Depicting as a Method of Communication

Herbert H. Clark
Stanford University

In everyday discourse, people describe and point at things, but they also depict things with their hands, arms, head, face, eyes, voice, and body, with and without props. Examples are iconic gestures, facial gestures, quotations of many kinds, full-scale demonstrations, and make-believe play. Depicting, it is argued, is a basic method of communication. It is on a par with describing and pointing, but it works by different principles. The proposal here, called staging theory, is that depictions are physical scenes that people stage for others to use in imagining the scenes they are depicting. Staging a scene is the same type of act that is used by children in make-believe play and by the cast and crew in stage plays. This theory accounts for a diverse set of features of everyday depictions. Although depictions are integral parts of everyday utterances, they are absent from standard models of language processing. To be complete, these models will have to account for depicting as well as describing and pointing.

Keywords: depictions, gestures, quotations, make-believe play, staging scenes

This article is about depicting as a basic method of communication. When we think of depictions, we usually think of artifacts created at one place and time and exhibited at another—paintings, sketches, blueprints, maps, statues, movies, animated cartoons, TV sitcoms, or radio plays. These are exhibited depictions. In everyday discourse, however, people depict things with their hands, arms, head, face, eyes, voice, and body, with and without props. Examples include iconic gestures, facial gestures, quotations of all kinds, full-scale demonstrations, and make-believe play. These are depictions that people create and display with a single set of actions at a single place and time. These I will call performed depictions (often simply depictions).

The goal of this article is to offer a general theory of performed depictions—what they are and, in outline, how they are used and interpreted. We need such a theory for at least three reasons.

First, depicting is a basic method of communicating on a par with describing and indicating (Peirce, 1932, 1974; see Atkin, 2010; Clark, 1996). In describing, people use arbitrary symbols (e.g., words, phrases, nods, and thumbs-up) to denote things categorically, and in indicating, they use pointing, placing, and other indexes to locate things in time and space. In depicting, people create one physical scene to represent another. Although there are rich, comprehensive theories for describing and indicating, there is nothing comparable for depicting.

Second, depicting contrasts most directly with describing, but is based on a fundamentally different mode of thinking. Consider an interview, reported in the New Yorker (August 1, 2011), in which Hollywood director WG told correspondent TF about having to stop filming in New York because of some falcons nesting on the ledge of a building:

“In L.A., they would have”—He leveled a finger at some imaginary nestlings and made a gun-cocking sound.

Although WG could have continued with the description “shot those falcons,” he continued with a depiction—a visible, audible model of shooting. In interpreting “shot those falcons,” TF would have imagined a scene based on the denotations of shoot, those, and falcons. For the depiction, he would have imagined a scene based instead on the scene WG created—the shape, orientation, and time course of his hand movements and tongue-click.

The contrast seems clear. With descriptions, people rely mainly on their knowledge of the vocabulary, syntax, and semantics of, say, English or Japanese and their ability to categorize. With depictions, they rely mainly on their visual, auditory, tactile, and proprioceptive knowledge of physical scenes and on their ability to use one scene in imagining another. Describing depends on knowledge of a language or code. Depicting does not.

Third, depicting is a common part of everyday language. In different studies of storytelling, narrators were found to produce iconic gestures 20 times per minute (Bavelas, Gerwing, & Healing, 2014), facial gestures in overlap with 30% of their words (Bavelas et al., 2014), and direct quotations 5.4 times per minute (Stec, Huiskes, & Redeker, 2015) or containing 17% of their words (Wade & Clark, 1993). As, noted later, depictions are the core of children’s make-believe play.

Despite its prevalence, depicting is absent from standard accounts of language processing (see, e.g., Gleitman & Liberman, 1995; Goldrick, Ferreira, & Miozzo, 2014; Harley, 2013; Spivey, McRae, & Joanisse, 2012). Why? The assumption seems to be that depictions are not part of the syntactic or semantic composition of an utterance, so they can be set aside or ignored. And yet WG used his depiction of falcon shooting as if it were the verb V in “In L.A. they would have V.” Depictions like his—and they come in many forms—are genuine parts of everyday utterances and, therefore,
ought to be included in accounts of language processing. *How* they are to be included is beyond the scope of this article.

What, then, are *depictions*? In the theory proposed here, called the *staging theory*, they are physical scenes that people stage for others to use in imagining the scenes depicted. Staging theory, as I will show, is an important supplement to the theory that functioned as the direct object of “hear.” Examples like this are traditionally called direct quotations or reported speech. In *Dee-duh-dum* and *Bartok*, the depictions were direct objects of “write” and “play.” All three functioned as if they were noun phrases. In *Oboist* and *Falcons*, the depictions functioned as if they were a noun (in “Now, you’ve got a *N*”) and a verb phrase (in “they would have *VP*”). All five assertions would have been incomplete without the depictions.

### Embedded Depictions

Embedded depictions are ones that function as parts of utterances—as if they were words, phrases, or other segments. Examples are shown in Table 1. In the one labeled *Booming*, MB depicted the actions of a former colleague named Smith, a depiction that functioned as the direct object of “hear.” Examples like this are traditionally called direct quotations or reported speech. In *Dee-duh-dum* and *Bartok*, the depictions were direct objects of “write” and “play.” All three functioned as if they were noun phrases. In *Oboist* and *Falcons*, the depictions functioned as if they were a noun (in “Now, you’ve got a *N*”) and a verb phrase (in “they would have *VP*”). All five assertions would have been incomplete without the depictions.

### Indexed Depictions

Indexed depictions are ones that are introduced by expressions such as “this long,” “like this,” or “here” (see Fillmore, 1997; Klipple & Gurney, 2002). Two examples are listed in Table 2. In *Mozart*, MB depicted a passage from Mozart, which he indexed with “this.” In *Raised-voice*, he depicted Smith’s dramatic style of speaking by *fusing* the depiction with his utterance, and he indexed the depiction with “like this.”

Indexing a depiction is different from other deictic references. In *Mozart*, MB said “This is what I want to write” while playing a passage on a piano. Later he said “This is a late eighteenth century Viennese piano” while pointing at a piano. The referent of *this* in *Mozart* was a depiction that MB’s created for the occasion. The referent of *this* in the second utterance was a piano he did not create for the occasion. In *Mozart*, he used his hands and body to depict something; for the Viennese piano, he used his hands and body to locate something. Both were communicative acts, and both were composite parts of the utterances. Depicting things, however, is different from locating things (Atkin, 2010; Clark, 1996; Peirce, 1932).

### Adjunct Depictions

Adjunct depictions are ones that are adjoined to parts of utterances as nonrestrictive modifiers. Two examples are shown in

<table>
<thead>
<tr>
<th>Label</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Booming</td>
<td>[MB was speaking about a former academic colleague named Smith.] And you would go by his office and hear, “(MB turns his head and gaze 20 degrees left, looks at an imaginary person, and speaks in a booming voice with dramatic gestures) Well, I can see you Wednesday at three o’clock.”</td>
</tr>
<tr>
<td>Dee-duh-dum</td>
<td>[MB was discussing a measure in a Mozart sonata.] But then he writes “(gazing at audience and singing) <em>Dee-duh dum.</em>” That is very expressive.</td>
</tr>
<tr>
<td>Oboist</td>
<td>[Oboist RK was speaking to an oboe student.] Now, you have got a <em>(singing in rhythm) <em>Dee dah deh, (speaking) oom, (singing) bah duh dah, bah duh dah deh duh duh.</em> (speaking) oom, (singing) bah duh dah.</em> You have got to feel that.</td>
</tr>
<tr>
<td>Bartok</td>
<td>[MB was discussing a piece played by Bela Bartok.] He does not play <em>(demonstrates four measures on the piano while singing), but rather he plays—and he does it better than I do—</em>(demonstrates the same four measures while singing, but differently).</td>
</tr>
<tr>
<td>Falcons</td>
<td>[Film director WK was speaking to reporter TF] “We were going to shoot on top of the MetLife Building . . . but falcons had nested there, so they shut it down. In L.A., they would have—” He leveled a finger at some imaginary nestlings and made a gun-cocking sound. (New Yorker)</td>
</tr>
</tbody>
</table>
Table 2
Examples of Indexed and Adjunct Depictions

<table>
<thead>
<tr>
<th>Label</th>
<th>Examples from MB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mozart</td>
<td>I am now Mozart. Forgive my immodesty. (Sits down at piano.) (Plays two measures from Mozart sonata) This is what I want to write (2.8 s). (Stands.) (1.6 s) How can I do it?</td>
</tr>
<tr>
<td>Raised voice</td>
<td>At the University of Illinois where I used to teach, there was a voice (raising voice with RH gesturing dramatically toward audience) professor who said you should always speak like this. [This was followed by Booming.]</td>
</tr>
<tr>
<td>Tapering</td>
<td>It means simply (tracing a 30 cm long sideways V in the air with R thumb and forefinger near his R shoulder) what we would call the tapering.</td>
</tr>
<tr>
<td>Disgust</td>
<td>[MB had just played an excerpt from Beethoven on the piano with the wrong phrasing.] (showing disgust on his face and in his voice) It could not be that. I wouldn’t think so.</td>
</tr>
</tbody>
</table>

Table 2. In Tapering, MB’s depiction was timed to overlap with the phrase “what we would call the tapering,” and in Disgust, the depiction was timed to overlap with “It couldn’t be that. I wouldn’t think so.” The expressions they are adjoined to are called affiliates (Schegloff, 1984). The depictions elaborate on their affiliates as if they were non-restrictive relative clauses—hence the term adjunct.

Adjunct depictions are different from indexed depictions. The carrier utterance in Mozart (“This is what I want to write”) would not have been complete without the depiction, but the carrier utterance in Tapering (“It means simply what we would call the tapering”) would have. Indexed depictions are mandatory, but adjunct depictions are not. And yet, adjunct depictions, like all nonrestrictive modifiers, add content to the discourse (see De Ruiter & Wilkins, 1998).

A critical difference between indexed and adjunct depictions is in what indexes what. In Mozart, the word “this” indexed MB’s demonstration on the piano. The direction of indexing was words-to-demonstration. In Tapering, it was the timing of the depiction that indexed the phrase “what we would call the tapering”: its timing, or temporal placement, was used as a temporal index (see Clark, 1996, 2005; Clark & Fox Tree, 2002). Here the direction was depiction-to-words.

Independent Depictions

Other depictions make independent contributions to the discourse. Two examples are shown in Table 3. In Audition, Morgan used the first quotation as a separate contribution to her story. In a corpus of oral narratives studied by Tannen (1986), 26% of the quotations were independent depictions (as in Audition), and the rest were embedded depictions (as in Booming). And in Manifest, the piano student answered LT’s question with a demonstration on the piano, so it, too, was an independent depiction (see Clark, 2012).

Content and Function

Every linguistic expression has both content and a function. Take the bracketed unit in MB’s “I will now illustrate [a passage from Mozart].” Its content was a noun phrase formed from a passage, from, and Mozart. Its function was its use as the direct object X1 in “I will illustrate X1.” The function of the entire clause, in turn, was its use as a contribution to MB’s lecture. With utterances like this, it is descriptions all the way down.

Whenever there are depictions, it is not descriptions all the way down. Compare “I will illustrate X1,” with “He doesn’t play X1,” in Bartok. The units X1 and X2 have roughly the same function—both are direct objects of transitive verbs—but their contents are patently different. X1 is a description, and X2 is a depiction. Table 4 lists the content and function of eight of the examples cited so far.

Here, then, is the first hint of difficulty for models of language processing. None of them are capable of processing constituents that consist of playing the piano, singing an oboe score, or pretending to shoot falcons. Nor can they process adjunct depictions that are overlapped or fused with descriptions. These models will remain incomplete until they deal with all four types of depictions.

