The role of schemas and scripts in pictorial narration

Abstract: The theoretical debate on the nature of narrative has been mainly concerned with literary narratives, whereas forms of non-literary and especially pictorial narrativity have been somewhat neglected. In this paper, however, I shall discuss narrativity specifically with regard to pictorial objects in order to clarify how pictorial storytelling may be based on the activation of mentally stored action and scene schemas. Approaches from cognitive psychology, such as the work of Schank, Roger C. & Robert P. Abelson. 1977. *Scripts, plans, goals and understanding*. Hillsdale, NJ: Lawrence Erlbaum; Mandler, Jean Matter. 1984. *Stories, scripts, and scenes: Aspects of schema theory*. London/Hillsdale, NJ: Lawrence Erlbaum; Schank, Roger C. 1995. *Tell me a story: Narrative and intelligence*. Evanston, IL: Northwestern University Press, suggest that cognition crucially depends on the storage and retrieval of *action scripts or schemata*, that is, narrative structures, which may occur at various levels of abstraction. These schemas incorporate generalized knowledge about event sequences, such as the order in which specific events will take place; causal, enabling, or conventionalized relations between these events, and what kind of events occur in certain action sequences. There also are *scene schemas* that are characterized by spatial rather than temporal relations. Further kinds of schemas seem also to play a decisive role. Drawing upon considerations from schema and script theory, I will focus on some concrete examples of pictorial narration, more specifically depictions of the *Annunciation of the Virgin Mary*, where narrative schema structures become involved and, indeed, the comprehensibility of the pictures as such presuppose mental script representations.

Keywords: Annunciation; cognitive psychology; narrativity; pictures; schemas; scripts
1 (Pictorial) narration and the indeterminacy of meaning

Theorists of narrative have traditionally been concerned with literary texts, where narrativity has been approached in a relatively analytical, extensional, and fixed way, distinguishing stories from non-narrative texts via a set of defining criteria, such as temporal sequentiality, emplotment, eventfulness, causality or causal agency, particularity (instead of generality), and so forth. Thus, rhetorical modes such as arguments, descriptions, expositions, or explanations may not properly be regarded as narratives, due to the fact that they seem to be static, atemporal, and non-sequential. The narratologist Gerald Prince, for example, proposed the following minimal condition for narrative: “the representation of at least two real or fictive events or situations in a time sequence, neither of which presupposes or entails the other” (Prince 1982: 4). However, among narratologists no consensus exists regarding the exact definition of narrative. Frequently, the overt representation of a sequence or series of events is considered essential (cf. Rudrum 2005: 203), but perhaps even the representation of only a single event, as suggested by Gérard Genette (1982), might be a sufficient criterion for a narrative, a question to which we will return later on.

Moreover, although most studies have focused on oral or textual narratives, there is no reason to believe that stories are limited to verbal language, i.e., a semiotic system built on conventional signs (symbols), grammar, and praxis. On the contrary, there are good reasons to assume that stories also may be expressed by other semiotic resources, such as pictures, i.e., iconic signs (icons) that are interpreted by way of their resemblance to what they depict. This capacity for conveying narratives also exists in media which do not always show any (clear) temporal division on the expression side, such as static pictorial representations. Still, art historians have rather scarcely discussed and employed narratological approaches, at least not explicitly, apart from some few exceptions, such as Mieke Bal (1985, 1991; cf. also Kemp 1996: 68). Semioticians, on the other hand, being concerned with any kind of sign processes, have examined narrative issues from an intermodal perspective. Some noteworthy attempts have been put forward by prominent semioticians, such as Roland Barthes (1990) and Algirdas Greimas (1983 [1966], 1987). It should be noted, though, that narratology has developed into a major field in its own right during the last few decades, and thus not necessarily has been (or has to be) studied within semiotic frameworks (cf. Chandler 2007: 114).

