Emojis as pictures[*]

Emar Maier
University of Groningen

Abstract I argue that emojis are pictures, not a species of words, gestures, or expressives. 🎁 means that the world looks like that, from some viewpoint. I formalize this in terms of geometric projection with stylization. Since such a pictorial semantics delivers only very minimal contents I add an account of pragmatic enrichment, driven by coherence and metaphor. 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. Combined with the idea that the speaker's facial expressions thus depicted by face emojis are themselves expressive, I derive the intuition that face emojis can be used to express emotional states, while maintaining that they are pictures.

Keywords: emojis, pictorial semantics, symbolic vs iconic, geometric projection, pragmatic enrichment, coherence, metaphor, expressives, facial expressions, use-conditional content

Contents

1 Introduction

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

> 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)

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,[2] 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), this could be taken as suggesting the view that such face-drawings, like gestures and facial expressions, are *expressives*, i.e. 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); Pierini (2021);

---

1 E.g. https://qz.com/1261293/ludwig-wittgenstein-was-the-great-philosopher-of-the-20th-century-he-also-invented-the-emoji/
2 Close reading suggests a 'rebus' account, where emojis stand for words, rather than their meanings. We'll disregard this reading.

Pasternak & Tieu (2021), or that face emojis in particular are expressives (Grosz et al. 2021b,a). There are now also empirical studies purporting to show that emojis more generally are like words added to the lexicon (Tang et al. 2020; Scheffler et al. 2021). Interestingly these linguistic views tend to skip over the fact that emojis, like Wittgenstein's face sketches, are actually pictures, and as such should have an iconic semantics.[3]

I propose and defend a thoroughly pictorial semantics for a large class of emojis. In short, this paper argues that emojis are not new words, typed gestures, or expressives, but simply little pictures. That is, like drawings and photographs, they semantically convey information about 'what the world looks like'.

There are limitations to this 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, some very common emojis are clearly not pictures in the strict sense implied above (and to be made formally precise in terms of geometric projection below). ♻ is the fifth most common emoji on Twitter, according to emojitracker.com. It is typically used as a *symbol* denoting recycling or, more commonly, retweeting, not a *picture* of either of these activities – just like the word 'Belgium' or the Belgian flag refer to Belgium symbolically rather than pictorially. Simplistically put, a symbol doesn't resemble or depict its denotation but refers to it by convention or stipulation (Giardino & Greenberg 2015).

Many emojis in the Emojipedia[4] category 'Symbols' are indeed symbols in the above sense, though some fall in the grey area between picture and symbol – is the heart emoji just a conventional symbol of love, or is it perhaps in some sense a stylized (section 2.2) picture of a human heart, which by metaphoric extension (section 3.5) is associated with certain positive emotions? I will have nothing interesting to say here about symbolic emojis, nor propose a clear boundary between pictorial and symbolic emojis.

Some other emojis are arguably pictorial, but include some highly symbolic elements. For instance, 😪 depicts a sleepy face, with a giant snot bubble coming from the left nostril – a convention borrowed from 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, e.g., as described for speech balloons and other picture–symbol mixtures in comics by Maier (2019). I will not discuss this matter further here.

Finally, I'll be restricting attention to the use of emojis as separate discourse units, i.e. inserted after a sentence or as a free-standing discourse move, disregarding

---

3 I'm of course not claiming originality for the claim that emojis are pictures. Danesi (2016) and Cohn et al. (2019), for instance, seem to assume a pictorial account of emojis, but they don't explicitly argue for this position or make precise what it means to say that something is a picture.

4 https://emojipedia.org

'pro-speech emojis' (Pierini 2021), i.e. emojis that are syntactically integrated into a sentence and 'replace' a specific word or concept, like 'love' and 'present' in (1-a), 'happy' in (1-b), and sometimes even a specific (English) word sound or shape, in rebus-like fashion, like in (1-c).

(1)   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. 2021)

I will also not discuss emoji sequences and the interesting potential for syntactic and semantic composition therein Cohn et al. (2019); McCulloch & Gawne (2018).

In sum, I qualify my claim by accepting that some emojis may be used to replace specific words within a sentence, and moreover some emojis are inherently symbolic, and some are in the grey area between symbol and picture, and/or combine elements of both. What remains is 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 a strict, geometric sense to be explicated below.

Within that large class of pictorial emojis, I'm assuming with Grosz et al. (2021b) a semantic distinction between those 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 (or rather, what they present themselves as looking 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 expressive gestures and that pictorial representation in this case is somewhat transparent.

I will make my proposal precise by combining (i) a semantic account of pictorial content in terms of geometric projection (section 2), (ii) an account of pragmatic enrichment via the inference of coherence relations and metaphorical interpretation (section 3), and (iii) an extension of the Kaplanian account of use-conditional meaning components for pictures, gestures, and their composition (i.e., pictures of gestures) (section 4).

2 Pictorial semantics

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 could say that a picture is true or false with respect to a certain 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 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: the picture is true of a world iff it resembles part of that world. On reflection, resemblance turns 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 of resemblance, also in linguistic semantics (Abusch 2020).

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, Π, mapping a possible world w and a viewpoint v (a vector located at a certain spatiotemporal location, intuitively representing the gaze direction of someone located somewhere in the world) onto a picture p: $\Pi(w,v) = p$.

There are many different such recipes that all qualify as geometric projection functions. One well-known example Π 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 line drawing.

(2)
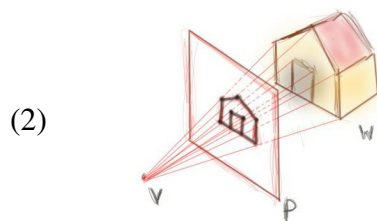
(=diagram illustrating linear perspective b/w line drawing, to be drawn)

More complicated projection functions might include rules for representing color, shadow, and even some additional distortion, abstraction, stylization transformations to create depictions that deviate from photorealistic projection in different ways.

5

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

When we know how to turn some part of the world into a picture, given a Π and a v, we can say when a picture is true:

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

Here we're assuming the projection function Π to be fixed, i.e., provided by the context, just as in the linguistic domain we assume the language to be given presemantically, i.e. before computing the truth value of an utterance. Projection is the pictorial analogue of language (Greenberg 2013; Giardino & Greenberg 2015).