Foundations of Depicting

Depicting has been distinguished from describing in literary studies for over two millennia. In Plato’s Republic, depicting, or mimesis, was contrasted with diegesis, or describing (see Bakhtin, 1981; Lodge, 1990). According to Lodge, mimesis is used in modern novels for “direct speech of characters,” and diegesis for “authorial report, description, summary, commentary.” For Plato, the epitome of mimesis was the theater. In comedies and tragedies, the cast and crew create nonactual scenes as a way of depicting scenes that are or could be actual.

Forms of Depicting

Depicting has also been distinguished from describing in scientific studies, though in five quite separate traditions:

1. Iconic gestures, as in Tapering, have been studied as a type of manual gesture (e.g., Cienki & Müller, 2008; Enfield, 2009; Kendon, 1980, 1994, 2004; McNeill,
Table 4

<table>
<thead>
<tr>
<th>Example</th>
<th>Content</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Booming</td>
<td>Enactment of Smith speaking to student</td>
<td>NP in “You would hear NP”</td>
</tr>
<tr>
<td>Bartok</td>
<td>Performance of Bartok on piano</td>
<td>NP in “He doesn’t play NP”</td>
</tr>
<tr>
<td>Mozart</td>
<td>Performance of Mozart on piano</td>
<td>Indicatum of “this”</td>
</tr>
<tr>
<td>Raised voice</td>
<td>Realization of a raised voice and dramatic gestures</td>
<td>Indicatum of “like this”</td>
</tr>
<tr>
<td>Tapering</td>
<td>Enactment of a person grimacing</td>
<td>Adjunct of “what we call the tapering”</td>
</tr>
<tr>
<td>Disgust</td>
<td>Enactment of Gary speaking</td>
<td>Adjunct of “It couldn’t be that. I wouldn’t think so.”</td>
</tr>
<tr>
<td>Audition</td>
<td>Performance of a phrase on piano</td>
<td>Independent contribution</td>
</tr>
<tr>
<td>Manifest</td>
<td></td>
<td>Independent contribution</td>
</tr>
</tbody>
</table>

Dividing depictions into these five classes, however, has led to parochial and often incompatible accounts. Theories that take quotations to be linguistic objects, for example, do not extend to iconic gestures, facial gestures, or make-believe play. And theories that take iconic gestures to be “visible actions” (see Kendon, 2004) do not extend to quotations, demonstrations, or make-believe play. Depictions deserve a single, comprehensive account.

### Depictions as Physical Analogs

Depictions, I argue, are physical analogs of what they represent. They are visible, audible, tactile, or proprioceptive models of things that one could actually see, hear, touch, or feel. The gesture in Tapering was a physical analog of a diminuendo sign, and the facial gesture in Disgust was a physical analog of a person showing disgust.

Performed depictions, like their exhibited cousins, are subject to two well-established principles:

1. **Pas-une-pipe principle.** A depiction is not what it depicts.
2. **Double-reality principle.** A depiction has two realities: its base, or raw execution; and its appearance, the features that are intended to be depictive.

The first principle is named for Magritte’s painting of a briar-wood pipe, “Ceci n’est pas une pipe,” or “This is not a pipe.” The second is based on Gregory’s (1968, 1970) analysis of pictures into two realities: “A picture has a curious double reality, for we see its shapes, colors and textures as more-or-less familiar objects, while at the same time, we know we are seeing blobs of paint” (Gregory, 2003, p. 1233; see also Gombrich, 1960; Maynard, 1994; Walton, 1990). By these principles, depictions entail three scenes—a base scene, a proximal scene, and a distal scene—with mappings from one scene to the next.

### Proximal and Distal Scenes

According to the pas-une-pipe principle, a depiction is different from what it depicts. In Booming, the depiction was a scene in which MB produced “(in a loud voice and with rhetorical gestures) Well, I can see you Wednesday at three o’clock.” What he depicted, however, was a scene in which Smith made an appointment with a student. These are the proximal and distal scenes:

**Proximal scene (the depiction) ⇒ distal scene (what is depicted)**

The proximal scene is in the here-and-now, and the distal scene in a there-and-then—in a world that is displaced in place, time, or reality.

---

1992, 2000, 2008; Streeck, 2008; Streeck, Groethues, & Villanueva, 2009). In their purest form, they contrast with (a) emblems, conventional gestures such as thumbs-up and crossed fingers, (b) deictic gestures, such as pointing or nodding at things, and (c) beat gestures, which mark the timing of certain speech events (see McNeill, 1992).

2. **Facial gestures**, as in Disgust, are often treated as expressions of emotion to be read off the face (e.g., Ekman, 1993; Ekman & Friesen, 1971). Many of these displays are really gestures, genuine acts of communication (Bavelas & Chovil, 2000; Chovil, 1991a,b).

3. **Spoken quotations**, as in Booming, have long been of interest to linguists, philosophers, and psychologists because of the problems they pose for theories of syntax, semantics, logic, and language processing (Cappelen & Lepore, 1997, 2003, 2012; Clark & Gerrig, 1990; Davidson, 1979; De Brabanter, 2005; Goodman, 1974; Partee, 1973; Potts, 2007; Recanati, 2001, 2008; Saka, 1998; Wade & Clark, 1993). In this literature, direct quotations are contrasted with indirect quotations. The first are depictions and the second are descriptions.

4. **Full-scale demonstrations**, as in Bartok, have been studied by psychologists and educators as a method of teaching (e.g., Csibra & Gergely, 2009; Gergely, Egyed, Király, 2007; Shakhashiri, 1989; Sokoloff & Thornton, 2004). Demonstrations like these contrast with narrative descriptions on the same topic.

5. **Make-believe play** has been studied mainly by developmental psychologists (e.g., Bretherton, 1989; Farver, 1992; Fein, 1981; Garvey, 1990; Garvey & Kramer, 1989; Harris & Kavanauagh, 1993; Leslie, 1987; Lowe, 1975; Mitchell, 2002). Children 24 months old are able to perform both independent depictions, such as “I want my teddy bear [whining in a petulant voice],” pretending to be a baby, and embedded depictions, such as “Say, ‘go to sleep now’” (Garvey, 1990, pp. 83–84).
Both of these are physical scenes—configurations in space and time of physical elements that we could, in principle, perceive with our senses. These elements may include people (e.g., MB or Bartok), physical objects and substances (e.g., a piano), people’s actions (MB singing, Smith speaking), other physical events (a metronome ticking), or physical states (the size of the room).

The proximal scene is a physical analog of the distal scene. When physicist Ernest Rutherford first characterized the atom, he chose the analogy “The atom is like the solar system” (Gentner, 1983): The solar system is a physical analog of the atom. The mapping from proximal to distal scenes, denoted by double arrows (\(\Rightarrow\)), is shown in Table 5. Performed depictions are no different. In Booming, MB intended the proximal scene, his dramatization, to be mapped into the distal scene, Smith making an appointment.

All analogies are selective—often highly selective (see Gentner, 1983; Holyoak & Thagard, 1989). The proximal scene represents only part of the distal scene; for example, there are no elements in the solar system that map into protons or neutrons. Furthermore, only some aspects of the proximal scene map into the distal scene, where an aspect is a feature that holds for an entire scene. The aspect “3D geometry,” for example, maps from the solar system to the atom, but “size” does not. Rutherford took full advantage of this selectivity. The analogy he chose did what he wanted it to do and nothing more. People choose depictions for the same reasons.

### Base and Proximal Scenes

By the double-reality principle, every proximal scene is built on a base scene, as illustrated for the sketch of a hydrogen atom in Figure 1:

1. The base scene is the raw, observable, yet-to-be-interpreted physical features of the depiction as executed. It is what the depiction actually is. In my sketch, it includes the paper, pencil marks, and smudges. Without explanation, these could mean anything.

2. The proximal scene is a construction, one that is based on the elements, aspects, and features of the base scene that are depictive. It is what the depiction appears to be. In my sketch, the proximal scene includes—in their idealized form—the large disk, small spot, and circle as props for the hydrogen nucleus, electron, and electron’s orbit. It does not include the border, the initials, or the aspect “black-and-white.”

The mapping from base to proximal scenes is functional: It specifies how the parts of the base scene are to function in the proximal scene. Here functional mappings are denoted with single arrows (\(\rightarrow\)) to distinguish them from analog mappings (\(\Rightarrow\)). Table 6 shows a functional mapping for the sketch in Figure 1.

Every base scene is construed according to an interpretive framework—the assumptions the creator intended to be used in the interpretation of the depiction. I established an interpretive framework for the sketch in Figure 1 by labeling it a “hydrogen atom,” implying the mapping in Table 6. If I had labeled it a “bicycle wheel,” I would have implied these rules instead:

- Large disk \(\rightarrow\) prop for axle
- Small spot \(\rightarrow\) prop for tire valve
- Curved line \(\rightarrow\) prop for tire

There is no such thing as a depiction simpliciter. One cannot know what a base scene depicts without knowing or inferring what its creator intended it to depict—the interpretive framework (Bloom, 2010; Gelman & Bloom, 2000; Gregory, 1968, 2005; Preissler & Bloom, 2008).

### Imagining What Is Depicted

The functional and analog mappings of a depiction (illustrated in Figure 2) are really a system of mappings: Each one constrains

Table 5

<table>
<thead>
<tr>
<th>Proximal element</th>
<th>(\Rightarrow)</th>
<th>Distal element</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sun</td>
<td>(\Rightarrow)</td>
<td>Nucleus</td>
</tr>
<tr>
<td>Planet</td>
<td>(\Rightarrow)</td>
<td>Electron</td>
</tr>
<tr>
<td>Attracts(sun, planet)</td>
<td>(\Rightarrow)</td>
<td>Attracts(nucleus, electron)</td>
</tr>
<tr>
<td>Orbits(planet, sun)</td>
<td>(\Rightarrow)</td>
<td>Orbits(electron, nucleus)</td>
</tr>
<tr>
<td>(\emptyset)</td>
<td>(\Rightarrow)</td>
<td>Proton</td>
</tr>
<tr>
<td>Aspect(3D-geometry)</td>
<td>(\Rightarrow)</td>
<td>Aspect(3D-geometry)</td>
</tr>
<tr>
<td>Aspect(size)</td>
<td>(\Rightarrow)</td>
<td>(\emptyset)</td>
</tr>
</tbody>
</table>

Table 6

<table>
<thead>
<tr>
<th>Base element</th>
<th>(\rightarrow)</th>
<th>Proximal element</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large disk</td>
<td>(\rightarrow)</td>
<td>Prop for nucleus</td>
</tr>
<tr>
<td>Small spot</td>
<td>(\rightarrow)</td>
<td>Prop for electron</td>
</tr>
<tr>
<td>Curved line</td>
<td>(\rightarrow)</td>
<td>Prop for orbit (electron, nucleus)</td>
</tr>
<tr>
<td>Rectangular frame</td>
<td>(\rightarrow)</td>
<td>(\emptyset)</td>
</tr>
<tr>
<td>“HHHC”</td>
<td>(\rightarrow)</td>
<td>(\emptyset)</td>
</tr>
<tr>
<td>Aspect(black-and-white)</td>
<td>(\rightarrow)</td>
<td>(\emptyset)</td>
</tr>
</tbody>
</table>

Figure 1. Sketch of hydrogen atom.
the other. Establishing this system takes both perceptual and conceptual imagination.

