Coherence in Film. *Philosophers' Imprint*, 17(1), 1–29.
- Danesi, Marcel (2016). *The Semiotics of Emoji: The Rise of Visual Language in the Age of the Internet*. Bloomsbury.

- Darwin, Charles (1872). *The Expression of the Emotions in Man and Animals*. John Murray.
- Davidson, Donald (1967). The Logical Form of Action Sentences. In Nicholas Rescher (Ed.), *The Logic of Decision and Action* (81–95). University of Pittsburgh Press.
- Ebert, Cornelia. The Non-at-Issue Contributions of Gestures and Speculations about Their Origin. Slides. Retrieved from author's homepage <http://www.cow-electric.com/neli/talks/CE-demonstration-stuttgart.pdf>
- Ekman, Paul, Richard Davidson, and Wallace Friesen (1990). The Duchenne Smile: Emotional Expression and Brain Physiology: II. *Journal of Personality and Social Psychology*, 58(2), 342–53. <https://doi.org/10.1037/0022-3514.58.2.342>
- Fillmore, Charles (2008). Frame Semantics. In Dirk Geeraerts (Ed.), *Cognitive Linguistics: Basic Readings* (373–400). De Gruyter Mouton.
- Gawne, Lauren and Gretchen McCulloch (2019). Emoji as Digital Gestures. *Language@Internet*, 17(2). <https://www.languageatinternet.org/articles/2019/gawne>
- Geurts, Bart, David Beaver, and Emar Maier (2020). Discourse Representation Theory. In Edward Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/archives/spr2020/entries/discourse-representation-theory/>
- Giardino, Valeria and Gabriel Greenberg (2015). Introduction: Varieties of Iconicity. *Review of Philosophy and Psychology*, 6(1), 1–25. <https://doi.org/10.1007/s13164-014-0210-7>
- Ginzburg, Jonathan, Chiara Mazzocconi, and Ye Tian (2020). Laughter as Language. *Glossa: A Journal of General Linguistics*, 5(1), 1–51. <https://doi.org/10.5334/gjgl.11152>
- Goodman, Nelson (1976). *Languages of Art: An Approach to a Theory of Symbols*. Hackett Publishing.
- Greenberg, Gabriel (2013). Beyond Resemblance. *Philosophical Review*, 122(2), 215–87. <https://doi.org/10.1215/00318108-1963716>
- Greenberg, Gabriel (2019). Tagging: Semantics at the Iconic/Symbolic Interface. *Proceedings of the Amsterdam Colloquium*, 22, 11–20.
- Greenberg, Gabriel (2021). Semantics of Pictorial Space. *Review of Philosophy and Psychology*, 12(1), 847–87. <https://doi.org/10.1007/s13164-020-00513-6>
- Grosz, Patrick, Elsi Kaiser, and Francesco Pierini (2021). Discourse Anaphoricity and First-Person Indexicality in Emoji Resolution. *Proceedings of Sinn und Bedeutung*, 25(1), 340–57.
- Grosz, Patrick, Gabriel Greenberg, Christian De Leon, and Elsi Kaiser (2023). A Semantics of Face Emoji in Discourse. *Linguistics & Philosophy*, 46, 905–957. <https://doi.org/10.1007/s10988-022-09369-8>
- Gutzmann, Daniel (2015). *Use-Conditional Meaning: Studies in Multidimensional Semantics*. Oxford University Press.
- Hagen, Margaret (1986). *Varieties of Realism: Geometries of Representational Art*. Cambridge University Press.
- Harris, Jesse and Christopher Potts (2010). Perspective-Shifting with Appositives and Expressives. *Linguistics and Philosophy*, 32(6), 523–52. <https://doi.org/10.1007/s10988-010-9070-5>
- Hobbs, Jerry (1979). Coherence and Coreference. *Cognitive Science*, 3(1), 67–90. [https://doi.org/10.1207/s15516709cog0301\\_4](https://doi.org/10.1207/s15516709cog0301_4)
- Hobbs, Jerry (1990). *Literature and Cognition*. CSLI.
- Kaiser, Elsi and Patrick Grosz (2021). Anaphoricity in Emoji: An Experimental Investigation of Face and Non-Face Emoji. *Proceedings of the Linguistic Society of America*, 6(1), 1009–23. <https://doi.org/10.3765/plsa.v6i1.5067>