From the truth definition in (4) we can define various notions or levels of pictorial content. A natural analogue of classic propositional content results from existential closure over the viewpoint: given a pictorial language Π, a picture expresses the proposition that there is a viewpoint from where the world projects onto that picture:

(5) $$\left[\!\!\left[\; \text{} \;\right]\!\!\right]^{\Pi} = \left\{ w \;\middle|\; \exists v.\Pi(w,v) = \text{} \right\}$$

Alternatively, for different purposes we may need other notions of content, e.g. analogues of centered/diagonal propositions (set of world–viewpoint pairs, Rooth & Abusch 2017) or horizontal contents (sets of worlds, assuming a fixed viewpoint). Below we'll introduce a dynamic notion of pictorial content, in terms of information states.

It may be possible and 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.[5] To avoid potential difficulties delimiting and quantifying over the space of possible pictorial (or lin-

---

5 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. Although metalinguistic diagonalization may be useful (e.g. in semantic accounts of mixed quotation Shan 2011; Wiślicki 2021), it comes at a high price (Maier 2016).

guistic) languages, I will refrain from this route here and simply assume that a Π is presemantically given. How an interpreter arrives at this Π is then a matter of contextual pragmatic inference that we will only talk about informally.

2.2 Stylization

If emojis are pictures, they are not very 'realistic', but rather 'abstract' or 'stylized'. In the projection framework the differences between a line drawing and a full color photograph can be thought of as corresponding to different parameter settings in 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. Various qualitatively distinct objects in various possible worlds could thus give rise to the same 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. Slightly crooked edges and invisible imperfections might be ignored and represented by perfectly straight lines on the picture plane. We could also decide to approximate any shape projected on the plane with the closest simple polygon (with less than 10 sides, say).[6] We'll consider such approximative geometric transformations part of the projection function (Abusch 2012; Greenberg 2021). With such a stylized projection function then, a simple cube drawing like ⌗ would be true of not just different geometric worlds where we're looking at actual floating Platonic cubes (of various shapes, colors, and sizes), but also of worlds like ours where we're looking at shapes that are roughly cube-like, like a sugar cube, or 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 transformations on top of a 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.

(6)

      Apple   OpenMoji   HTC
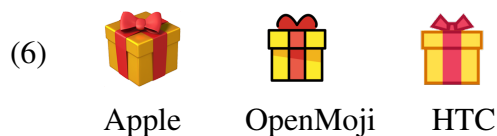
Apple provides one of the more photorealistic instantiations, and HTC one of the most stylized. In our projective semantics terminology, we'd say that Apple's

---

6 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 Abusch (2020).

projection function, Π_{Apple}, seems to involve a standard linear perspective (the edges in the back are slightly shorter than the parallel ones in front, which are moreover not quite parallel, suggesting a vanishing point); 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 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 is chosen.[7] HTC, finally, uses the same canonical viewpoint and builds similar color, texture, and symmetry abstractions as OpenMoji into its projection function, though with a slightly different edge processing, and ignoring all shadows.

Note that these informal inferences about the nature of the three Π'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 using 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. Also, all three pictures might in principle be parallel projections of flat objects, or even photorealistic linear projections of more or less abstract drawings of objects.

As discussed also at the end of section 2.1 we'll ignore this general projection underdetermination problem and assume that context, common sense and experience with pictures of common objects pragmatically (or in any case, pre-semantically) narrow down the space of possible parameter settings to something like the assumptions laid out above. On those assumption we might say that Apple's emoji gives the most information about color, and about the different sides of the object, and hence would exclude some worlds where we're looking at, say, an oblong box with a rough, ragged, unevenly colored cardboard texture where the bow is located way in the back, which would be compatible with the more stylized OpenMoji and HTC emoji.

This is pretty much all I have to say about the semantics of object emojis, treated as pictures. Note that this purely projective semantics is probably the most minimal kind meaning we can endow emojis with. The wrapped gift emoji does not denote the concept of gifts nor the activity of gift giving, it just depicts the world as containing at some point in space and time an object and a nearby viewpoint from where that object 'looks like' (i.e., 'projects onto') this: 🎁 .

---

7 Given some familiarity with common early 21st century photography, drawing, and computer graphics styles we can assume it to be either parallel or linear perspective.

Most of the interesting meanings that these emojis actually communicate in the context of a tweet or other text message come about on the basis of pragmatic enrichment of this minimal content. As will become clear, the relative abstractness of HTC vs Apple cashed out at the level of our minimal semantics above will be overshadowed by these enrichments, making the various emoji sets roughly equally expressive in actual communicative practice.

3 Pragmatic enrichment

The semantics I have proposed for emojis is incredibly minimal:

(7) $\left[\!\left[\, 🎁 \,\right]\!\right]^{\Pi_{Apple}} = \left\{ w \,\middle|\, \exists v \,.\, \Pi_{Apple}(w,v) = 🎁 \right\} \approx$ the proposition that there's a viewpoint somewhere in space and time from where the world looks like this: 🎁 .

As we've seen, already some defeasible pragmatic reasoning about the underlying Π_{Apple} is required to get even this much. A lot more pragmatics is needed to turn this basic pictorial content into something worth adding to an actual tweet or text. I follow Grosz et al. (2021b) who appeal 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 proper.

3.1 Coherence in verbal and visual language

Hobbs (1979) 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 (8):

(8)   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 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 involved we can infer different relations between them. While in (8) we inferred a relation commonly known as Explanation, in (9) we'll likely infer a different one called Result.

(9)   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 coherence relations Mann & Thompson (1988); Kehler (2002). In

all 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) 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 our formal semantic machinery, is called Segmented Discourse Representation Theory (SDRT, Asher & 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 DRT (Kamp 1981). The relata are elementary discourse units, typically sentences or clauses that can be said to describe eventualities or express propositions. Using special propositional discourse referents (π_1, π_2, ...) to label these elementary units and using DRT to represent their semantic content, we get familiar SDRS representations like (10)

(10)
$$\pi_1: \boxed{\begin{array}{c} \underline{e_1 \quad x} \\ \mathrm{miss}(e_1) \\ \mathrm{agent}(e_1,i) \\ \mathrm{meeting}(x) \\ \mathrm{theme}(e_1,x) \end{array}} \quad \pi_2: \boxed{\begin{array}{c} \underline{e_2} \\ \mathrm{fired}(e_2) \\ \mathrm{theme}(e_2,i) \end{array}}$$
$$\mathrm{Result}(\pi_1,\pi_2)$$
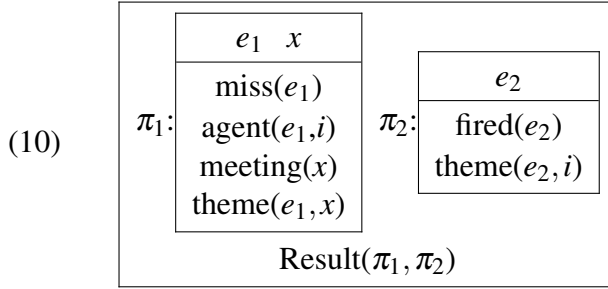
The interpretation of coherence conditions can be formalized as an extension of the semantics of DRT. For instance, Narration holds between two units if the information carried by both units is true (or, more properly dynamically: both units update the common ground) and the main event contributed by the first (e_{π_1}) immediately precedes (or 'occasions', $\prec$) the main event contributed by the second. Note that this semantics presupposes that both units introduce a main event. Notation: $\oplus$ stands for DRS merge, the DRT way of dynamic updating at the representational level of DRSs.