In perceptual imagining, we imagine seeing, hearing, touching, or feeling things (Walton, 1990). We can imagine seeing friends, listening to music, swallowing oysters, running, falling out of a window, feeling tense, and feeling sad. These are quasi-perceptual experiences (Shepard, 1984), and they activate the corresponding perceptual systems in the brain (Ganis, Thompson, & Kosslyn, 2004; Kosslyn, Ganis, & Thompson, 2001; Halpern & Zatorre, 1999; Zatorre & Halpern, 2005). In Tapering, the physical base of MB’s depiction helped his audience imagine seeing a large “V” on its side.

In conceptual imagining, on the other hand, we imagine that certain propositions are true (see also Gendler, 2011). We might imagine that today is the first of July 1620 or that Churchill was a secret Baptist. In Tapering, it was the interpretive framework that led MB’s audience to imagine that the sideways V was a prop for a diminuendo sign. Evidence has long shown that perceptual imagining is constrained by conceptual imagining, and vice versa (e.g., Bruner, Busiek, & Minturn, 1952; Carmichael, Hogan, & Walter, 1932), though there is much debate about how.

Perceptual imagining is often bound to physical stimuli in a process I will call percept-bound imagining. In reading tea leaves, fortune tellers “allow the imagination to play around the shapes suggested by them” (Melton, 1996). Once they discover a wedge shape (a base scene), they might construe it as a prop for Cupid’s arrow (the proximal scene) and predict romance. If, instead, they discover a circle, they might construe it as a prop for a wedding ring, or the circle as a prop for an arrow. This is one way conceptual imagination is constrained by perceptual imagination, and vice versa.

Depictions rely on percept-bound imagination. In Tapering, MB designed his gesture (the base scene) to bind his audience into imagining a sideways V. He could not have done that with an up-and-down or circular gesture. At the same time, he relied on the interpretive framework to get his audience to construe the V-ness of his gesture as a prop for a diminuendo sign.

### Depictions as Staged Scenes

Depictions may be physical analogs, but this is only a necessary condition. A depiction is an act of communicating the appearance of a distal scene, and people do that by staging a proximal scene that is a physical analog of the distal scene. In Booming, MB played the role of Smith and enacted him facing a student and booming out “I can see you Wednesday at three o’clock.” In Booming, MB did not “really or actually or literally” make an appointment (Goffman, 1974; see Bateson, 1955). He created the mere appearance of someone doing so. These scenes range from rich, vivid, and detailed, as when MB demonstrated Bartok on the piano, to sparse, incomplete, and skeletal, as when MB sang “dee-duh dum” for three notes from Mozart. As with any communicative act, MB intended his audience and him to mutually recognize his intention in staging these scenes (Grice, 1957).

### Staging Scenes in Make-Believe Play

Staging a scene should be a thoroughly familiar notion because it is what children do in make-believe play. Consider the episode called Go-to-sleep in Table 7 (from Garvey, 1990, pp. 83–84). Sally began by pretending that she was Mother, the doll was Baby, and Mother was putting Baby to bed. She then pretended that Alex was Father, and Father was telling Baby to go to sleep. And she pretended that she was Baby asking Father “Why?” Alex joined her and pretended that he was Father speaking to Baby.

What Sally and Alex did was stage a scene—a brief play about Mother and Father putting Baby to bed. As stage manager, Sally assigned the roles of Mother, Baby, and Father to Alex and herself, and introduced the doll and cradle as props. As script writer, she invented the dialog, and later, as actor, she played Mommy putting Baby to bed and Baby whimpering “Why?” Alex accepted Sally’s stage directions and, as actor, played Father. The children engaged in two types of activities: (a) as stage-managers outside the scene, they established what went into the scene; and (b) as actors within the scene, they played the roles of Mother, Father, and Baby.

The heart of the episode was really an elaborate depiction. The proximal scene was a series of actions in which Sally played Mother and Baby and Alex played Father, and it mapped into a distal scene in which Mother put Baby to bed, Baby whimpered, and Father spoke. The episode was planned and produced with individual depictions. Sally performed embedded depictions in “Say, ‘go to sleep now’” and “No, say ‘Because (emphatically)’.” She also performed independent depictions of Mother putting Baby to bed.

<table>
<thead>
<tr>
<th>Table 7</th>
<th>Go-To-Sleep (an Episode of Make-Believe Play)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Sally (3;3) was holding a baby doll and Alex (2;9) was watching her put it to bed.]</td>
<td></td>
</tr>
<tr>
<td>1. Sally</td>
<td>Say, “Go to sleep now.”</td>
</tr>
<tr>
<td>2. Alex</td>
<td>Go to sleep now.</td>
</tr>
<tr>
<td>6. Alex</td>
<td>Because.</td>
</tr>
<tr>
<td>7. Sally</td>
<td>No, say “Because (emphatically)”</td>
</tr>
<tr>
<td>8. Alex</td>
<td>Because. (emphatically)</td>
</tr>
</tbody>
</table>
Make-Believe Play as Theater

What, then, is make-believe play? In most accounts, it is taken to be a type of theater. It was described as “socio-drama” by Piaget (1962), “dramatic play” by El’konin (1969), and enactments of social roles by Vygotsky (1978), three pioneers in the field. It has since been characterized with a range of concepts from the theater, as in these examples (variously from Bretherton, 1984, 1989; Dunn & Dale, 1984; Garvey, 1990; Giffin, 1984; Haight & Miller, 1993; Harris, 2000; Lillard, Pinkham, & Smith, 2011; Miller & Garvey, 1984; Nelson & Seidman, 1984; Sawyer, 1997; Strömqvist, 1984; Taylor, 1999; Taylor & Mannering, 2006):

1. **Roles**: Children are said to take the “roles” of such “characters” as mothers and cowboys, Santa and Cookie Monster, drivers and passengers.

2. **Acting**: Children are said to “enact” roles or do “pretend actions of play characters.” They might “play” doctor or fireman.

3. **Props**: Children are said to use “props” such as dolls, toy beds, and blocks, objects that stand in for real objects.

4. **Scripts**: Children are described as establishing “story lines” or “scripts,” such as “playing house” or “playing cops and robbers,” which result in “improvised theater.”

5. **Staged versus real actions**: Children distinguish play from nonplay. They treat play as a “pretend world,” “subjunctive world,” or “alternative world,” as “non-literal,” or a “suspension of objective truth.”

6. **Acting versus stage managing**: Children distinguish “acting” (e.g., Sally’s whining “Why?”) from “stage directions” (e.g., Sally’s “Say, ‘Go to sleep now’”).

Make-believe play changes as children grow up (see, e.g., Garvey, 1990). The earliest forms are primitive, and are initiated and sustained by caregivers. Later forms are elaborate, self-initiated, and self-sustained. Children and young adults take part in ever more sophisticated and institutionalized forms of play—ritualized games, games with rules (Garvey, 1990), computer games (Goldstein, 1994), and, of course, theater (Goffman, 1974). The assumption is that the theater is really a highly evolved, institutionalized form of make-believe play.

The theater, like make-believe play, is a family of forms. There are stage plays, silent mime shows, puppet shows, radio plays, play readings, improv theater, street theater, and skits. All of these depict events. There are also tableaux vivants, an obsolete type of theater, that depict states, or instantaneous time-slices of scenes. Theater pieces may have many actors, one actor, or none, a large stage crew or none, many props or none, a script or no script, and they may be fiction or nonfiction. The theater represents not only what is common to staged scenes, but how they vary.

**Logic of Staging Theory**

The evidence I will cite for staging theory is based on this logic. If depictions are a form of make-believe play, and the theater is an evolved, institutionalized form of make-believe play, then depictions should reflect both what is common and what varies in staging scenes in the theater. The strategy here is to show how principles from the theater account for many features of everyday depictions. These principles fall into six areas:

1. **Participants**, such as actors, prop-managers, and spectators;

2. **Actor versus prop depictions**;

3. **Props**, both physical and imaginary;

4. **Spatial frames of reference** within scenes and within the discourse;

5. **Temporal frames of reference**, also within scenes and within the discourse;

6. **Actual and displaced worlds** in discourse.

As a characteristic example of the theater, I will refer to a performance I saw in London of Samuel Beckett’s *Waiting for Godot* in which the characters Estragon and Vladimir were played by Ian McKellen and Patrick Stewart.

**Actors and Props**

Those who execute scenes in the theater are either **actors** or **prop-managers**. The actors enact characters *inside* the scenes, working on stage in plain sight. The prop-managers realize props from *outside* the scenes. Even when the prop-managers are visible or audible, they are treated as invisible and inaudible. Actors have identities; prop-managers do not.

Staged scenes also typically include **spectators**. When I saw *Godot*, I, too, was a participant. As Jorge Luis Borges (1960/1999, see Walton, 1978) put it: “the actor...on the stage plays at being another before a gathering of people who play at taking him for that other person (p. 300).” Therefore, while McKellen and Stewart were playing Estragon and Vladimir, we in the audience were playing our own roles.

What, then, are these roles? The stage is usually treated as having a fourth wall. The pretense is that the actors on stage are unaware of the spectators, yet the spectators can see and hear everything that happens on stage. In the physical analog, then, the spectators play the role of **covert observers** of the distal scene. At *Godot*, we in the audience played characters who covertly watched Estragon and Vladimir as they bantered in the countryside. The general mappings are these:

- Cast member → actor → character
- Cast member’s action → enactment ⇒ character’s action
- Prop-manager → ø
- Recipient → spectator ⇒ covert observer
Actor and Prop Depictions

If depictions are staged scenes, they, too, should be performed by actors or prop-managers, and they are. The two basic formats are illustrated by Booming and Tapering, repeated in Table 8. In Booming, MB placed himself inside the proximal scene and enacted what Smith said. In Tapering, MB placed himself outside the proximal scene and realized a V-shaped prop inside it. The first I will call an actor depiction, and the second, a prop depiction:

An actor depiction is the selective enactment, by a producer within the staged scene, of a character in a distal scene.

A prop depiction is the selective realization, by a producer outside the staged scene, of one or more props for a distal scene.

Actor and prop depictions are named for the main element the producer controls—an actor, or props. If the producer is an actor, the depiction is an actor depiction; if not, it is a prop depiction.

Actor depictions can be further classified by the characters they enact (by whose persona, or mask, they put on): the producer himself or herself, an addressee, or a third person. These are 1P, 2P, and 3P actor depictions. Every character in a distal scene is associated with an I, here, and now, which anchor the character and his or her actions to that scene (see Bühler’s, 1934 notion of origo. In Booming, the person MB was quoting (“I”) was Smith, so this was a 3P actor depiction. In Audition, when Morgan said, “And I was kind of like ‘Whatever’ you know,” the person she was quoting was Morgan herself, so this was a 1P actor depiction (for further divisions, see Engberg-Pedersen, 2015).

The characters depicted may be robots, bears, dogs, or other creatures. In make-believe play, one child enacted a baby fish with “Hello, I’m a baby fish” (said in a high squeaky voice while making the fish swim toward the partner) (Garvey, 1990, p. 88).

**Iconic Gestures**

Actor and prop depictions have been distinguished in iconic gestures for over a century (Kendon, 2004, Chapter 6). They have been called, among other things, kinetographic versus iconographic gestures (Efron, 1941), kinetic versus pictorial illustrators (Ekman and Friesen, 1969), and gestures with a character versus an observer viewpoint (McNeill, 1992; for more on viewpoints, see later).

The difference between the two types is clear. In retellings of Snow White recorded by McNeill (1992), one narrator enacted Snow White sweeping the floor by pretending to hold a full-sized broom and making sweeping motions. This was a 3P actor depiction. Another narrator depicted a person walking by wiggling two fingers pointed downward across an imaginary surface. This was a prop depiction. Both gestures depicted people, but the first was by an actor inside the scene, and the second, by a prop-manager outside the scene. As McNeill put it, the first “incorporates the speaker’s body into the gesture space,” and the second “excludes the speaker’s body from the gesture space” (p. 119). This is a criterial difference between actor and prop depictions.