Be that as it may, many examples of pictorial storytelling can undoubtedly be found throughout history and across the world (e.g., in ancient Egypt,
Greece, the Renaissance, as well as in India, China, and South America; cf. Ranta 2011: footnote 2). At a first glance, we may distinguish between at least three forms of pictorial storytelling (cf. Ranta 2011, 2013; Ryan 2011: Section 3.4.1; Wolf 2002, 2005):

1. Serial pictures, i.e., multiple distinct images, each of which is showing a single scene or event, narratively linked with a fixed reading order, often horizontally or vertically arranged. A clear-cut example would be the triptych “Life of Christ and the Virgin Mary” (c. 1580) by Frans Franken the Elder, consisting of 16 panels in a narrative series (Figure 1). The second picture in the upper row shows the Annunciation of the Virgin Mary, a theme to which we shall return in Section 3.

2. Single pictures that show disparate events and persons in the same pictorial space, sometimes called “continuous narratives,” “simultaneous succession,” or “polyphase pictures.” As an illustrative example, we might here refer to the highly complex painting “Scenes from the Advent and Triumph of Christ” (c. 1480) by Hans Memling (Figure 2). This panoramic landscape painting, a so-called *Simultanbild*, consists of only one pictorial space, but at the same time also displays a sequence of 25 disparate scenes from the lives of Christ and Mary (for a detailed analysis, see Whitman Coleman 2013). One of the scenes, in the left background, shows the Annunciation of Mary, who, easily identifiable in her blue dress, appears several times in the landscape setting (Figure 3).

3. Last, we might mention single pictures where an entire story has been reduced to, or implied by, an isolated scene or event. These, obviously narratively demanding, images are sometimes referred to as “monophase” or “monoscenic” pictures (see Figures 6–10).

---

**Figure 1:** Frans Franken the Elder: “Life of Christ and the Virgin Mary” (c. 1580; cropped), Museum of Fine Arts, Ghent. (Photograph: https://commons.wikimedia.org/wiki/File:Frans_Francken_I_-_Life_of_Christ_and_the_Virgin_Mary.jpg).
Despite these frequently occurring forms of pictorial narrativity, one might still concur that storytelling is most favorably manifested by “genuine” temporal arts, such as poetry, drama, literature in general, and also motion pictures, which inherently have a sequential structure. Pictures seem to lack narrative precision compared to literary texts, e.g., in rendering the agents’ internal states and motivations, their hopes and fears, and so forth. Monophase images also seem only to indicate changes or successions of events, which the beholder has to infer, rather than explicitly expressing these changes. The representation of causal or teleological relations seems to be particularly problematic for such images. It could
further be argued that pictorial media only to a very limited extent can display so-called “disnarrated” elements, such as detailed alternative courses of action or the agents’ unfulfilled intentions (see Wolf 2005: 434). Especially monoscopic works are thus narratively challenging, because they have to compress or refer to an entire story arc by displaying only one single scene.

Among art historians, however, the narrative potential of monophase pictures has usually been taken for granted, frequently from a descriptive, interpretative, and historical perspective. A more outspoken semiotic account on pictorial narrativity had already been put forward in the eighteenth century by Gotthold Ephraim Lessing (1957 [1766]), who attempted to outline the characteristic features of painting vs. poetry qua signs. According to Lessing, painting essentially is a spatial art concerned with the rendering of bodies, while poetry is an art of time, the latter being privileged in narrating actions, that is, the succession of events in time: “Objects which exist side by side … are called bodies. Consequently bodies with their visible properties are the peculiar subjects of painting. Objects which succeed each other … are actions. Consequently actions are the peculiar subjects of poetry” (Lessing 1957 [1766]: 91). Painting is thus basically descriptive, while poetry is more disposed to tell stories. The former can only represent momentary events, or slices in time. It has been argued, though, that Lessing’s view on perception was somewhat narrow-minded (Wollheim 1998: 223; see also Nanay 2009: 122). If we in real life can see temporally extended actions (due to processes of retention and protention, in Husserl’s terminology) and not only time-slices, it seems unreasonable to claim that our perception of depicted actions would be radically different in that respect. So pictures, it seems, do not have to depict two or more events in order to represent an action; at least in some cases only the depiction of one single event might be sufficient. Thus it would be odd to say that we see an image of a man in the air and imagine him landing on the ground, rather than saying that we actually see a man jumping in the image, although only a temporal limited part is visible (cf. Nanay 2009: 122).