(11) $[\![\mathrm{Narration}(\pi_1,\pi_2)]\!] = [\![\pi_1 \oplus \pi_2 \oplus \boxed{e_{\pi_1} \prec e_{\pi_2}}]\!]$

The introduction of coherence relations in the discourse interpretation process 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 (formally, axioms in a so-called Glue Logic Asher & Lascarides 2003). For instance, if one unit π_1 introduces a state and a subsequent, structurally accessible unit π_2 introduces and event, then all else being equal we can

add 'Background(π_1, π_2)' to the discourse structure. We'll skip over all details of context change composition, accessibility, complex discourse units, and Glue Logic axioms.

We're interested in the application of coherence theory, and SDRT in particular, as a way to model pragmatic enrichment with partly or wholly pictorial discourses. First let's rephrase Maier & Bimpikou's (2019) DRT style analyses of purely pictorial narratives like (12) into the SDRT framework, by viewing the panels as elementary discourse units.[8]

(12)
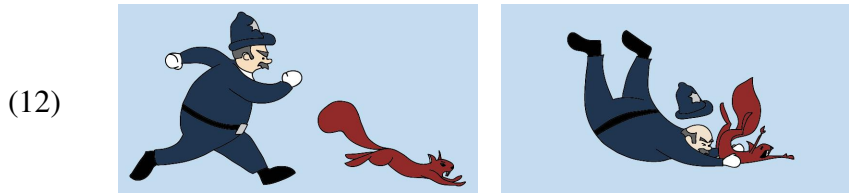
The basic assumption behind Maier & Bimpikou's (2019) PicDRT is that pictures are like elementary discourse units, i.e., they express information about what the world looks like. To turn a sequence of such propositional units into a coherent narrative would mean that we add coherence relations. In this case, and in many panel transitions in many comics, the relation must be Narration: the policeman is chasing a squirrel *and then* he catches it.
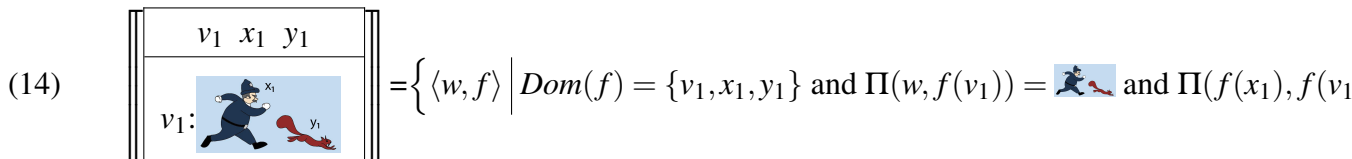
(13)

π_1 : π_2 :

Narration(π_1, π_2)
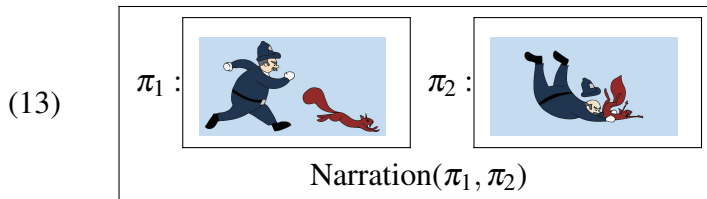
Zooming in on the pictorial DRS components in (13), Maier & Bimpikou (2019), inspired by Abusch (2012), add a presemantic level of interpretation where pictures are labeled with viewpoint referents (v_1, v_2), and salient regions of interest in the picture are labeled with individual discourse referents (x_1, y_2). A preliminary DRS representation of the first panel, with 2 salient regions introducing discourse referents, would be modeltheoretically interpreted as in (14).

(14)

$$\boxed{\begin{array}{c} v_1 \;\; x_1 \;\; y_1 \\ v_1 : \end{array}} = \left\{ \langle w, f \rangle \;\middle|\; Dom(f) = \{v_1, x_1, y_1\} \text{ and } \Pi(w, f(v_1)) = \text{🏃🐿} \text{ and } \Pi(f(x_1), f(v_1 \right.$$

_____

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

Paraphrasing informally, the DRS in (14) contributes the information (formally, an 'information state', defined as a set of world–assignment pairs) that (i) there is a certain viewpoint from which the world looks like the whole picture (i.e., $\Pi(w, f(v_1)) =$ 🏃🐿️); (ii) there's an individual that, when projected from that same viewpoint, looks like the bluish region (i.e., $\Pi(f(x_1), f(v_1)) =$ 🏃) ; and (iii) there's another individual that looks like the smaller brownish region (i.e., $\Pi(f(y_1), f(v_1)) =$ 🐿️).

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)'), and different discourse referents from different pictures can be equated (e.g., $x_2 = x_1$ – a pragmatic pictorial analogue of anaphora resolution, Abusch 2012). In addition to the discourse unit labels (π_1, π_2) and coherence relations ('Narration(π_1, π_2)') sketched in (14), we now further add to the post-semantic enrichment stage the introduction of eventuality discourse referents $(e_1, e_2 \dots)$. Note that this last enrichment is crucially driven by the semantics of Narration, which, as defined in (11), presupposes that both units introduce an event discourse referent.