**Quotations**

Everyday quotations also come as both actor and prop depictions. The prototypical quotation is reported speech, a 3P actor depiction of what someone said. In Booming, MB enacted Smith, a third party, booming out, “Well, I can see you Wednesday at three o’clock.” Other quotations are 1P or 2P actor depictions. In one narrative, a speaker depicted herself screaming and gesturing dramatically, “I went out of my mind and I just screamed I said ‘Take that out! That’s not for me!’” (Polanyi, 1989, p. 92). In Go-to-bed (see Table 7), Sally told Alex “Say ‘Go to sleep now’” to which he responded “Go sleep now.” Her quotation was a 2P actor depiction, and his line was a 3P actor depiction.

Other quotations depict actions that are not linguistic. In Bartok, MB produced two 3P enactments, not of a person speaking, but of a person playing the piano. His lecture was filled with such quotations, including 1P actor depictions such as “I might play [demonstration on the piano]” and 2P actor depictions (as in Bink-bink cited later). Some of these were embedded depictions; others were independent depictions.

Still other quotations are prop depictions—depictions of objects, states, or events by a producer outside the scene. The examples in Table 9 are from callers talking to the hosts of the American radio program Car Talk. Both callers depicted car noises, the first embedded as a noun in “the car has a kind of a X,” and the second embedded as an adverb in “it goes X.” And in Dee-duh-dum, MB depicted three notes from a Mozart sonata embedded as the direct object in “he writes X.”

Examples like these are further evidence against the widely held reproduction theories of quotation (e.g., Cappelen & Lepore, 1997, 2003, 2012; Davidson, 1979; Geurts & Maier, 2003; Goodman, 1978; Pagin & Westerståhl, 2010; Partee, 1973; Potts, 2007; Saka, 1998). According to these theories, quotations are the mention or reproduction of linguistic expressions. The materials quoted in Bartok, Dee-duh-dum, Worry-noise, and Puttrng, of course, are not linguistic expressions, or even forms of communication (see Clark & Gerrig, 1990; De Brahanter, 2005; Holt, 2009; Recanati, 2001, 2008; Tannen, 1989; Wade & Clark, 1993).

**Hybrid Depictions**

In the theater, a single person can perform two or more roles at once. In Punch and Judy shows, the puppeteer is both a prop-manager, manipulating the puppets, and an actor, enacting the puppets’ speech. Ventriloquists are also both prop-managers and actors as they manipulate and speak for their dummies. These are hybrid depictions, and in everyday discourse, they can be actor–actor, actor–prop, or prop–prop depictions.

**Table 8**

<table>
<thead>
<tr>
<th>Paradigmatic Actor and Prop Depictions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Actor depiction (Booming)</strong></td>
</tr>
<tr>
<td><strong>Prop depiction (Tapering)</strong></td>
</tr>
</tbody>
</table>
Hybrid depictions are ubiquitous in make-believe play. In *Go-to-bed* (see Table 7), Sally managed the doll and, at the same time, whined “Why?” in the doll’s voice. Like a puppeteer, she was both prop–manager and actor, creating an actor–prop depiction. Another girl pushed a toy car toward her brother and said “Beep. Brmm” (Dunn & Dale, 1984, p. 154) This was a prop–prop depiction; one prop was the car and its movements, and the other was its sounds.

Hybrid, or dual, depictions are also well documented in adults (see Dudis, 2004; McNeill, 1992; Parrill, 2009; Perniss, 2007; Perniss & Özyürek, 2008; Stec, 2012). Three examples are shown in Table 10:

1. In *Bink-bink* MB enacted a person striking piano keys (a 2P actor depiction) and, simultaneously, depicted the sounds of the keys being struck, “Bink bink bink bink” (a prop depiction).
2. In *Conductor*, MB sang “dee-duh dum” (a prop depiction) while gesturing like an orchestra conductor (a 3P actor depiction). (MB’s in-breath was a prop depiction of the silence between two notes.)
3. In *Lift-your-chin*, Sevigny enacted herself lifting her chin (a 1P actor depiction) and, simultaneously, enacted a cinematographer placing the tips of his fingers under her chin and saying, “Lift your chin” (a 3P actor depiction in two parts).

How, then, do people perform and interpret actor and prop depictions? Let us start with actor depictions.

**Actor Depictions**

Actors are usually described as playing characters, but that is not right: As with any analog, they enact only a selection of the characters’ actions. In *Godot*, McKellen did not play Estragon as a whole. All he enacted were Estragon’s vocal, facial, and bodily actions. In a radio play, all he would have enacted would have been Estragon’s vocal actions. Everyday depictions are just as selective. In *Booming*, MB enacted Smith’s posture, head movements, arm, hand, and facial gestures, sentence content, and voice amplitude, and yet moments earlier, in *Raised voice*, all he enacted was Smith’s hand gestures and voice amplitude.

So producers select a subset of a character’s actions to enact, but which ones? Many selections are formed around a section of the body—the voice, face, head, hands and arms, torso, or entire body.

**Vocal Depictions**

In depicting a person’s vocal action, producers have options in which of its aspects to enact. Although the quotations in Table 11 look simple enough, they are not what they seem.

1. In *Tiffany’s*, the narrator had memorized verbatim what the clerk had said, and yet she enacted the force of the clerk’s question and not his wording. Why? To fit the style of her ongoing narrative (Wade & Clark, 1993; see also Holt, 2000, 2009).
2. In *Rolled-r*, the speaker did not enact any of Paisley’s actual utterances. He invented an utterance with many r’s to depict the way Paisley rolled his r’s (Clark & Gerrig, 1990, p. 799).
3. In *Zschau*, what Ed Zschau enacted was not a particular utterance, but an instance of the type of question he had often been asked (Clark & Gerrig, 1990, p. 773).
4. And in *Haydn*, MB enacted what Haydn said, not in German, but in an English version his audience could understand. Here the aspect “language of producer” was not depictive, but supportive.

**Table 9**

<table>
<thead>
<tr>
<th>Label</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Worry-noise</td>
<td>My son and I noticed that the car has a-uh, a worry noise, kind of a “(low growling noise)” like when you used to put a baseball card in a-uh, the spokes of your bike.</td>
</tr>
<tr>
<td>Purring</td>
<td>And then, about one and a half or two seconds into my foot being on the accelerator, it goes “[gurgling sound].” has a real deep lion-purring sound to it.</td>
</tr>
</tbody>
</table>

Table 10

**Examples of Hybrid Depictions**

<table>
<thead>
<tr>
<th>Label</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bink-bink</td>
<td>[MB was speaking to a woman in the audience.] All you have to do is say, “Look, there are four notes here, but it’s not really very natural to go (thrusting R index finger four times in rhythm onto an imaginary piano keyboard) bink bink bink bink.”</td>
</tr>
<tr>
<td>Conductor</td>
<td>[MB was speaking about a phrase from a Mozart sonata.] But “(singing and directing with hands) dee-duh dum” has this little “(in-breath)” in it, which is so nice.</td>
</tr>
<tr>
<td>Lift-your-chin</td>
<td>[Actress Chloe Sevigny was describing how a cinematographer told her to hold her chin up during a scene he was shooting.] Before every take, the (directing her hand, palm down, sagittally toward her throat, putting her chin from underneath, and lifting her chin with each pat) cinematographer would be like, “Lift your chin.” So when I watched the movie, I’m . . . (from the Conan O’Brien Show)</td>
</tr>
</tbody>
</table>
Speakers can also select the dialect, lisping, drunkenness, and other aspects of a person’s voice (Clark & Gerrig, 1990). In narratives, speakers often go further and dramatize a character’s attitude, emotion, or tone of voice (see later).

**Facial Depictions**

Most facial gestures are actor depictions. In *Disgust*, MB had just finished a bad rendition of Beethoven when he turned to his audience and, putting on a look of disgust, said, “It couldn’t be that. I wouldn’t think so.” He used the 1P actor depiction as an adjunct to his assertions. Note that he depicted his disgust, not during the bad rendition of Beethoven when he might have been disgusted, but during his assertions about it. Facial gestures of emotions are generally performed in the absence of those emotions (Bavelas & Chovil, 2000; Chovil, 1991a, 1991b). They are there to depict emotions, not express them.

Everyday conversations are filled with facial gestures. In a sample of 1,000 such gestures (Chovil, 1991a), 50% of them depicted how people felt. Two examples are shown in Table 12. In *Desserts* (Chovil, 1991a, p. 180), the speaker performed a 1P actor depiction of the face as an adjunct to her assertion about desserts. And in *Looked-down* (Chovil, 1991a, p. 182), the speaker “re-enacted the [guy’s] look by tilting his head down, lowering his eyes, and forming a straight mouth” as a 3P actor depiction, which he embedded in his utterance as an adjective. Other facial gestures depicted “slight disgust,” “disbelief and surprise,” “amazement,” and “pain.”

The facial gesture in *Disgust* belongs to a class of 1P actor depictions that make public what the producer is thinking. Not all of these depict affect. One type is the *thinking face*, as in Table 12 (from Goodwin & Goodwin, 1986). Allan was speaking to Barbara when he gazed to one side and put on a conventional expression of someone thinking hard. He initiated the gesture at “uhm” and ended it at “tch!” to tell Barbara that he was searching for a word. Another such gesture is the so-called *funny face* to depict uncertainty (Krahmer & Swerts, 2005; Swerts & Krahmer, 2005). Both of these gestures are marked by features such as depressed lip corners, stretched lips, widened eyes, moving eyebrows, and averted gaze (see also Chovil, 1991a).

Facial gestures can also be 2P actor depictions. In an experiment by Bavelas, Black, Lemery, and Mullett (1986), individual students watched a research assistant carry a large TV set into the lab room and, in a carefully contrived accident, drop it on his finger. The assistant, feigning pain, then did one of two things: He looked up at the student, or he hunched down without looking at the student. As captured on video, most students began forming a grimace a fraction of a second after the onset of the accident. When the assistant looked up, the nascent grimaces turned into full grimaces, but when he did *not* look up, they quickly faded to nothing.

The full blown grimaces were 2P actor depictions. If the students themselves had experienced pain, *all* of the nascent grimaces should have become full blown, but they did not. The students used the full blown grimaces to enact the emotion they attributed to the assistant (“you”) as a way of communicating sympathy—“I feel your pain” (Bavelas et al., 1986; Bavelas, Black, Chovil, Lemery, & Mullett, 1988; Bavelas, Black, Lemery, & Mullett, 1990). A few students combined their grimaces with “Ouch!” in composite 2P actor depictions.

**Broader Depictions**

Vocal, facial, and manual enactments are often part of broader depictions, as illustrated in Table 13. In *Schubert*, DN did more than enact the tenor singing a few words from Schubert’s “Winterreise.” He combined the singing with gestures and posturing toward an imaginary concert audience. MB did much the same in *Booming*. And in *Come-out*, Lauren enacted not only Lo’s conspiratorial speech style, but her conspiratorial face and three “come here” gestures. Dramatizations like these are common in narratives (Bavelas, Gerwing, & Healing, 2014; Holt, 2000, 2009; Stec, Huiskes, & Redeker, 2015; Tannen, 1989; Wade & Clark, 1993).