Lessing was of course aware that painting, though to a lesser extent, also can display narratives, namely, as he claimed, indirectly by depicting the most fruitful or “pregnant moment” in a presumed story arc. An exemplary case discussed by Lessing would be the famous Hellenistic sculpture group “Laocoön and his Two Sons” (probably first century AD). A beholder familiar with the myth in question might very well see this sculpture as a significant part of a story about the priest Laocoön and his tragic destiny. On Lessing’s account, then, the representation/perception of actions in painting is not impossible per se; rather the reconstruction of actions demands more active efforts or is less “convenient” in painting compared to poetry.
But even without any specific knowledge of the myth, viewers would tend to interpret the scene as a violent struggle, filled with agony, between defending humans and combative serpents. Quite spontaneously, one might be tempted to reconstruct a minimal, although imprecise, narrative from the sculpture, simply due to our lifeworld experience (e.g., our knowledge about the danger which wild animals, such as serpents, may constitute) and the ability to decode pictorial representations (e.g., of body language and facial expressions) as such. So even without any literary scaffolding, pictorial renderings of actions can be interpreted as narratives, where the beholders have to fill in detailed gaps as a result of their background knowledge (cf. also Gombrich 1977 [1960]: 181). We may note, though, that also literary works involve gaps, presupposed but unstated facts or assumptions, presumed causal relations, points of indeterminacy, and so on; all of these voids likewise require interpreters to fill in or complete the structures at issue. Indeed, all texts demand active efforts of meaning or coherence making. Hence the contrast between the narrative-supporting or -enabling potential of verbal language versus that of static pictures would rather be a question of degree than a question of kind (cf. Wolf 2002).

On a basic structural level, as suggested by Nicholas Rescher (Rescher 1966; cf. also Herman 2002: 62), any action description ideally involves the following constituents:

1. **The Agent** (WHO did it)
2. **The Act-type** (WHAT did he or she do?)
3. **The Modality of Action** (HOW did he or she do it?)
4. **The Setting of Action** (IN WHAT CONTEXT did he or she do it?)
5. **The Rationale of Action** (WHY did he or she do it?)

All these action characteristics can of course be specified further, e.g., as to the means involved in the action, the temporal or spatial aspects of the setting, or the intentions or causes leading to the action. Actually, an action can be described in countless details, perhaps even indefinitely. And all narrative texts are inevitably incomplete and cannot specify or characterize every facet of a storyworld. Take, for example, the following sentences:

(i) “Finally, it had stopped raining, and Lì Wang decided to walk his dog near the lotus pond.”

(ii) “When driving his bicycle on the Champs-Élysées, the postman lifted his arm to signal his approaching right turn.”

As to sentence (i), readers would spontaneously assume that it had been raining for a longer period of time, thus restraining the agent’s outdoor activities, that the agent intends to move his feet and look for bushes or the like, will hold a leash,
maybe has a Chinese background, and so on. The action described in sentence (ii) could be presumed to take place in the traffic-dense urban environment of Paris, where an employee of a mail delivery service carries out his duties by using a pedal-driven vehicle with two wheels and taking safety precautions while changing his direction. But these action specifications are by no means exhaustive. In the first sentence, we also take for granted that rain consists of water droplets falling under gravity, the exposure to which may feel uncomfortable, that the agent repeatedly raises his left and then his right foot about 10 cm over the ground while swinging them forward, and that mammals have to urinate regularly. The action of the second sentence could also more extensively describe the agent’s movements as the result of certain electric impulses traveling from his brain to his limbs, that the location of his arm shifted from 1.5 to 1.7 m over the street level, that he was breathing seven times during that time, and so on (cf. Herman 2002: 69).