(15)

v_1 x_1 y_1 e_1	v_2 x_2 y_2 e_2
π_1: v_1: [picture]	π_2: v_2: [picture]
policeman(x_1) squirrel(y_1) chase(e_1) agent(e_1, x_1) theme(e_1, y_1)	$x_2 = x_1$ $y_2 = y_1$ catch(e_2) agent(e_2, x_2) theme(e_2, y_2)
Narration(π_1, π_2)	
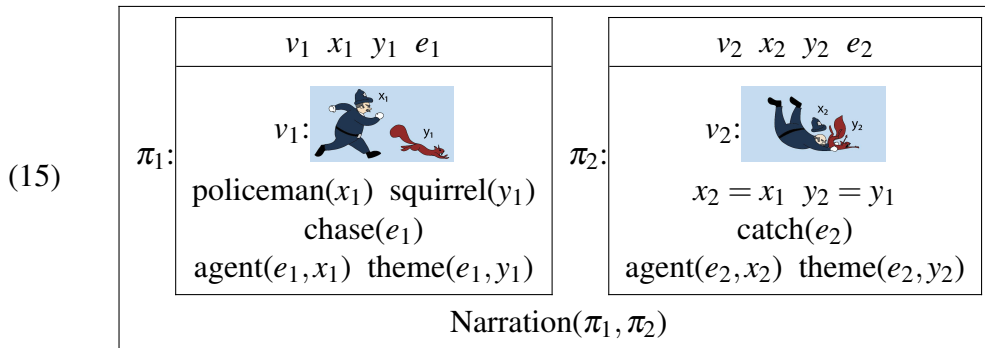
The modeltheoretic interpretation of (15) is a straightforward extension of (14) involving only standard DRT and SDRT semantics.

In sum: the semantics proper of a single picture is very minimal, viz. the world looks like this at some point in space and time. When presented with a few pictures in a seemingly deliberate sequence we want to go beyond mere conjunction of those minimal propositions (the world looks like this at some point and also like that at some point), just like we do when presented with a series of utterances (see Hobbs' 1979 discussion of 'John took a train from Paris to Istanbul. He likes spinach'). The sequencing triggers a cognitive enrichment process that crucially involves the inference of various coherence relations in order to satisfy a global desire for maximal coherence.

Since the coherence-driven enrichment mechanism 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 (Maier 2019; Wildfeuer 2019), illustrations with captions (Rooth & Abusch 2019) or tags (Greenberg 2019), or, say, illustrated children's books or instruction manuals. 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

To recap: the minimal, pictorial content we've assigned to object emojis like 🎁 can be roughly paraphrased as the proposition (or more precisely, a dynamic information state) that there is a viewpoint near where there is some object that looks like that.

Note first that this semantic object is a proposition, or, if we take the DRT approach sketched in the previous section, the dynamic equivalent of a proposition, an information state. Arguably, this is fundamentally the wrong semantic type for analyzing 'pro-speech 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.

Following Lascarides & Stone's (2009) original analysis of speech–gesture integration, but more directly following Grosz et al.'s (2021b) analysis of activity emojis, these propositional emoji uses can be analyzed as discourse units in their own right, alongside the textual units.

(16) π_1:Happy Birthday! π_2:I'm coming over this afternoon π_3:🎁

Maximizing coherence means inferring coherence relations between these discourse units. π_2 contributes the existence of an event of the speaker coming over in the afternoon of the utterance day, while π_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 Explanation:

(17)
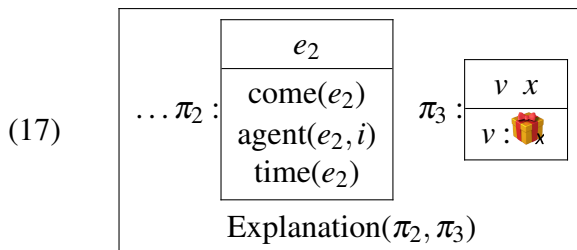
The semantics of Explanation, like Narration, (11), demands two events, and requires that the second unit's main event causes the first's. Thus, the inference of an

Explanation (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?[9]

At this point I defer to something like general cognitive schemas, scripts, or frames: things that look like that, i.e., nicely wrapped boxes, typically serve as 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 (12) is (probably) a police officer and that the inferred position of his arms legs are (probably) snapshots of him chasing something (rather then assuming a weird pose and floating in the air).

With all the enrichments above, the coherent interpretation of the tweet in (16) now looks something like this:

(18)
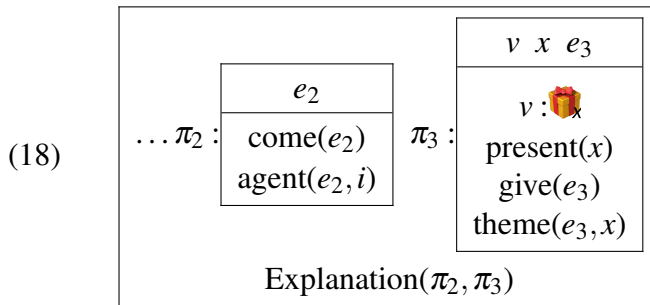
Paraphrasing the interpretation of the SDRS in (18): there's an event of the speaker coming over and an event of giving a present that looks like this 🎁 and the latter event explains the former.

But does the present really have to look just like that? Of course, we're assuming that Apple's projection function includes some abstraction leading to some indeterminacy about the actual size, shape, color, lighting of the present depicted, but what if it's a blue box with a yellow bow, or a ball wrapped in newspaper without a bow, or just an electrocnic gift certificate? As it stands, the pictorial account would make the speaker a liar if she meant to give such a gift, which is obviously absurd. Any gift will do, it doesn't have to look like this 🎁 .

---

9 Altshuler & 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. I disagree, at least if we look at the total pragmasemantic contribution of the picture in context, i.e. after the various enrichments, as opposed to the purely semantic, pictorial content, which is, arguably, either stative or devoid of eventualities.

3.3 The pictorial overdetermination challenge

To drive the point home and illustrate the framework once more, consider another example tweet:

(19)   I'm coming over this afternoon 🚗

In this variation of (16) the emoji likewise 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.[10]

(20)

	v_2 x_2 e_2
e_1	v_2: 🚗^{x}
$\dots \pi_1$: come(e_1) π_2 :	car(x_2)
agent(e_1, i)	drive(e_2)
	theme(e_2, x_2)
Elaboration(π_1, π_2)	
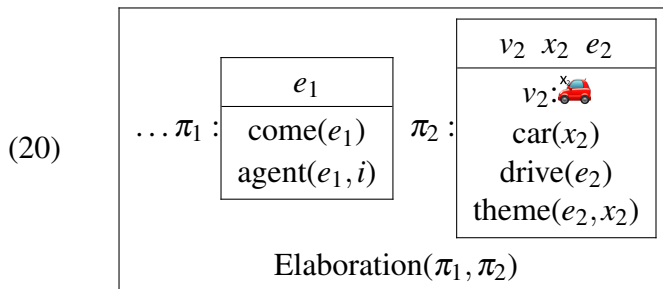
In words: there's an event of the speaker coming over and that event includes the event of driving a car that looks like this. The phrase 'looks like' as always has to be understood in terms of projection. Since Apple's emojis tend to have various different colors that in many ways seem to reflect the actual colors of the depicted objects (in canonical lighting conditions) with some degree of faithfulness (e.g. think of the emojis for entities like footballs, trees, sheep, but also people and hand emojis with various skin tones, see section 4.4), we might reasonably assume that Apple's projection function approximates actual color (within a finite set of fixed color codes, presumably). But then, counterintuitively, (19) would be false if the speaker drives a silver car.[11]

A first attempt at addressing the 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 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 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

---

10 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.