---

**Table 11**

**Direct Quotations of Vocal Actions**

<table>
<thead>
<tr>
<th>Label</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tiffany’s</td>
<td>[Speaker was narrating a scene from Breakfast at Tiffany’s and knew the script verbatim. In that scene, the clerk said, “May I ask how limited?”] And the clerk says, “Well, how much can you spend?”</td>
</tr>
<tr>
<td>Rolled-r</td>
<td>Old Jim Paisley goes, “oh dear oh dear it must be the beer,” he always says that.</td>
</tr>
<tr>
<td>Zhau</td>
<td>Many people have come up to me and said, “Ed, why don’t you run for the Senate?”</td>
</tr>
<tr>
<td>Haydn</td>
<td>[MB was speaking to DN.] Do you really think that Haydn went back from London to Vienna and said, “Now listen everybody. I wrote all these pieces in London, <em>(raising pitch of voice)</em> and I don’t want any of you to play them on the pianos that you have here?”</td>
</tr>
</tbody>
</table>

---

**Table 12**

**Facial Gestures as Actor Depictions**

<table>
<thead>
<tr>
<th>Label</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Desserts</td>
<td><em>(Speaker “raised her eyebrows, then raised one side of her upper lip, and squinted her eyes, which was seen as illustrating her dislike of the dessert”)</em> I hate, I hate desserts with alcohol in them.</td>
</tr>
<tr>
<td>Looked-down</td>
<td>[Speaker was describing “how a person had given him a particular kind of ‘look.’”] . . . and the guy just sort of looked <em>[facial gesture]</em> you know, sorta looked down on me</td>
</tr>
<tr>
<td>Thinking-face</td>
<td>He put <em>(gazing to one size with a thinking face)</em> uhm, (.7) tch! Put crabmeat on th’bo:dum [bottom].</td>
</tr>
</tbody>
</table>
Table 13
Actor Depictions With Gestures

<table>
<thead>
<tr>
<th>Label</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schubert</td>
<td>[DN was telling MB about a tenor in a recital.] He found he could do his “(breaking eye-contact, turning 60 degrees left, gesturing dramatically toward an invisible audience, and singing) Fremd bin ich ein…” in a much more conversational way...</td>
</tr>
<tr>
<td>Come-out</td>
<td>[Lauren, a California high school student, was telling an interviewer about her friend Lo.] I could see her through the window and she was trying to get my attention, and she was like “(breathlessly, with L hand and arm pointing upward, making three “come-here” motions toward her R shoulder) Come out, come out.” (from Laguna Beach)</td>
</tr>
</tbody>
</table>

In staging theory, what producers are trying to do is stage scenes and not simply collections of aspects. If so, the aspects they combine should form coherent scenes, and they seem to. In Schubert, DN surely did not enact the tenor’s voice, gestures, and posture as three separate actions and then synchronize them. He enacted the bundle of actions as a whole.

Joint Depictions

In the theater, actors sometimes violate the fourth wall and enlist the audience as coactors. In American melodramas, actors invite the audience to cheer the hero and boo the villain, and in British pantomimes, some actors interact with the audience, and others do not (Taylor, 2007). In immersive theater, the audience is often treated as coactors (see White, 2012), and in Disneyland, the actors who play Mickey Mouse and Snow White speak freely with visitors.

The same happens in everyday depictions. In Manager (see Table 14), Ellen incorporated Kate into the proximal scene of her depiction and treated Kate as if she were the person Lorna was speaking to. That is, Ellen, in her role as Lorna, enlisted Kate to play the role of Lorna’s addressee. For similar examples, imagine Allan telling Sam “She was so mad that she shook me like this (shaking Sam by the shoulders)” or “The ball grazed my shoulder like this (brushing Sam’s shoulder with one hand).” The first is a 3P actor depiction, and the second, a prop depiction.

Joint depictions are a common part of many demonstrations. In a self-defense class analyzed by Stukenbrock (2012), the instructor demonstrated body throws by enlisting students and confederates as coactors. Furthermore, joint depictions are the core of much make-believe play. In Go-to-bed (see Table 7), Sally and Alex performed joint enactments throughout. Children are able to do that from about two-and-a-half (Garvey, 1990; Miller & Garvey, 1984).

Table 14
Example of a Joint Depiction

<table>
<thead>
<tr>
<th>Label</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manager</td>
<td>[Ellen was speaking to Kate.] And my friend Lorna, if something’s wrong like that, the first thing she’d say, “(pointing index finger at Kate, and looking Kate in the eye) Give me the manager. Let me talk to your manager.”</td>
</tr>
</tbody>
</table>

Note. Example from video-recording by Kashmiri Stec (personal communication).

Physical Props

In the theater, most staged scenes have props. These are elements in proximal scenes (other than the actors and enactments) that map into real elements in distal scenes. Props can be physical or imaginary. In Godot, the physical props included a road, a tree, greasepaint, scenary, a moon rising, evening light. The imaginary props included a person McKellen created by looking off stage at someone in the distance. Physical props may be: both visual and auditory, as in Godot; visual only, as in silent mime shows; auditory only, as in radio plays; or even visual, auditory, tactual, and proprioceptive, as in immersive theater. And in silent mime shows, most props are imaginary.

Props can be further classified as center, side, or back props. Center props are those that are realized by prop-managers alone, such as the offstage noises and rising moon in Godot. Side props are the ancillary objects or events actors use in the course of their enactments, such as McKellen’s boots and bowler hat. Back props are those that do not change during a scene, such as the mound, tree, road, and scenery in Godot. Furthermore, physical props can range from genuine to figural, as illustrated in Table 15. If depictions are staged scenes, they, too, should include both physical and imaginary center, side, and back props, and they do. I begin with physical props.

Status of Props

First, a warning. It is often stipulated that for an action to be an iconic gesture, or a quotation, it must be performed without physical props. In a theory of depictions, this stipulation makes no sense. When I say “George holds his pen like this,” I could mold my hand as if I were grasping a pen, or I could mold my hand around a physical pen. The two actions are surely different, and yet both depict George holding a pen. The second is no less a depiction—or a gesture à la Kendon (2004)—simply because the prop is physical.

Furthermore, physical props are not tools. People can use pencils, sticks, or chalk for drawing pictures, and computers for creating pictures or sounds. These are tools, not props, because they do not map into elements in the distal scene. The focus here is on depictions that do not require tools.

Center Props

Center props are those that are realized by producers from outside the scene, and the depiction is centered on the prop itself. The props people use depend in part on availability. In a study by Engle (2000), tutors were asked to instruct students on how a door
lock worked. Two of their depictions are shown in Table 16. When the tutors had an actual lock and key in hand, they tended to use them as genuine props, as in Cotter-pins. When they did not, they often created somatic props, as in Grooves, in which the tutor used his knuckles as a prop for a key’s grooves. In Protractor, a math instructor used a pen and protractor as ad hoc props for a man and a woman (Radford, 2009, p. 117).

In face-to-face dialog, somatic props are often the only option available. Some of these are novel, like the knuckles in Grooves, but others are more or less conventional, such as these:

a. A hand-phone made with an extended thumb and little finger;

b. Hand scissors fashioned from two fingers opening and closing;

c. A hand pistol made from a vertical fist and extended index finger;

d. A hammer or gavel made from a fist plus forearm; and

e. A person walking formed from two fingers pointing down and wiggling on a surface.

Other somatic props are auditory, such as hissing to depict a snake and a hand clap to depict a door slam. Still others are tactile, such as drawing a finger across a friend’s arm to depict a spider.

Make-believe play relies heavily on props that are replicas (Bretherton, 1984; Garvey, 1990; Harris, 2000). Many of these are toys specifically manufactured to be used as center props—dolls, doll houses, train sets, toy cars, and toy animals. And many of these are replicas not only in shape, but in sound (e.g., dolls that cry, toy cows that moo) or in touch (e.g., soft animals, rubbery worms).

Children also use center props that are ad hoc. In one example (Harris, 2000), a 2-year-old put a teddy bear into a bathtub (a shoebox), rubbed its back with soap (a wooden block), and then, because “He’s all wet,” wrapped it in a towel (a piece of paper). As a rule, children find it easier to pretend with replica than with ad hoc props (Fein, 1975; Mann, 1984).

Side Props

People enacting characters often use physical props as ancillary support—as side props. In Godot, McKellen used a genuine boot in enacting Estragon trying to take off his boot, and in Shakespeare, actors use fake daggers, fake swords, and fake poison. In everyday depictions, side props range from genuine to ad hoc.

A great many side props are somatic, as illustrated in Table 17. In Hand-phone and Falcons (repeated from Table 1), the producers J and WK used a conventional hand-phone and hand-pistol. In Arm-viola, a music tutor created a novel prop, using his forearm as the neck of a viola and, at the same time, producing spoken syllables for its sounds in a hybrid depiction.

In make-believe play, children’s side props are often replica props. In Go-to-bed (see Table 7), Sally enacted Mother putting Baby to bed by using a doll and a toy bed as side props. Many toys are designed to be side props—Halloween costumes, water pistols, pedal driven cars, and hobby horses.

Genuine props are especially common in full-scale demonstrations. In Bartok, MB demonstrated on an actual piano. In other examples, music tutors performed demonstrations with actual violins, oboes, and marimbas; tennis coaches with actual rackets and balls; chefs with actual bowls, spoons, flour, eggs, water, and frying pans; and carpenters with actual wood, saws, hammers, screws, and screwdrivers.

Back Props

Back props are found in both actor and prop depictions, and they, too, range from genuine to ad hoc. In demonstrating how a lock works (Engle, 2000), one tutor used her finger to depict a key that she inserted into an actual lock. Her finger was a side prop, and the actual lock was a buck prop.

Table 16
Prop Depictions With Physical Props

<table>
<thead>
<tr>
<th>Label</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cotter-pins</td>
<td>You can probably hear it. (Slowly moves an actual key into the keyhole; three clear separate clicks can be heard). (Finishes sliding the key in with a fourth click) You can hear each pin go up?</td>
</tr>
<tr>
<td>Grooves</td>
<td>And these (lifting up L fist with largest knuckles on top) grooves are specially designed. . . .</td>
</tr>
<tr>
<td>Protractor</td>
<td>Okay, look (4.36 s) Okay, hey, I got it. She . . . That is him, this is her. (The speaker uses a pen and a protractor to simulate his and her motions on the desk) He walks more quickly, so it does this (the distance between the pen and the protractor increases), and there is a bigger distance.</td>
</tr>
</tbody>
</table>
Many back props are figural props such as maps, photographs, and diagrams. In one study (Clark, unpublished), one student was instructed to tell another how to get from one place on campus to another. When the two of them shared a map, the first described the route while tracing it on the map with her finger, as here:

So what you’re gonna do is you’re gonna (tracing to a stop with R index finger) come up (1.03 sec) (tracing to a stop) on, uh, Fremont. [Uh-huh.] (tracing to a stop) Electioneer. [Uh-huh] (tracing to a stop) Up Campus Drive. . . .

The student treated the map as a back prop and depicted the route on top of it. In other examples, people traced mechanical sequences on diagrams (Hegarty, Mayer, Kriz, & Keehner, 2005; Tversky & Suwa, 2009), and changes to architectural plans on blueprints (Heath & Healey, 2012; Murphy, 2004, 2005).

Imaginary Props

Imaginary props are invisible, inaudible, impalpable props that producers bring into existence in the very performance of their depictions. In Godot, McKellen used a physical hat in enacting Estragon doing routines with his hat. In a silent mime version, he would have used an imaginary hat. Everyday depictions are no different. In Bartok, the piano MB used was visible and audible to the audience, but in Bink-bink, there was no piano until he created one—an imaginary one—by acting as if it were visible and audible to the audience.

Imaginary props are found in all types of depictions, but I will focus on iconic gestures, where they have been studied in remarkable detail (e.g., Cienki & Müller, 2008; Enfield, 2009; Kendon, 2004; McNeill, 1992; Streeck, 2008; Streeck, Groethues, & Villanueva, 2009, among many others).