In all those (actual or possible) respects, then, action descriptions (or narratives) are unavoidably underspecified, whether verbal or pictorial. Even the most prolif-ically detailed representations of actions are bound to be incomplete, and over-specified descriptions might also easily be experienced as containing superfluous and narratively disturbing elements. Text interpretations require in general creative processes of supplementing missing or implied information (available due to the readers’ background knowledge as well as suggested by the text itself; cf. Herman 2002: 68), thus, e.g., complementing canonical action slots as listed by Rescher.

Here we might point to the pioneering motion studies by the photographer Eadweard Muybridge (1830–1904). Using multiple cameras, he captured motion in stop-motion photographs of a number of subjects, such as humans running, jumping, walking up- or downstairs, throwing discs, as well as running or walking animals, e.g., bison, elephants, or horses. Especially his photo series of galloping horses gave him public recognition. As the human eye cannot break down the fast action segments of galloping horses, most artists had traditionally painted them at a trot with always at least one foot on the ground, or at a full gallop with both front and hind legs stretched outwards (cf., for example, Théodore Géricault’s “The 1821 Derby at Epsom,” 1821). Muybridge, however, was able to prove that neither was the case; for an instant during a gallop, all four hooves are off the ground, though collected beneath the body (Figure 4).

It has been argued, that Muybridge’s studies have changed subsequent practices of depicting galloping horses. But his photos of action sequences also suggest general insights how pictorial representations of action sequences are influenced by mental schematizations of them; in this case incorrect ones. Moreover, they also illustrate how missing segments in such sequences are spontaneously filled in by the beholder, so that extended serial depictions of actions, as in Muybridge’s photo series, usually would be experienced as superfluous and irrelevant.
To sum up, then, we may note that gap-filling in narrative comprehension occurs in both verbal and pictorial media, although the latter may require more elaborate and active interpretation efforts compared to the former (cf. Wolf 2002). Even so, pictures can also have narrative strengths: they can provide us with more vivid and palpable impressions of spatial settings, some kind of “knowledge how” rather than “knowledge that” of the visual layout of storyworlds. Further, they can directly display the agents’ emotional expressions and their body language, as well as their visual appearance in general, giving them perceptual “flesh and blood,” so to speak, and thereby enhancing their noteworthiness or “tellability” (cf. Ranta 2019; Ryan 2011).

2 Schemata, scripts, and narrative comprehension

Activities of gap-filling and inevitable “spots of indeterminacy” in literary fiction had already been discussed by literary scholars, such as Roman Ingarden (1973 [1931]) and Wolfgang Iser (1978 [1976]). Apart from these theorists, also researchers within Artificial Intelligence and cognitive psychology have investigated similar aspects involved in perceptual and cognitive processing in general.
In cognitive science, the notion of so-called schemata has been employed to describe general cognitive structures and patterns of behavior organizing our experiences and social interaction. It has long been claimed that prior knowledge, and schematic cognitive representations in particular, influence memory formation and retrieval, and that these in turn have been constituted by personal experiences. Within philosophy, Immanuel Kant seems to have at least anticipated the notion of schemata, and in psychology, Frederick Bartlett referred to schemata in order to explain constructive and altering memory processes: we remember rather the *gist* of a story rather than its details (Bartlett 1932; Kant 1963 [1787]).

Schema theoretic considerations have also more recently been put forward concerning the nature of pictorial representation. During the last decades, the common-sense view, according to which pictures basically just (ought to) reproduce or imitate an external reality by means of visual similarity, has been refuted as untenable by numerous scholars (Black 1972; Bryson 1983; Eco 1979 [1976]; Goodman 1976; Walton 1993). However, historically seen, such imitation or mimesis views, conceived as straightforward copy theories, have not been as prominent as some of its critics seem to have claimed. Mimesis has also frequently been considered to be the imitation of universals, abstractions, essences, or types, or as rendering certain idealizations (e.g., in terms of morality or beauty; see Ranta 2000).