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

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 the parallel problem of shape. What if the 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 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 a pictorial/iconic account.

Before we explore ways to solve the overdetermination challenge without going symbolic, let's briefly explore this alternative symbolic approach and its limits.

3.4 Symbol vs picture

As alluded to in the introduction, the standard view of emoji semantics treats emojis as essentially an extension of the lexicon of a certain genre of written language. In Grosz et al. (2021b), 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 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, i.e. symbolic, lexical entry, rather than a strictly pictorial semantics like the one I'm defending here.[12]

Interestingly, the symbolic approach runs into a different kind of overdetermination problem. While it avoids overdetermining what the object looks like, it overdetermines the conceptual classification. In some cases, 🎁 really *does* denote something that looks like that without being a present. For instance, in the tweet in (21) 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.

---

12 See for instance Abusch's (2012) for the claim that there is no real anaphoricity in pictures. Purely semantically, in the police/squirrel comic, 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*), nothing in the second panel itself tells us to look for an antecedent to bind the new discourse referents to.

(21)   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 fixes the meaning symbolically using the concept of a
gift,[13] would predict that (21) would be infelicitous if the author intends to include
people simply ordering their stuff for themselves rather than as a gift.

   In this particular case, the symbolist might object that the use in (21) *is* infelici-
tous – the author should have chosen 📦 to illustrate the general concept of parcels
or delivery. 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).

(22)   Thanks to anonymous for chatting to me about reaching a global audience
         online during lockdown with his stunning cello performances 🎻

In light of (22) the symbolist might weaken their lexical entry to accommodate cellos
(e.g. to 'bowed classical instrument'), but there may always be new, unforeseen use
cases that don't quite fit any proposed lexical definition (e.g. *bring your ukulele* 🎻).

   A more dramatic illustration of the flexibility of emoji meaning involves the
well-known use of e.g. 🍆 and 🍑 to denote somewhat taboo body parts and
events involving them. On a symbolic account, do we really want to give the lexical
semantics of 🍆 as including both eggplants and penises? Probably a more intuitive
approach would have eggplants in the literal meaning and derive the other use as a
secondary,[14] non-literal meaning.

3.5   The eggplant metaphor

Lakoff (1993) uses the term 'image metaphor' to describe a metaphorical interpre-
tation based on visual resemblance between the literal meaning ('source') and the
metaphorical interpretation ('target'). He illustrates the pheomenon with linguistic
examples like (23):

(23)   a.   My wife... whose waist is an hourglass. (André Breton, cited by

---

13 Extrapolating Grosz et al.'s remarks, they would have a lexical entry of the form:
   $\llbracket 🎁 \rrbracket = \lambda e \exists y.(\text{present}(y) \wedge \text{agent}(e,x) \wedge \text{theme}(e,y))$.
14 Of course, in this case 'secondary' does not mean less frequent. Also, like metaphor and metonymy,
   non-literal 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.

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 (23-a) is not in any way conceptually related to an hourglass – it doesn't help keep time, for instance – it just looks like one, vaguely.

   Examples like (23) show the need for an account of image metaphor, wholly independently of emoji or pictorial semantics. Now if we'd assume such an independent account, proponents of both the pictorial and the symbolic account could appeal to it to explain the common interpretation of the eggplant emoji. But, since the symbolist makes the eggplant emoji roughly synonymous with '(there's an) eggplant' we might expect the eggplant–penis metaphor to occur linguistically – and occasionally it does:

(24)   The Warri pikin took to his IG account this morning to flaunt his eggplant in wet white underwear.[15]

Crucially, though examples like (24) are not really hard to google, they are not that common either – especially comparing the proportion of literal vs. figurative uses of the word with that of the emoji. In fact, it's quite possible that most if not all the linguistic instantiations of this particular image metaphor are entirely derivative on the widespread emoji version.

   On my pictorial account of 🍆 , these observations are not surprising at all: the semantics of the emoji literally tells us what the world looks like, and hence, at the level of pragmatic enrichment, is readily extended with further image-based inferences. In cognitive processing terms, the pictorial semantics would predict engagement of the visual system in semantic processing, which we'd expect to prime image-metaphoric pragmatic processing.[16]

   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, i.e., facing the pictorial overdetermination challenge of section 3.3: no geometric projection function will map both a small red Honda coupe and a big black BMW sedan unto this picture 🚗 (unless it's essentially concept-based and hence not really pictorial).

---

15 gossipnaija.ng/2019/12/tuoyo-is-at-again-as-he-flaunts-his-eggplant-in-wet-white-speedo/

16 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.

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, properties, events, but also, typically, finding a figurative meaning in cases where the literal pictorial content does not appear to fit well in the given context. In the eggplant case, deriving this figurative meaning involves a rather pure image metaphor, e.g. mapping the depicted vegetable to a body part on the basis of a visual resemblance. In other cases, the metaphoric extension may be partly based on conceptual similarity. Thus, the interpreter may map the depicted red car shaped object to a black BMW on the basis of a mixture of (vague) resemblance and conceptual similarity (viz., both being cars).

Since Lakoff & 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 metaphor or metonymy we find in text or speech can in principle be applied to emojis as well, because ultimately, on the projective account, pictures and linguistic utterances express similar semantic contents (viz. information states). Hence, in the case of the wrapped gift emoji used to denote an electronic gift certificate, we have to move beyond image metaphor and assume a purely conceptual mapping (from box with bow to gift card, on the basis of both being gifts).

Beyond metaphor proper, there are also 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 more generally, while 🥑 depicts a half avocado, but can be used to denote avocados generally, or healthy vegan food more generally, or even early 21st century hipsters.

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. This is not to deny that pragmatic emoji interpretation on my pictorial account – as on a symbolic alternative – may also invoke (partly) conceptual metaphor, or metonymy. All kinds of metaphoric and otherwise 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 similariy', 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 over time, is all well beyond the scope of this 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.