Center Props

When people are outside the proximal scene, they create imaginary props according to an as-if-evidence principle: “From outside the proximal scene, behave as if there is evidence for the physical props inside the scene.” Most of this evidence takes one of four forms: shape, location, movement, and sound.

Shape. Producers can trace an imaginary prop’s shape or exterior surface, as in Tapering. They can delimit part of a prop’s exterior surface, as in the gesture for “I caught a fish this long.” Or they can pretend to handle parts of its exterior surface.

Location. Producers can establish an imaginary prop’s location. In an example from Kendon (2004, p. 148), “M acts as if to sketch a large rectangular object lying on the table in front of him, understood as a depiction of the Christmas cake he refers to verbally.” M locates the cake by tracing its shape at a location in the larger scene.

Motion. Producers can imply an imaginary prop by depicting its motion. In Sparks in Table 18, the man interviewed by a TV reporter accompanied his description of what happened with vivid gestures. He used the second and third gestures to depict the motion of objects emitting sparks and exploding.

Sound. Producers can also imply an imaginary prop by depicting its sound. In Dee-duh-dum, MB created a piano and its sounds by singing “dee-duh-dum.” In Worrying-sound and Purr-ing, the callers implied the presence of a car engine by depicting its sounds. And in Fire in Table 18, the man implied the presence of the fallen electrical pole by depicting its audible aftereffects.

Side Props

When people are inside a proximal scene, they create imaginary props, instead, according to an enacting-as-if principle: “Perform your enactments as if the physical props were present.” The props created by this principle look distinctly different from those created by the as-if-evidence principle.

The true experts in enacting-as-if are silent mimes. To study their techniques, I examined video-recordings of sketches by the mimes Marcel Marceau (MM) and Red Skelton (RS). Segments of two sketches are described in Table 19. As actors in these sketches, MM and RS never once relied on the as-if-evidence principle.

Table 17
Actor Depictions With Physical Props

<table>
<thead>
<tr>
<th>Label</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hand-phone Falcons</td>
<td>[Film director WK was speaking to reporter TF “We were going to shoot on top of the MetLife Building . . . but falcons had nested there, so they shut it down. In L.A., they would have—” He leveled a finger at some imaginary nestlings and made a gun-cocking sound. (New Yorker)]</td>
</tr>
<tr>
<td>Arm-violin</td>
<td>[Violist DM was telling a student about fingering on a viola.] We have to get this cleaner. [Arm-violin]</td>
</tr>
</tbody>
</table>

Table 18
Imaginary Props About a Car Struck by a Falling Electrical Pole

<table>
<thead>
<tr>
<th>Label</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sparks</td>
<td>There was (thumbs pointing off to one side) <strong>arcin’</strong>, (two hands going up) <strong>sparkin’</strong>, (moving the two hands upward and outward in front of him) <strong>blowing up</strong>, <strong>Fire</strong> And then (pronouncing each word with an explosive phonation and with wild head and hand gestures) <strong>‘arc are bam’</strong> fire was coming from everywhere.</td>
</tr>
</tbody>
</table>
Their use of enacting-as-if fell mainly into three classes: acting on, reacting to, and communicating with.  

**Acting on.** In *Smoker*, MM created an imaginary cigarette by picking it up, lighting it, and puffing on it. In another sketch, RS created imaginary swim goggles, waders, suspenders, and swim fins by putting them on.  

**Reacting to.** In other sketches, both MM and RS moved their heads and eyes as they followed imaginary birds, airplanes, lions, bulls, horses, and passers-by. In *Parade*, RS created a marching band in a parade by watching it pass by while marching in place.  

**Communicating with.** In other sketches, MM and RS “spoke” to, “yelled” at, gestured for, saluted, and held hands out for imaginary people.  

The same techniques should occur in everyday discourse, and they do. In *Bink-bink*, MB created an imaginary piano by acting on it—thrusting his right index finger onto its imaginary keys. In *Falcons*, WG depicted the location of the imaginary falcons by leveling his finger in their direction. And in *Booming*, MB depicted an imaginary student by addressing her as if she were there.  

To sum up, imaginary props are created in both prop and actor depictions. Center props are brought into existence by prop-managers displaying as-if evidence of their presence. Side props are brought into existence by actors performing enactments as if the imaginary props were physically present.

### Spatial Frames of Reference

In the theater, people keep track of several frames of reference (FRs) at once. They have to. At any moment in a play, there are three scenes, a base, a proximal, and a distal scene, and each has its own FR—its own spatial coordinates, or *place-frame*, and its own temporal coordinates, or *time-frame*. People represent the elements in each scene with respect to these frames. In this section, I focus on place-frames.  

Typically, it is easy to map the place-frames from one scene to the next. If McKellen is facing Stewart in the base scene, then the actor playing Estragon is facing the actor playing Vladimir in the proximal scene, and Estragon is facing Vladimir in the distal scene. Where, however, is the audience in all this?  

The base scene is itself part of a larger *discourse scene*. In *Godot*, the discourse scene included not only the stage (for the actors and props), but the backstage (for the crew) and the auditorium (for the audience):  

\[
\text{Discourse scene} \supset \text{base scene} \rightarrow \text{proximal scene} \Rightarrow \text{distal scene}
\]

Base scenes have very different configurations in the discourse scene as one goes from standard theaters to theaters-in-the-round, Punch-and-Judy shows, street theater, immersive theater, radio plays, and other venues.

### Spatialization

In everyday circumstances, people have to configure each depiction within the discourse scene, and they do that by a process I will call *spatialization*. To spatialize a depiction is to establish the dimensionality, scale, location, orientation, and granularity of the base scene within the discourse scene (see Table 20).  

The idea behind spatialization is easy to illustrate. At the opening of *Godot*, a lone cast member (McKellen) was sitting on a prop mound with a prop tree behind him. This was the base scene. That scene, however, was oriented in such a way that McKellen was downstage facing the audience and the tree was upstage. If the scene had been rotated 180 degrees, McKellen would have been upstage facing away from the audience with the tree downstage. The base scene would have been the same, but its spatialization would have been different.  

People spatialize depictions largely to suit recipients. In *Godot*, the base scene (McKellen sitting in front of a prop tree) was designed to reflect the distal scene (Estragon sitting in front of a tree). Its orientation (with McKellen downstage and the tree upstage) was chosen to be optimal for the audience to see and hear. Everyday depictions are spatialized for the same reasons.

### Orientation

In spatializing a depiction, people have a choice of orientation. Consider DN’s actions in *Schubert* in Table 21. In Line 3, DN enacted Enrico singing “Fremd bin ich ein-” to an imaginary audience. In doing so, he could have continued to gaze at MB, but if he had, MB might have thought he was singing to him. To prevent this, DN broke his gaze with MB and reoriented his body, head, and gaze 60 degrees to his left (Line 2), and only then did he gesture and sing the four words (Line 3). Even before he had finished singing, however, he had returned his body, head, and gaze to MB (Line 3). These are *orientation shifts*. DN used them

### Table 20 Five Parameters in the Spatialization of Depictions

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Example values of parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Dimensionality</td>
<td>0D, 1D, 2D, 3D</td>
</tr>
<tr>
<td>2. Scale</td>
<td>Miniature, life-sized, enlarged</td>
</tr>
<tr>
<td>3. Location</td>
<td>In front of producer, surrounding producer</td>
</tr>
<tr>
<td>4. Orientation</td>
<td>Rotate left 60 degrees, horizontal-to-vertical</td>
</tr>
<tr>
<td>5. Granularity</td>
<td>Full detail, much detail, skeletal</td>
</tr>
</tbody>
</table>
to make clear the proximal scene did not include MB (see Sidnell, 2006).

Producers use the same orientation shifts in follow-up depictions. In Line 6 of Table 21, DN went on to depict Enrico “pitching himself” over a lot of noise. To mark that scene as having the same place-frame as the first, DN used the same orientation shifts, turning 60 degrees left before the enactment (Line 5) and turning back after it (Line 8).

Producers have a choice in which segments of their body to reorient, as in these examples:

1. Head, eye gaze, body, and gestures (Booming, Schubert, Bink-bink, Smoker, and Parade)
2. Head, eye gaze, and gestures (Lift-your-chin)
3. Head and eye gaze (Thinking-face, Funny-face)
4. Gestures alone (Raised-voice, Come-out)
5. Eye gaze alone (Audition, MTV in Table 22)

In sign languages, quotations are marked by reorienting the entire body in what are called role shifts (Emmorey, 2002; Engberg-Pedersen, 1993).

Orientation shifts go with the characters enacted. In MTV (in Table 22), Lauren depicted alternating-camera-like shots of Lo herself, and Lo by means of slight orientation shifts of her eye gaze left, right, and then left again (along with shifts in her voice). We see and hear the sequence not as three unrelated actor depictions, but as a single conversation (see also Goodwin, 1990). Orientation shifts in eye gaze are especially common in oral narratives (Sidnell, 2006; Stec, Huiskes, & Redeker, 2015; Thompson & Suzuki, 2014).

Placement

People spatializing a depiction also have a choice of placement. People usually perform hand gestures in an imaginary 3D bubble in front of the body called a gesture space (McNeill, 1992). In McNeill’s studies, 95% of the iconic gestures were performed there. People sometimes create imaginary stages within this space, about chest high, on which they perform proximal scenes, much as signers do in sign languages (Emmorey, 2002; Liddell, 1995, 2003). For people who are face to face, this is the space where the gestures are most visible (Gullberg & Holmqvist, 1999).

By the same logic, producers should place gestures in other locations if that would be better for recipients. In Tapering, MB was speaking to about 50 people, so he placed his gesture above his shoulders and left of his head where it was highly visible to everyone. And in a study by Özyürek (2002; see also Campisi & Özyürek, 2013), people who narrated stories to people on their left or their right accommodated by shifting their gesture spaces to their left or their right.

Producers also use placement in marking follow-up depictions. In an example analyzed by Enfield (2009, p. 119), a Lao fisherman described a fish-trap called a sòön5 (translated from Lao):

1. As for the sòön5, they make it (using both hands to form a funnel facing upward) fluted at the mouth
2. and they (holding LH in place while using RH to depict smallness of the opening) make it small
3. (holding LH in place but now using RH to depict a fish swimming into the opening) going in.

In line 1, the fisherman depicted the mouth of the trap as a funnel facing upward (a somatic prop). In line 2, he kept his left hand in place to maintain half of the funnel, and depicted the small opening at the bottom of the funnel and then, in line 3, a fish swimming into that opening. He marked the continuity of these depictions by holding his left hand in place and depicting the opening and the fish with respect to it. One of Enfield’s depictions went through 11 steps (for similar examples, see Kendon, 2004, e.g., pp. 161–163).

And, finally, people use placement in creating private depictions—depictions for their own use. In an unreported experiment by Teenie Matlock and myself, people were given a map and asked to locate routes and landmarks for partners who had their own maps. The instructors tended to gesture over the maps in a private gesture space before performing similar gestures in a public space for their partners. They intended only the public gestures for communication.