The art historian Ernst Gombrich, in his influential work *Art and Illusion* (first published 1960), was also a well-known critic of such simple copy theories. According to Gombrich, neither artists nor beholders have an “innocent eye”: there are historically/culturally variable belief systems and conventions which direct our way of seeing and depicting something. Influenced by Karl Popper’s view on scientific progress as a continuous process of conjecture and refutation, Gombrich proposed a corresponding model for the development of visual representation, i.e., as an interplay of “schemata and corrections” or “making and matching.” In his view, schemata might roughly be described as models for depiction and visual vocabularies, but also as perceptual and cognitive expectations. Such schemata are supposed to precede corrections: artists have some kind of preliminary pictorial/perceptual models of what they intend to depict. Visual stimuli are interpreted according to these models and transferred into pictorial representations. Subsequently, these representations are then compared with the depicted objects in question, after which imperfections of the picture and its underlying schemata may become apparent (Gombrich 1977 [1960]: 24, 63–64, 99). This interaction of making and matching is thus a historical process explaining the development and change of pictorial styles (ultimately striving for faithfulness or naturalism).
However, what are schemata more exactly? In Gombrich’s account, clear definitions are largely absent; instead, circumscriptions, synonymous expressions, and examples prevail (cf. Wollheim 1974: 286–289; for a fuller discussion, see Ranta 2000: 77–82). Furthermore, conceptions of art as progressing towards increased realism have not been uncommon in traditional art history (cf. general textbooks such as Gombrich 2006 [1950]; Honour and Fleming 2013 [1982]; Janson 1991 [1962]). More specifically, such accounts tend to focus upon the representation of (rather static) objects or of space/perspective, while a possible progress regarding the depiction of actions and narratives has not been as thoroughly considered.

Gombrich developed his ideas in the 1950s and 1960s, but since then, not surprisingly, research on schema theory has evolved considerably. Within cognitive psychology, stringent theoretical models and experimental procedures have been elaborated for clarifying the cognitive processes underlying categorization and schema formation (for a detailed account, see Ranta 2000: 175–219). Basically, schemas have been outlined as general, higher-level mental constructs that incorporate representations of similarities or commonalities across events or objects, rather than the specificity that make those events or objects unique. A number of experimental findings suggest that perception and cognitive activities are hierarchically structured (for some overviews, see, e.g., Ghosh and Gilboa 2014; Gilboa and Marlatte 2017; Nishida 1999; Rumelhart 1980). New information is compared with and assimilated into broader mental schemata or categories which are necessary for object recognition, explanations, predictions, and communicative activities. This means that we are able to generate mental representations which have something like a type-character, stored in long-term memory with which external objects are compared. Essential for schemas are that they are based upon multiple episodes or encounters with specific objects or situations, that they lack unit detail or specificity, and that they are flexible. They must be general in order to organize information, but also capable of incorporating details of new particular episodes. This alterability of schemas may, using Jean Piaget’s notions, be referred to as accommodation and assimilation. Encounters with objects or events that do not fit in with one’s schematic representations necessitate modifications of (or adaptations to) the latter. While assimilation just requires incorporating elements within a schema without challenging it as such, accommodation refers to the modification of a schema, when too deviating elements are too incompatible with the existent one (cf. Ghosh and Gilboa 2014: 108; Gilboa and Marlatte 2017: 619).

Not only objects, but also events and actions are subject for schematic processing. Event is here referred to as an overarching concept, which not only includes goal-related actions performed intentionally by conscious agents, but
perhaps also inanimate or natural phenomena, such as earthquakes or thunderstorms. In general, any representation of a change of state, explicitly or implicitly, might be called a representation of an event. Thus events are temporally extended and may be chronologically ordered (for a more detailed account on events, see Zacks and Tversky 2001). Schematizations of action sequences and their constituents have been called scenarios, frames, scripts, event schemas, or plans within cognitive psychology (see, e.g., Brewer 1987; Emmott and Alexander 2014; Schank and Abelson 1977). For example, a script is a temporally structured schema, based on knowledge of stereotypical goal-oriented event sequences “that define a well-known situation” (Schank and Abelson 1977: 422). This includes general internalized, experientially based knowledge about the order in which specific events will take place; causal, enabling, or conventionalized relations between these events; and what kind of events are common in these action sequences (see, e.g., G. Mandler 1984). Moreover, there are also scene schemas, which are rather constituted by spatial than temporal relations. These may include mentally stored inventory information, that is, what kinds of objects mostly appear in such situations, and relational information regarding the normal spatial layout of a scene (J. Mandler 1984: 13–17). For example, we have certain expectations concerning the outlook of environments such as rooms, streets, and buildings, where certain activities, such as going to a wedding or going to a dentist, take place.