4 Face emojis and expressives

4.1 The special status of face emojis

Up until this point we have been focusing entirely on a specific subclass of emojis, viz. those depicting familiar, concrete objects. In actual usage, object emojis however are decidedly less common than face emojis. According to emojitracker, 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 emoji, which I already put aside as 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 & Bimpikou 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. (2021b); Kaiser & Grosz (2021); Grosz et al. (2021a), face emojis are expressive, meaning that they are used to express the speaker's emotional state, roughly the same way verbal expressives do. Thus, the two utterances in (25) mean roughly the same: Kate said that Sue sent the report and I have a negative emotional attitude about that.

(25)   a.   kate said that sue sent the report to ann 😠 (Grosz et al. 2021b)
        b.   kate said that sue sent the fcking report to ann (Grosz et al. 2021b)

Potts (2007) lists a number of common 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 a. Harris & Potts 2010; Amaral et al. 2007); and they are infinitely gradable (e.g., by varying intonation or repetition).

Emojis satisfy these characteristics. Ad (i), the 😠 in (25-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. Ad (ii), Grosz et al. (2021b) show with corpus data 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.

(26)   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} by car

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

(27)   a.   If I had gone, I'd have missed Ada 😊
       ⤳ I'm happy (that I didn't go)
       ⤳̸ If I'd gone, I'd have been happy (because then I'd have missed Ada)
     b.   By now, Sue hasn't trained for months 🏄 (Grosz et al. 2021b)
       ⤳ surfing is part of the training[17]

Furthermore, Kaiser & Grosz (2021) show experimentally that face emojis are not always anchored to the actual speaker, but 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:

(28)   Ana admired Betty 😊
       ⤳ I'm glad about that
       ⤳ Ana has a positive attitude

Ad (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):

(29)   Omgggggg 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

---

17 Pierini (2021) follows Schlenker's (2018) analysis of co-speech gestures as contributing co-suppositions, i.e. (27-b) licences the inference that if Sue had trained it would have involved surfing. Grosz et al. 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.

I reconcile this with our 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 can be seen as expressives, and that face and hand emojis are use-conditional pictures of the speaker's expressive gestures.

4.2 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 (Potts 2007) to a more abstract semantic notion like use-conditional content (Gutzmann 2015; Kaplan 1999; Predelli 2013; Charlow 2015). While Grosz et al. (2021a) opt for a more Pottsian analysis (defining emotive content in terms of real intervals signifying emotional valence) I'll introduce and adopt the latter, more minimal approach to expressive content in order to avoid making 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)

According to Wittgenstein, "I am frightened" is not amenable to a standard compositional semantic treatment in terms of reference and truth. On the intended reading it's not so much an 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, i.e. 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 $[\![\alpha]\!]$ to denote the descriptive content of α.

To deal with indexicals, Kaplan 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, (30-b), as well as the more abstract 'character' modelling the descriptive linguistic meaning of the sentence, (30-c). Notation: sp_c and ad_c denote the speaker/writer of context c and hearer/reader/interpreter/addressee of c, respectively):

(30)   a.   truth condition: $[\![\text{I see you}]\!]_w^c = 1$ iff $\langle [\![\text{I}]\!]_w^c, [\![\text{you}]\!]_w^c \rangle \in [\![\text{see}]\!]_w^c$ iff $\langle sp_c, ad_c \rangle \in$ $[\![\text{see}]\!]_w^c$ iff sp_c sees ad_c in w.

       b.   truth-conditional content: $[\![\text{I see you}]\!]^c = \{w \mid sp_c \text{ sees } ad_c \text{ in } w\}$

       c.   truth-conditional character: $[\![\text{I see you}]\!] = \{\langle c, w \rangle \mid sp_c \text{ sees } ad_c \text{ in } w\}$

The starting assumption of Kaplan's 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. In the Kaplanian formal framework we can naturally think of use-conditional content as a set of contexts (those contexts where the expression is felicitously used). One of Kaplan's central examples of a use-condition is (31-a). The corresponding use-conditional content can be modeled set-theoretically as in (31-b).

(31)   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: $[\![\text{oops}]\!] = \{c \mid sp_c \text{ observed a minor mishap in } w_c\}$

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 LF's that puts expressives (and their use-conditional interpretations) on top, and descriptions (and their truth-conditional interpretations) below.

(32) $\quad \llbracket \text{Oops, I did it again} \rrbracket = \left\llbracket \dfrac{\text{Oops}}{\text{I did it again}} \right\rrbracket = \dfrac{\llbracket \text{Oops} \rrbracket}{\llbracket \text{I did it again} \rrbracket}$

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

The full linguistic meaning of a sentence is thus a pair consisting of a use-conditional content and a truth-conditional character. 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 content, useful for validating different intuitive inference schemas. For instance, as it stands we can't account for the evident (Moorean) paradoxicality of statements like (33), where the levels of meaning conflict:

(33) $\qquad$ 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 invalidate (33), 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.

4.3 Expressive gestures

Emblematic gestures like the middle finger or waving goodbye are sometimes said to be expressives (Ebert 2014). Indeed, we can easily verify this by checking off Potts's criteria, the way we did for emojis above. For instance, the meaning of the middle finger gesture is surely negative but hard to define descriptively, and can be graded continuously by exaggerating or repeating the gesture (or combining it with facial expressions or verbal expressives).[18] Since, these emblematic gestures are as much conventional, intentional, symbolic, and hence as 'language-like' as Kaplan's 'ouch' and 'oops', it makes sense to analyze them semantically on a par, i.e. as contributing use-conditional content.

(34) $\quad$ a. $\quad$ use condition: use of middle finger gesture is felicitous iff the speaker is very annoyed at the addressee

$\qquad$ b. $\quad$ use-conditional content: $\llbracket \texttt{<middlefinger>} \rrbracket =$

$\qquad\qquad\qquad\qquad = \left\{ c \mid sp_c \text{ is very annoyed at } ad_c \text{ in } w_c \right\}$

---

18 The '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.

Crucially, as with expressive linguistic utterances, 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 a convention, as is someone who gives a stranger the middle finger because she herself is angry at her partner. Hence, the use-conditional content provided by (34) is non-trivial.

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 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.[19] Notation: I'm using an 'overline' notation to denote co-speech gestures.

$$(35) \quad \left[\!\!\left[\overline{\text{I'm coming over}}^{\texttt{<smile>}}\right]\!\!\right] = \frac{\left[\!\!\left[\texttt{<smile>}\right]\!\!\right]}{\left[\!\!\left[\text{I'm coming over}\right]\!\!\right]} =$$

$$= \frac{\left\{c \,\middle|\, sp_c \text{ has a friendly disposition towards } ad_c \text{ in } c\right\}}{\left\{\langle c,w\rangle \,\middle|\, sp_c \text{ is coming over in } w\right\}}$$

4.4 Face emojis as use-conditional pictures

Facial expressions and face emojis both behave like expressives, but only face emojis are at the same time pictures. I propose to view 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 overcome this gap we could 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 2006, 2003). 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.[20] Linguistic description is not usually transparent: a linguistic description of a linguistic description of a mountain is a description of a sentence or discourse, not of a mountain. What we really need for our current purposes is a cross-medial transparency principle that allows us to infer that a picture of a sentence or gesture expresses what that sentence or gesture expresses.