- Kamp, Hans (1981). A Theory of Truth and Semantic Representation. In Jeroen Groenendijk, Theo Janssen, and Martin Stokhof (Eds.), *Formal Methods in the Study of Language* (277–322). Mathematical Centre Tracts.
- Kaplan, David (1989). Demonstratives. In Joseph Almog, John Perry, and Howard Wettstein (Eds.), *Themes from Kaplan* (481–614). Oxford University Press.
- Kaplan, David (1999). The Meaning of ‘Ouch’ and ‘Oops’: Explorations in the Theory of Meaning as Use. Unpublished manuscript. <http://eecoppock.info/Pragmatics-SoSe2012/kaplan.pdf>
- Kehler, Andrew (2002). *Coherence, Reference, and the Theory of Grammar*. University Of Chicago Press.
- King, Alex (2018). A Plea for Emoji. *American Society for Aesthetics Newsletter*, 38(3), 1–3.
- Kulvicki, John (2003). Image Structure. *The Journal of Aesthetics and Art Criticism*, 61(4), 323–40.
- Kulvicki, John (2006). Pictorial Representation. *Philosophy Compass*, 1(6), 535–46.
- Kulvicki, John (2013). *Images*. Routledge.
- Lakoff, George (1993). The Contemporary Theory of Metaphor. In A. Ortony (Ed.), *Metaphor and Thought* (202–51). Cambridge University Press.
- Lakoff, George and Mark Johnson (2003). *Metaphors We Live By*. University of Chicago Press.
- Lascarides, Alex and Matthew Stone (2009). A Formal Semantic Analysis of Gesture. *Journal of Semantics*, 26(4), 393–449. <https://doi.org/10.1093/jos/ffp004>
- Maier, Emar (2019). Picturing Words: The Semantics of Speech Balloons. *Proceedings of the Amsterdam Colloquium*, 22, 584–92.
- Maier, Emar and Sofia Bimpikou (2019). Shifting Perspectives in Pictorial Narratives. *Sinn und Bedeutung*, 23(2), 91–106. <https://doi.org/10.18148/sub/2019.v23i2.600>
- Mann, William and Sandra Thompson (1988). Rhetorical Structure Theory: Toward a Functional Theory of Text Organization. *Text*, 8(3), 243–81.
- McCulloch, Gretchen and Lauren Gawne (2018). Emoji Grammar as Beat Gestures. *Proceedings of the International Workshop on Emoji Understanding and Applications in Social Media* 1(1), 1–4.
- Nouwen, Rick, Adrian Brasoveanu, Jan van Eijck, and Albert Visser (2022). Dynamic Semantics. In Edward N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/archives/fall2022/entries/dynamic-semantics/>
- Pagin, Peter (2014). Pragmatic Enrichment as Coherence Raising. *Philosophical Studies*, 168(1), 59–100. <https://doi.org/10.1007/s11098-013-0221-8>
- Pasternak, Robert and Lyn Tieu (2022). Co-Linguistic Content Inferences: From Gestures to Sound Effects and Emoji. *Quarterly Journal of Experimental Psychology*, 75(10), 1828–43. <https://doi.org/10.1177/17470218221080645>
- Peirce, Charles S. (1868). On a New List of Categories. *Proceedings of the American Academy of Arts and Sciences*, 7(1), 287–98.
- Pierini, Francesco (2021). Emojis and Gestures: A New Typology. *Proceedings of Sinn und Bedeutung*, 25, 720–32. <https://doi.org/10.18148/sub/2021.v25i0.963>.
- Potts, Christopher (2007). The Expressive Dimension. *Theoretical Linguistics*, 33(2), 165–98. <https://doi.org/10.1515/TL.2007.011>
- Predelli, Stefano (2013). *Meaning without Truth*. Oxford University Press.
- Recanati, Francois (2010). Pragmatic Enrichment. In Delia Fara and Gillian Russell (Eds.), *Routledge Companion to Philosophy of Language* (67–78). Routledge.

- Rooth, Mats and Dorit Abusch (2017). Picture Descriptions and Centered Content. *Proceedings of Sinn und Bedeutung*, 21(2), 1051–64.
- Rooth, Mats and Dorit Abusch (2019). Indexing across Media. *Proceedings of the Amsterdam Colloquium*, 22, 612–24.
- Scheffler, Tatjana, Lasse Brandt, Marie de la Fuente, and Ivan Nenchev (2022). The Processing of Emoji-Word Substitutions: A Self-Paced-Reading Study. *Computers in Human Behavior*, 127(1), 1–11. <https://doi.org/10.1016/j.chb.2021.107076>
- Schlenker, Philippe (2018). Visible Meaning: Sign Language and the Foundations of Semantics. *Theoretical Linguistics*, 44(3–4), 123–208. <https://doi.org/10.1515/tl-2018-0012>
- Shin, Sun-Joo, Oliver Lemon, and John Mumma (2018). Diagrams. In Edward N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/archives/win2018/entries/diagrams/>.
- Stevenson, C. L. (1944). *Ethics and Language*. Yale University Press.
- Tang, Mengmeng, Bingfei Chen, Xiufeng Zhao, and Lun Zhao (2020). Processing Network Emojis in Chinese Sentence Context: An ERP Study. *Neuroscience Letters*, 722(1), 134815. <https://doi.org/10.1016/j.neulet.2020.134815>
- Viebahn, Emanuel (2019). Lying with Pictures. *The British Journal of Aesthetics*, 59(3), 243–57. <https://doi.org/10.1093/aesthj/ayz008>
- Wildfeuer, Janina (2014). *Film Discourse Interpretation: Towards a New Paradigm for Multimodal Film Analysis*. Routledge.
- Wildfeuer, Janina (2019). The Inferential Semantics of Comics Panels and Their Meanings. *Poetics Today*, 40(2), 215–34. <https://doi.org/10.1215/03335372-7298522>
- Wittgenstein, Ludwig (1958). *Philosophical Investigations*. Basil Blackwell.
- Wittgenstein, Ludwig (1966). *Lectures and Conversations on Aesthetics, Psychology and Religious Belief*. Cyril Barrett (Ed.). Basil Blackwell.