Scale and Granularity

To spatialize a depiction is also to select its scale. In the theater, one reflection of scaling is in stage voices and stage gestures. When McKellen produced “Nothing to be done,” he directed his speech at Stewart a few steps away (in a base place-frame), but in a voice loud enough to be heard in the balcony (a discourse place-frame). And when he gestured to Stewart, it was with move-

Table 22

<table>
<thead>
<tr>
<th>Depiction of Dialogue About MTV</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Lauren was recounting a conversation with her friend Lo about producers from MTV coming to their school.]</td>
</tr>
<tr>
<td>1. I don’t know. I had to go to the bathroom and went out there,</td>
</tr>
<tr>
<td>2. and she’s all, “(shifting gaze slightly left, in excited voice) You’ll never guess who’s in the office.”</td>
</tr>
<tr>
<td>3. and I was like, “(shifting gaze slightly down, in second excited voice) Who?”</td>
</tr>
<tr>
<td>4. and she’s like, “(shifting gaze slightly left, in first excited voice) MTV.”</td>
</tr>
</tbody>
</table>

Note. From Laguna Beach (MTV).
ments large enough to be seen in the balcony. The audience would be mistaken if they thought it was Estragon who was exaggerating his speech or gestures.

Producers follow the same logic in everyday depictions (Bavelas, Gerwing, Sutton, Prevost, 2008; Campisi & Özürek, 2013; Debrešljoska, Özürek, Gullberg, & Perniss, 2013; Gerwing & Bavelas, 2004; Holler & Wilkin, 2009). In Tapering, MB turned a 3 cm long diminuendo sign into a 30 cm long gesture to make it visible to his audience. And in Fish-trap, the fisherman reduced the size of the fish-trap to fit his gesture space. People also use smaller gestures when they are on the telephone (Bavelas, Gerwing, Sutton, & Prevost, 2008).

Spatializing a depiction also includes granularity. In one study (Gerwing & Bavelas, 2004), people were asked to tell others how a toy whirligig worked. Their initial depictions were large, detailed, and precise because the information about the toy was new. Their later depictions were smaller, less detailed, and less precise because, by then, the whirligig was already known and taken as given (see also Holler & Wilkin, 2009).

**Dimensionality**

Depictions can be 0D, 1D, 2D, or 3D. MB’s depiction in Dee-duh-dum was effectively 0D, because it was produced by his vocal apparatus at a single point. Someone saying “I caught a fish this long” would use a 1D gesture of length. MB’s depiction of the diminuendo sign in Tapering was a 2D drawing. Most of the other depictions were 3D.

How people spatialize a depiction depends on its dimensionality. For a 0D depiction, producers must select a location, but not a size or orientation. For 1D, 2D, and 3D depictions, they must fix all three parameters. Producers sometimes have a choice of dimensionality. In Come-out, Lauren quoted her friend Lo using “come-here” gestures in a 3D depiction. If all she had enacted was Lo’s speech, it could have been a 0D depiction.

**Viewpoints**

To interpret what people are doing generally requires taking account of their viewpoints. By viewpoint I mean the position of a person in a scene—the location and direction from which he or she sees, hears, feels, and manipulates things. In face-to-face conversation, speakers and addressees tend to display their roles in part by the way they position their bodies, shoulders, heads, eye gaze, and actions in relation to each other (Goodwin, 1981; Kendon, 1990; Schegloff, 1998). These are discourse viewpoints.

Depictions require additional viewpoints. In Schubert, there were only the speaker’s and the addressee’s viewpoints as long as DN and MB were discussing pianos. Though once DN began singing, “Fremd bin ich ein!” he introduced three more. With his 60 degree orientation shift, he created distinct viewpoints (a) for the actor playing Enrico, (b) for the imaginary audience he was facing, and (c) for MB in the role of spectator, viewing his enactment from the side. These were proximal viewpoints.

In distal scenes, every character has a distinct viewpoint. In Schubert, DN and MB imagined a recital hall in which Enrico was facing an actual audience with a covert observer looking on from the side. The viewpoints of Enrico, his audience, and the covert observer were distal viewpoints.

Are all of these viewpoints really needed? In general, yes. Discourse viewpoints are needed for expressions such as I, you, over there, and the other side. Proximal viewpoints are needed for interpreting the spatialization of a depiction; with the orientation shift, DN separated the actor’s, audience’s, and spectator’s viewpoints from the speaker’s and addressee’s. Distal viewpoints are needed for identifying the I, here, and now of the characters depicted. When MB uttered “Well, I can see you Wednesday at three o’clock” in Booming, the referents of I and you were not MB and his audience, but Smith and the student in the distal scene.

**Temporal Frames of Reference**

Every scene also has temporal coordinates, or a time-frame. Early in Godot, McKellen’s “Nothing to be done” was followed by Stewart’s “I’m beginning to come round to that opinion.” We in the audience represented the timing of these two events with respect to the base scene. We also represented the timing of the base scene as a unit with respect to the time-frame of the discourse scene—the play as performed in London that evening.

**Internal Timing**

The events in a base scene typically map moment by moment into the distal scene. In Dee-duh-dum, the tempo and rhythm of MB’s three syllables mapped into the tempo and rhythm of Mozart’s three notes. And yet base timings can be modified when there is reason to do so. Here are several techniques:

1. **Holds:** A producer can pause, or hold, a depiction at particular junctures. In demonstrating a tennis serve, a coach might add a hold at the beginning, the top, and the end of the swing.

2. **Slow-motion:** A producer can perform an entire depiction at half speed.

3. **Condensation:** A producer can depict an entire scene in a condensed form.

4. **Truncation:** A producer can truncate a depiction in the middle.

Holds and slow-motion are common in instructional demonstrations, as in this example:

```
[Pianist NT is asking a student about the phrasing of three bars from Brahms’ first piano concerto.] Are you going (playing piano slowly and singing softly) "(D# E E C#) (rest) (C# D D D) (A A B B C# C#)?
You could also do (playing and singing softly) "(D# E E C#) (rest) (C# D D A) (rest) (A A B B C# C#)."
```

NT played the three bars in slow motion, first in one phrasing and then in another. She added the hold (a “rest”) to mark the difference between them. Condensation is illustrated by the hand gestures called “markers” that ballet dancers use in discussing and rehearsing dance routines (Kirsh, 2010, 2011). Each marker is a compressed depiction of a specific ballet movement. And truncation is illustrated by “Fremd bin ich ein!” in Schubert.
Phases of Discourse

For most depictions, the discourse scene divides into three phases: preparation, execution, and follow-up. In Go-to-bed (see Table 7), Sally and Alex prepared their scene by setting up props (e.g., the doll and doll’s bed) and issuing stage directions (e.g., Sally’s “Say, ‘Go to bed’”). They distinguished these from their enactments, such as Alex’s “Go to bed” and Sally’s “(whining) Why?” And in a follow-up phase, they coordinated in closing down their play and leaving.

In everyday discourse, the three phases are usually local to the utterance produced. In Schubert (see Table 21), the first clause of DN’s utterance divided as follows:

1. Preparation phase: (DN and MB are facing each other two meters apart. DN is speaking.) He (turning his body, head, and gaze 60 degrees left) found he could do his

2. Execution phase: (gesturing and singing to an imaginary audience) Freund bin ich ein.

3. Follow-up phase: (at ein-, returning his body, head, and gaze to BM) in a much more natural way . . .

DN produced the entire clause as part of DN and MB’s discussion of pianos. Within that clause, however, he timed the depiction locally to fit the N slot in “He found he could do his N in a much more natural way.”

Discourse Timing

For local timing, a depiction can be: (a) slotted within or between utterances; (b) overlapped with an utterance part; or (c) fused with an utterance part. Depictions, of course, cannot overlap descriptions produced with the same body part, or be fused with ones produced with a different body part. Otherwise, the three techniques are used for different purposes.

Embedding. The function of an embedded depiction is specified by the slot in which it is embedded. In Schubert, the depiction filled the slot for the nominal N in “his N.” Other depictions cited in this article occupied slots for noun phrases, verbs, adjectives, and adverbs. The same logic applies to independent depictions. The student in Manifest placed her depiction in the slot for an answer to a question.

One way to embed depictions is with quotatives such as “he said,” “they go,” “she’s like,” or “Mozart wrote.” In MTV, Lauren, speaking about her friend Lo, said: “and she’s all ‘you’ll never guess who’s in the office.’” Her quotative (“she’s all”) helped establish not one, but two functional mappings:

Lauren as base producer → actor playing Lo

Lauren executing “You’ll never guess who’s in the office” → enactment of Lo saying “You’ll never guess who’s in the office”

Indexing. The function of an indexed depiction is specified by the phrase that indexes it. In Mozart, MB’s demonstration on the piano was indexed by “this,” the subject noun phrase of “NP is what I want to write.” In his lecture, MB placed his indexed depictions in various locations: (a) in a slot just before the current utterance; (b) in overlap with the indexing phrase, as in Mozart; (c) fused with the indexing phrase, as in Raised voice; and (d) in a slot just after the current utterance. That is, MB placed indexed depictions near, but not necessarily cotemporal with, the phrases that indexed them.

Adjoining. The function of an adjunct depiction is specified by the word or phrase it is timed to go with. Table 23 shows two examples. In Window, as analyzed by Kendon (2004, p. 117), a woman was about to say “window displays” when she started to produce a gesture to go with it. As Kendon argued, however, the woman had not planned the gesture by the end of “wonderful,” so she delayed “window displays” by 0.6 s so that she could perform the “stroke” of her gesture in overlap with it. In other examples, speakers delayed the gesture to overlap the yet-to-be-formulated affiliate (Kendon, 2004, pp. 134ff; see also Chu & Hagoort, 2014). A gesture should overlap its affiliate, and speakers can achieve that by delaying either the gesture or the affiliate.

In Running, a woman depicted the running speed of the gunmen she was describing by accelerating her speech within the words take off and running. That is, she fused her depictions with these two words to mark them as the affiliates (for other examples of fusion, see Perlman, Clark, & Johansson Falck, 2015; Shintel, Nusbaum, & Okrent, 2006).

Producers of a depiction, in brief, pay close attention not only to its spatialization, but to its timing. They choose the timing within a depiction for strategic reasons and use the timing of the depiction as a unit to specify its function in the discourse.

Displaced Worlds

In discourse, people engage in the actual world, but during each depiction, they also engage in a displaced world—a world displaced in location, time, or reality. In Schubert, the actual world contained the discourse scene (DN and MB discussing pianos), and within the depiction, the displaced world contained the distal scene (Enrico singing “Fremd bin ich ein”). Both worlds were larger than these scenes alone. Enrico did not stop singing after just four words, and he surely had a life both before and after his singing. Displaced worlds, like actual worlds, are assumed to be continuous and complete.

During a depiction, then, people jointly engage in two simultaneous layers of activity (Clark, 1996):

Layer 1: Engaging in the actual world (with its discourse scene)

Layer 2: Imagining the displaced world (with its distal scene)

Table 23

<table>
<thead>
<tr>
<th>Label</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Window</td>
<td>And they used to have wonderful (0.6 s silence, then outlining a window with two hands) window displays in there, didn’t they.</td>
</tr>
<tr>
<td>Running</td>
<td>[Woman was describing some gunmen she witnessed to a TV reporter.] I come out here to the steps [pause] [speeded pronunciation, hand moving quickly to her right] take off down that way . . . and then I see some people [speeded pronunciation, hand moving quickly to her left] running that way.</td>
</tr>
</tbody>
</table>
The participants in a discourse are always engaged in Layer 1. It is only during a depiction that they engage in Layer 2 as well. How they engage in Layer 2 depends on the displaced world.

Real and Fictional Worlds

In literature and film, works are either nonfiction or fiction. Histories contrast with novels, news stories with short stories, and documentaries with movie dramas. The two differ in their displaced worlds. Fictional worlds include elements that are not real. Nonfictional worlds reputedly do not.

Nonfiction also contrasts with fiction in the theater. In Shakespeare, the histories contrast with the comedies and tragedies, and in solo performances, actors doing Mark Twain or Abraham Lincoln contrast with satirists and imps. If so, nonfiction should also contrast with fiction in everyday discourse, and it does. In Schubert, the world of Enrico was real. In Go-to-bed, the world of Mother, Father, and Baby was fictional.