---

19 Sarah Zobel and Thomas Weskott p.c. suggest there may be two distinct types of smiles, one purely emotive, simply expressing happiness/contedness, 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. See also Ginzburg et al. (2020) for a detailed semantic analysis of smiles and laughter.

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

One complication in arguing for such a principle involves indexicality: a painting of an inscription that reads 'I love you', if it expresses anything about love, does not necessarily express the painter's love; a photo of Ada giving Stella the finger expresses not the photographer's negative emotion, but (at best) Ada's.

As argued in 4.1, 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 we could simply 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:

(36)   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), (36) 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.

(37)   use-conditional content: $\llbracket 😄 \rrbracket = \left\{ c \,\middle|\, \Pi(w_c, v_c) = 😄 \right\}$

Paraphrasing (37): the use-conditional content of the face emoji 😄 is the set of utterance contexts in which the speaker looks like this: 😄 . Obviously, an actual tweeter doesn't always wear a big grin on her face while typing 😄 . What (37) says is that she conveys to her addressee that she has such a facial expression. In other words, by using 😄 the speaker presents herself as looking that way, but that presentation may involve some pretense, or even deception.

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

(38) $\llbracket \text{I'm coming over } 😄 \rrbracket = \dfrac{\llbracket 😄 \rrbracket}{\llbracket \text{I'm coming over} \rrbracket} =$

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

This gives us the observed speaker orientation of face emojis, because the picture depicts the context from it's canonical viewpoint. But we don't yet see any of the expected emotional content in (38).

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 (38) and the use-conditional content of the smile. Let's consider what the use of the emoji in (38) 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, i.e. relative to Apple's highly stylized projection function, and perhaps also by metaphorical and/or metonymic extension. The speaker's looking like that smiley face plausibly entails that the speaker of the context was smiling.[21] 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, = (35)) we can infer that a (cooperative) speaker that is smiling has a positive disposition towards their addressee. In sum, we can naturally compose the use-conditional pictorial semantics of face emojis with that of facial expressions to derive exactly the kind of transparency we need.

The pictorial account presented above, though consisting of a few more moving parts, has a number of benefits over the lexicalized expressive account championed by Grosz et al. (2021b,a). First, the current account naturally predicts the kind of creative emoji usage that we also encountered with object emojis in section 3.4 (e.g. using 🎁 to illustrate shipping a package that is not a gift, or using 🎻 to denote a viola or cello rather than a violin, etc.). For instance, in the category of face emojis, we have 😣 , which can be used to indicate a host of psychologically unrelated but outwardly similarly looking states, from physical exertion to helplessness. The lexical expressive account then would make 😣 a multiply ambiguous item, while my pictorial account treats it 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.

---

21 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 3.4.

Second, on the lexical expressive account the recent rise of gender- and skin-tone specific emojis is quite puzzling. Let's look at a hand gesture emoji, 👍 , since face emojis don't have skin tone specific versions (yet), though both are equally expressive. For the lexicalist, 👍 literally – if use-conditionally or emotively – expresses that the speaker is giving approval. 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 onto a skin tone matching version than onto a default yellow version[22], 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 overspecification challenge now explain the use of skin tone matching emojis.[23]

5   Conclusions

I have argued that many emojis can be analyzed quite literally as 'little pictures.' Not lexical expressives, typographic gestures, or anaphoric event descriptions, but pictures that, like photos or drawings, inform us what the world looks like. I thus propose a formal semantics of emojis in terms of geometric projection, as used also to model the semantic interpretation of pictures and visual narratives (Greenberg 2013; Abusch 2012).

A lot of the communicative work that emojis do in computer mediated communication, e.g., 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 eventualities, 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

---

22 Recall from section 3.3ff that we had to appeal to either more complicated projection functions or even metaphoric meaning extension to explain the use of the default red car emoji to denote arbitrary cars

23 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 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).

inserted in text messages, like emoticons, ascii drawings, and kaomoji, 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.The expressive use of face and hand emojis however is mainly useful in interactive communication. 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 clearly 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 metaphor and metonymy 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 linguistic scrutiny already independently of emojis I will leave it at this for now.

References

Abusch, Dorit. 2012. Applying Discourse Semantics and Pragmatics to Co-reference in Picture Sequences. *Proceedings of Sinn und Bedeutung* 17.

Abusch, Dorit. 2020. Possible-Worlds Semantics for Pictures. In *The Wiley Blackwell Companion to Semantics*, 1–31. Wiley.

Altshuler, Daniel & Julian J. Schlöder. 2021. If pictures are stative, what does this mean for discourse interpretation? *Proceedings of Sinn und Bedeutung* 25.

Amaral, Patricia, Craige Roberts & E. Allyn Smith. 2007. Review of the logic of conventional implicatures by Chris Potts. *Linguistics and Philosophy* 30(6). 707–749.

Asher, Nicholas & Alex Lascarides. 2003. *Logics of Conversation*. Cambridge University Press.

Barnden, John. 2010. Metaphor and Metonymy: Making their connections more slippery. *Cognitive Linguistics* 21(1). 1–34.

Brasoveanu, Adrian & Jakub Dotlacil. 2015. Incremental and predictive interpretation. *Semantics and Linguistic Theory (SALT)* 25. 57–81.

Charlow, Nate. 2015. Prospects for an Expressivist Theory of Meaning. *Philosophers' Imprint* 15. 1–43.

Cohn, Neil. 2013. *The Visual Language of Comics: Introduction to the Structure and Cognition of Sequential Images.* A&C Black.

Cohn, Neil, Jan Engelen & Joost Schilperoord. 2019. The grammar of emoji? Con-

straints on communicative pictorial sequencing. *Cognitive Research: Principles and Implications* 4(1). 33. http://dx.doi.org/10.1186/s41235-019-0177-0.

Danesi, Marcel. 2016. *The Semiotics of Emoji: The Rise of Visual Language in the Age of the Internet*. Bloomsbury Publishing.

Ebert, Cornelia. 2014. The non-at-issue contributions of gestures and speculations about their origin.

Gawne, Lauren & Gretchen McCulloch. 2019. Emoji as digital gestures. *Language@Internet* 17(2).