Numerous experimental studies have investigated the formation and structure of such action schemas or scripts. Based on these findings, Roger Schank and Robert Abelson argued that our knowledge is usually organized around a large number of stereotypical situations including more or less routine activities (Schank and Abelson 1977). We acquire these scripts, along with culture-specific and idiosyncratic variations, through previous direct or indirect experiences of and about various situations. For example, a series of experiments by Gordon Bower, John Black, and Terrence Turner demonstrated that people usually agreed on the constituents and order of actions in routine activities like eating in a restaurant and visiting a dentist. Moreover, when recalling texts narrating script actions, the subjects tended to confuse actions that were presented with unspoken ones implied by the script. Subjects also used to recollect script actions in their common or familiar order; scrambled texts describing script actions out of order were usually recalled according to their canonical order (see, e.g., Bower et al. 1979).

Schank and Abelson’s well-know restaurant script, for example, would include knowledge of the action sequence of “entering,” “ordering,” “eating,” “paying bills,” and “leaving.” In addition to such sequences of events, there may be further “slots” within the script to describe typical roles (customers and waiters, etc.), inventory items (menu, table, food, bill, money, etc.), entry conditions (hungry customer, restaurant provides food, etc.), and results (customer is not hungry,
restaurant has less food, etc.; Schank and Abelson 1977). According to Schank, who elaborated his early work on action schemas, human intelligence itself seems to a considerable extent presuppose the formation and retrieval of scripts (Schank 1995, 1999). Moreover, such schemas may occur on various levels of abstraction, as we shall see below; even the identity of cultures and sub-cultures is basically shaped by shared low- and high-level narrative structures. Needless to say that many scripts are not universal, but frequently culture-specific or even idiosyncratic (regarding, for example, our individual morning routine habits). Some social scripts differ slightly, whereas in other cultural contexts differences may be more profound. For example, Western scripts for dining in a restaurant may differ, more or less, from comparable scripts in China (cf. Meng 2008).

When social interactions and communication with members of the same culture occur in certain situations a number of times, schemas are generated and internalized, becoming successively more abstract, organized, and stable, and thus more usable as efficient means for interaction. Based upon findings from cognitive science, it has been suggested that we may distinguish between at least the following “primary types for generating human behavior for social interactions” (Nishida 1999: 757):

1. **Fact-and-concept schemas** (cf. “knowing that”): pieces of general information concerning facts such as “Berlin is the capital of Japan,” and concepts such as “A train is a connected line of railroad cars with or without a locomotive.”
2. **Person schemas**: knowledge about different types of people, including their (dominant) personality traits (e.g., “Mary is humble”), sometimes also being prejudicial.
3. **Self schemas**: people’s cognitive generalizations about their selves, derived from past experience, not least social ones.
4. **Role schemas**: knowledge about social roles which denote behavioral patterns that are expected of people in particular social positions.
5. **Context schemas**: information about the situation and usual setting of behavioral frameworks, such as going to a restaurant, which may include guidelines about appropriate actions in order to achieve goals in the context (i.e., to activate procedure and strategy schemas).
6. **Procedure schemas** (cf. “knowing how”): knowledge about the appropriate sequences, specific steps to take, and behavioral rules in common action situations, derived from previous successful practice, or from indirect “experience” transmitted through instructions, narratives, or societal and other conventions.
7. **Strategy schemas**: knowledge about problem-solving strategies, such as hypothetico-deductive reasoning, which also may vary widely due to the agents’ past experiences.
8. *Emotion schemas*: information about affect and evaluation stored in long-term memory which is accessed when other schemas are activated. These schemas are constructed in social interactions throughout one’s life (see Nishida 1999: 757–759).