The difference is crucial. With nonfictional depictions, the recipient’s goal is to fit the distal scene to the real world. The direction of fit is depiction-to-world. With fictional depictions, in contrast, the goal is to create a world different from the real world, that fits the depiction. The direction of fit is world-to-depiction. Those of us at Godot created the world of Estragon and Vladimir as the play progressed. That world did not exist for us until it was created by the play. Still, evidence suggests that people assume that a displaced world is real unless it is marked otherwise.

Joint Access

The basic function of a nonfictional depiction is to establish joint access to a segment of the displaced world. In Schubert, DN drew MB’s attention to Enrico singing just four words (“Fremd bin ich ein”)—a brief segment of Enrico’s world. Other depictions access segments of past or future actions, events, and objects. Examples range from the quotation of Smith in Booming, the musical depiction in Dee-duh-dum, and the diminuendo sign in Tapering to the car noises in Worry-noise and Purring and the notches in the key in Grooves. Each of the worlds accessed is displaced in place or time, but not in reality.

The item depicted may be either a particular individual or event, or an example of a type of individual or event. The quotation in Tiffany’s represented what a particular speaker said on a particular occasion, but the quotation in Zchau was an example of the type of question Zchau was asked. The diminuendo sign in Tapering was also an example of a type, and so were the two measures depicted in Mozart. In full-scale demonstrations, tennis instructors are more likely to depict a type of serve than, say, Serena William’s last serve at Wimbledon in 2015.

Joint Pretense

Fictional depictions, in contrast, are designed to create joint pretenses. Children are explicit about this in make-believe play. In Go-to-bed, Sally and Alex would have said that they were just pretending (Garvey, 1990; Harris, 2000). As one two-and-a-half-year-old said, “I’m not realing. I’m pretending.” Joke tellers are usually just as explicit, as in this joke Sam told to Reynard (Svartvik & Quirk, 1980):

1. Let me tell you a story.
2. a girl went into a chemist’s shop and asked for contraceptives.
3. so he said “well I’ve got all kinds and all prices what do you want”
4. she said “well what have you got”
5. [proceeds with long joke]

In Line 1, Sam announced he was going to tell Reynard a “story.” and from Line 2 through the punch line, he enacted someone telling someone else in a fictional world about a girl who went into a chemist’s shop. The depicted worlds in both make-believe play and jokes are displaced in place, time, and reality.

In another class of depictions, however, the depicted worlds are displaced in reality, but not in time or space. One example is the tease recorded in Table 24. Conan O’Brien had just welcomed Chloe Sevigny as a guest on his late night TV show when, in Line 3, he appeared to assert that Will Ferrell, who was sitting next to Sevigny, “almost wore that outfit” but “decided that would not be cool.” He did not announce the pretense beforehand, but left Sevigny and his audience to discover it on their own. He enacted himself speaking to Sevigny at that moment, but in a fictional world in which Farrell “almost wore that outfit.” Ferrell recognized the pretense and added “Inches away,” and Sevigny showed her recognition by laughing (see Drew, 1987).

Certain pretenses, then, are designed to be discovered, and at least some are interpreted first as real (in Layer 1) and then reinterpreted as fictional (in Layer 2). Teasing belongs to a class of staged communicative acts that also includes verbal irony, hyperbole, understatement, rhetorical questions, banter, taunting, and ostensible speech acts (Clark, 1996, Chapter 12; Clark & Gerrig,

Table 24
A Joint Tease by O’Brien and Ferrell of Sevigny

<table>
<thead>
<tr>
<th>Line</th>
<th>Character</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.</td>
<td>Sevigny</td>
<td>Thank you.</td>
</tr>
<tr>
<td>3.</td>
<td>O’Brien</td>
<td>Will almost (two hands gesturing at dress) wore that outfit when he came here tonight.</td>
</tr>
<tr>
<td>4.</td>
<td>Ferrell</td>
<td>(leaning forward with thumb and finger inches apart) Inches away.</td>
</tr>
<tr>
<td>5.</td>
<td>Sevigny</td>
<td>[laughs]</td>
</tr>
<tr>
<td>6.</td>
<td>O’Brien</td>
<td>Then he decided that would not be cool.</td>
</tr>
</tbody>
</table>
1984; Gibbs, 2000; Isaacs & Clark, 1990; see also Clift, 1999). As Grice (1978) wrote, “To be ironical is, among other things, to pretend (as the etymology suggests), and while one wants the pretense to be recognized as such, to announce it as a pretense would spoil the effect” (p. 125).

Recursion

Layering is also needed to account for recursive depictions (see Clark, 1996). In Act 5 of A Midsummer Night’s Dream, Quince and five other characters put on a play called Pyramus and Thisby. The audience at Midsummer Night’s Dream would readily recognize: (a) that they themselves belonged to the actual world, (b) that Quince and his friends belonged to the Dream-world, and (c) that Pyramus, Thisby, and their friends belonged to the Pyramus-and-Thisby world. They would know they were seeing a fictional depiction of a fictional depiction.

If there is recursion in the theater, there could also be recursion in everyday depictions, and there is. Recall that in Bink-bink, MB told a woman:

All you have to do is say, “Look, there are four notes here, but it’s not really very natural to go ‘(thrusting R index finger four times in rhythm onto an imaginary piano keyboard) bink bink bink bink.’”

Starting with “look,” MB enacted the woman speaking, but starting with “bink,” he enacted the woman enacting a pianist playing four notes—a 2P actor depiction of a 2P actor depiction. And Sam’s joke became recursive when the fictional narrator quoted the chemist and the girl. Recipients had no trouble interpreting these recursions.

Conclusions

Depicting, I have argued, is a basic method of communication. It is on a par with describing and indicating, but it works by different principles. To describe something is to tell others about its properties—to represent it categorically. To indicate something (e.g., by pointing) is to locate it for others in space and time. To depict something, however, is to show others what it looks or sounds or feels like.

Depictions contrast most directly with descriptions. To highlight their differences, let me return to Falcons, TF’s report of WG telling him about shooting some falcons:

Version 1: “In LA, they would have shot those falcons.”

Version 2: “In LA, they would have—” He leveled a finger at some imaginary nestlings and made a gun-cocking sound.

Version 1 is what WG might have said. Version 2 is what he did say, with its embedded depiction. The question is this: How does the depiction in Version 2 differ from the description “shot those falcons” in Version 1?

Scenes

Pure descriptions are structured strings of arbitrary symbols, whereas depictions are physical scenes. In Version 1, WG produced a stream of sounds that was construed as the phrase shot those falcons. He used these words in turn to denote a composition of the categories “kill or wound with a bullet,” “distant in space,” and “diurnal bird of prey” for elements in the scene he was describing. In Version 2, in contrast, WG staged a scene by “leveling a finger at some imaginary nestlings and making a gun-cocking sound.” He intended the scene as a physical analog of the scene he was trying to depict. In depictions, space maps into space, time into time, actors into characters, and props into objects. In descriptions like “shot those falcons” there are no spaces, actors, or props to map.

Descriptions are formed by linearization, and depictions by spatialization. In Version 1 of Falcons, once WG settled on what he wanted to describe, he selected, ordered, and timed the words shot, those, and falcons according to the conventions of English. This is a process of linearization (Levelt, 1989). In Version 2, once WG settled on which elements to depict, he spatialized them in the discourse by selecting a placement, orientation, scale, granularity, and dimensionality for them. Spatializing a depiction and linearizing arbitrary symbols are fundamentally different processes.

Layers of Action

Descriptions are created in one layer of action, and depictions in two. As a result, descriptions have two active participant roles, and depictions have four.

In the primary layer of a discourse, every utterance has a speaker and at least one addressee. The speaker, equivalent to Goffman’s (1981) principal and Grice’s (1991) utterer, is the participant who means something by producing the current utterance. In both versions of Falcons, WG was the speaker, TG was the addressee, and WG’s goal was to jointly establish with TG what people in LA would have done to the falcons.

In a depiction, however, people engage in a second layer of activity as well, and there they take on additional roles—and viewpoints. In the depiction in Version 2, WG was also temporarily an actor (playing someone shooting falcons), and MB a spectator (playing someone covertly watching the shooting). In this layer, WG’s goal was for the two of them to jointly imagine a displaced world in which someone was shooting falcons. In prop depictions, the added roles are prop-manager and spectator (as in Tapering), and in joint depictions, they are actor and coactor (as in Manager). The roles of speaker and addressee originate in the discourse; the roles of actor, prop-manager, spectator, and coactor originate in the staging.

Imagined Worlds

Ordinarily people try to imagine a displaced world whether it is described or depicted. In Version 2 of Falcons, TF would have used WG’s enactment to imagine someone shooting falcons. In Version 1, he might have used “shot those falcons” to imagine a similar scene (see Clark & Van Der Wege, 2015; Gerrig, 1993; Zwaan, 2009, 2014). These are really two types of imagining.

Imagining a depicted scene is a percept-bound process. When WG “ leveled a finger at some imaginary nestlings and made a gun-cocking sound,” the scene TF imagined was bound by the depiction as he perceived it. It would have included: a rifle visually analogous to WG’s leveled finger; a gun-cocking sound auditorily analogous to WG’s tongue click; and a shooter aiming across at the falcons.
Imagining a scene from a description is, instead, a concept-bound process. For “shot those falcons,” the scene would have had to fit a composition of the categories “kill or wound with a bullet,” “distant in space,” and “diurnal bird of prey.” TF could not know whether the shot came from a pistol or rifle, or from above or below, so he would have filled in these details by guesswork. If the verb had been “done something,” it would all have been guesswork. Imagining distal scenes is an inherent part of interpreting depictions. It is not an inherent part of interpreting descriptions. Many descriptions, indeed, have no scenes to imagine.

### Composite Utterances

Everyday discourse is a mix of descriptions, depictions, and composites of the two. It is one thing to account for descriptions and depictions on their own and quite another to account for composites.

One issue is function. Embedded depictions are used as if they were nouns, verbs, noun phrases, or other linguistic units. In Version 2 of *Falcons*, the depiction of someone shooting falcons functioned as the verb V in “In LA, they would have V.” And yet an enactment of someone shooting falcons is manifestly not a verb. How do people infer word-like categories for depictions that allow them to function as nouns, verbs, adjectives, noun phrases, and adverbs?

Another issue is reference. In *Mozart*, MB told his audience “This is what I want to write” while playing two measures from Mozart. He used “this” to refer, not to features of his performance (the proximal scene), but to features of the notes he depicted (the distal scene). Later, when MB said “This is a passage from this sonata” and then played two measures from a Mozart sonata, he referred, not to his performance, but to the sonata represented in his performance. How do recipients identify these referents?

Still another issue is modification. Adjectival depictions are nonrestrictive modifiers of the words or phrases they are affiliated with. In *Tapering*, it was as if MB had said “what we would call the tapering, which looks like this” as he depicted the diminuendo sign. And in *Disgusting*, it was as if he had said “It couldn’t be that. I wouldn’t think so, which I consider like this” as he depicted himself in disgust. How do recipients interpret nonlinguistic depictions as modifiers of linguistic expressions?

Standard theories of language processing account for descriptions, but to be complete, they must also account for depictions. They must deal with scenes as well as symbol strings, layered as well as nonlayered actions, proximal and distal viewpoints as well as discourse viewpoints, and percept-bound as well as concept-bound imagination. And they must say how people combine describing and depicting in composites of the two.

### References


Received May 26, 2015
Revision received December 22, 2015
Accepted December 23, 2015