Giardino, Valeria & Gabriel Greenberg. 2015. Introduction: Varieties of Iconicity. *Review of Philosophy and Psychology* 6(1). 1–25. http://dx.doi.org/10.1007/s13164-014-0210-7.

Ginzburg, Jonathan, Chiara Mazzocconi & Ye Tian. 2020. Laughter as language. *Glossa: a journal of general linguistics* 5(1). http://dx.doi.org/10.5334/gjgl.1152.

Goodman, Nelson. 1976. *Languages of Art: An Approach to a Theory of Symbols*. Hackett Publishing.

Greenberg, Gabriel. 2013. Beyond Resemblance. *Philosophical Review* 122(2). 215–287. http://dx.doi.org/10.1215/00318108-1963716.

Greenberg, Gabriel. 2019. Tagging: Semantics at the Iconic/Symbolic Interface. *Proceedings of the Amsterdam Colloquium* 22.

Greenberg, Gabriel. 2021. Semantics of Pictorial Space. *Review of Philosophy and Psychology* http://dx.doi.org/10.1007/s13164-020-00513-6.

Grosz, Patrick G., Gabriel Greenberg, Christian De Leon & Elsi Kaiser. 2021a. A semantics of face emoji in discourse lingbuzz/005981.

Grosz, Patrick G., Elsi Kaiser & Francesco Pierini. 2021b. Discourse Anaphoricity and First-Person Indexicality in Emoji Resolution. *Proceedings of Sinn und Bedeutung* 25. http://dx.doi.org/None.

Gutzmann, Daniel. 2015. *Use-Conditional Meaning: Studies in Multidimensional Semantics*. Oxford Univeristy Press.

a. Harris, Jesse & Christopher Potts. 2010. Perspective-shifting with appositives and expressives. *Linguistics and Philosophy* 32(6). 523–552. http://dx.doi.org/10.1007/s10988-010-9070-5.

Hobbs, Jerry. 1990. *Literature and Cognition*. Stanford: CSLI.

Hobbs, Jerry R. 1979. Coherence and Coreference. *Cognitive Science* 3(1). 67–90. http://dx.doi.org/10.1207/s15516709cog0301_4.

Kaiser, Elsi & Patrick Georg Grosz. 2021. Anaphoricity in emoji: An experimental investigation of face and non-face emoji. *Proceedings of the Linguistic Society of America* 6(1). 1009–1023. http://dx.doi.org/10.3765/plsa.v6i1.5067.

Kamp, Hans. 1981. A theory of truth and semantic representation. In Jeroen Groenendijk, Theo Janssen & Martin Stokhof (eds.), *Formal Methods in the Study of Language*, 277–322. Amsterdam: Mathematical Centre Tracts.

Kaplan, David. 1999. The meaning of "ouch" and "oops": Explorations in the theory of meaning as use.

Kehler, Andrew. 2002. *Coherence, Reference, and the Theory of Grammar*.

Kulvicki, John. 2003. Image structure. *The Journal of Aesthetics and Art Criticism* 61(4). 323–340.

Kulvicki, John. 2006. Pictorial representation. *Philosophy Compass* 1(6). 535–546.

Lakoff, George. 1993. The Contemporary Theory of Metaphor. In A Ortony (ed.), *Metaphor and Thought*, 202–251. Cambridge University Press.

Lakoff, George & Mark Johnson. 2003. *Metaphors We Live By*. University of Chicago Press.

Lascarides, Alex & Matthew Stone. 2009. A Formal semantic analysis of gesture. *Journal of Semantics* 26. 393–449.

Maier, Emar. 2016. A plea against monsters. *Grazer Philosophische Studien* 93. 363–395.

Maier, Emar. 2019. Picturing words: The semantics of speech balloons. *Proceedings of the Amsterdam Colloquium* 22.

Maier, Emar & Sofia Bimpikou. 2019. Shifting perspectives in pictorial narratives. *Sinn und Bedeutung* 23.

Mann, William & Sandra Thompson. 1988. Rhetorical structure theory: Toward a functional theory of text organization. *Text* 8(3). 243–281.

McCulloch, Gretchen & Lauren Gawne. 2018. Emoji Grammar as Beat Gestures .

Pagin, Peter. 2014. Pragmatic Enrichment as Coherence Raising. *Philosophical Studies* 168(1). 59–100. http://dx.doi.org/10.1007/s11098-013-0221-8.

Pasternak, Robert & Lyn Tieu. 2021. Co-linguistic content inferences: From gestures to sound effects and emoji lingbuzz/005082.

Pierini, Francesco. 2021. Emojis and gestures: A new typology. *Proceedings of Sinn und Bedeutung* 25.

Potts, Christopher. 2007. The expressive dimension. *Theoretical Linguistics* 33(2). 165–198.

Predelli, Stefano. 2013. *Meaning without truth*. Oxford: Oxford University Press.

Rooth, Mats & Dorit Abusch. 2017. Picture descriptions and centered content. *Proceedings of Sinn und Bedeutung* 21.

Rooth, Mats & Dorit Abusch. 2019. Indexing across media. *Proceedings of the Amsterdam Colloquium* 22.

Scheffler, Tatjana, Lasse Brandt, Marie de la Fuente & Ivan Nenchev. 2021. In the <palm-emoji> of her hand: Do emojis encode words? Unpublished Ms.

Schlenker, Philippe. 2018. Iconic pragmatics. *Natural Language & Linguistic Theory* 36(3). 877–936. http://dx.doi.org/10.1007/s11049-017-9392-x.

Shan, Chung-chieh. 2011. The character of quotation. *Linguistics and Philosophy* 33(5). 417–443. http://dx.doi.org/10.1007/s10988-011-9085-6.

Tang, Mengmeng, Bingfei Chen, Xiufeng Zhao & Lun Zhao. 2020. Processing network emojis in Chinese sentence context: An ERP study. *Neuroscience Letters* 722. 134815. http://dx.doi.org/10.1016/j.neulet.2020.134815.

Wildfeuer, Janina. 2019. The Inferential Semantics of ComicsPanels and Their Meanings. *Poetics Today* 40(2). 215–234. http://dx.doi.org/10.1215/03335372-7298522.

Wiślicki, Jan. 2021. Scare quotes as deontic modals Unpublished ms.

Wittgenstein, Ludwig. 1958. *Philosophical Investigations*. Oxford: Basil Blackwell.

Wittgenstein, Ludwig. 1966. *Lectures and conversations on aesthetics, psychology and religious belief*. Oxford: Basil Blackwell.