Repetitions of schema-based behavior will more likely lead to the storage and enhancement of the schemas in one’s long-term memory. Once these schemas have been established, the cognitive processing of new information is directed by schema-driven or by top-down processes (rather than data-driven or bottom-up processing). Ambiguous information may either lead to the search for completing data (i.e., data-driven processing) or by using “default options” provided by the schemas at hand (i.e., schema-driven processing). The latter, which demands less time and effort, seems often to be preferred when the prevalent data are schematically salient, relatively unambiguous, or less significant for persons (Nishida 1999: 766). We may also note that retrieving information from memory involves the activation of various levels of the network of schemas, spreading from one schema to related schemas. Fact-and-concept, person, self, role, context, procedure, strategy, and emotion schemas mentioned above are interconnected with each other, forming a network of schemas which generate behavior. This means that if one part of the schema network changes, also other parts of the whole system, and thus our behavior, become affected (Nishida 1999: 765). All these findings from schema theory may, as we shall see in the next sections, help us understand the role of schemas and scripts for comprehending pictorial narratives.

Before we shall look at some examples of the latter, there is one further important aspect of schemas which should be considered, namely, that they are hierarchically structured. This does not only apply to objects, but to events and actions as well. Objects may be structured according to their parts and subparts, due to their spatial configurations. For example, a car consists of parts such as doors, wheels, seats, etc., which in turn consist of various subparts, which can be divided into further subparts. Such an hierarchical structure would be a *partonomy*. Further, objects may also be structured according to their *taxonomy*, that is, which kinds, types, or categories they belong to. Findings from cognitive psychology suggest that we here may differentiate between at least three levels of abstraction (i.e., the degree of inclusiveness of a category). According to, e.g., Eleanor Rosch, at the most salient, so-called *basic* level, objects – both biological entities and artefacts – are most commonly divided into categories (cf. Rosch and Mervis 1975: 586). Besides the basic level, Rosch’s research suggests that we may differentiate between at least two other levels of abstraction, namely, *superordinate* and *subordinate* levels. For example, “furniture” might count as a superordinate category, “chair” as a basic level one, and “kitchen chair” or “living-room
chair” as subordinates. The basic level seems to be psychologically different from superordinates and subordinates in several ways. In contradistinction to members on a superordinate level, which have relatively few cognitively or perceptually salient characteristics in common (according to studies where subjects had to list attributes of the objects), basic level members are regarded as resembling each other to a higher degree (i.e., they have more attributes in common, such as perceived shapes), and they are most quickly identified by subjects as belonging to a certain category (Rosch and Lloyd 1978; Rosch and Mervis 1975; cf. also Ranta 2000: 187–199; Zacks and Tversky 2001: 5).

Like objects, also events and actions may be organized into partonomic and taxonomic hierarchies. Studies from cognitive psychology suggest that within relatively homogeneous groups of participants, the identification of typical parts of everyday activities are usually agreed upon. Moreover, it may be argued that also event partonomies include distinctive, cognitively salient constituents. Such so-called “behavior episodes” are analogous to physical objects which can be seen with the naked eye” (Barker and Wright 1954: 6, my italics; quoted in Zacks and Tversky 2001: 6).

The boundaries of behavior episodes may be delineated by the following characteristics:
1. A change in the “sphere” of the behavior between verbal, social, and intellectual.
2. A change in the predominant part of the body.
3. A change in the physical direction of the behavior.
4. A change in the object of the behavior.
5. A change in the behavior setting.
6. A change in the tempo of the activity.
(Barker and Wright 1954: 6; quoted in Zacks and Tversky 2001: 6)

Also certain scene levels in familiar scripts are commonly regarded as privileged, easy identifiable event segments, such as the restaurant script mentioned earlier with constituents such as “entering,” “ordering,” “eating,” and “exiting” (cf. Zacks and Tversky 2001: 6). Events may furthermore be organized as taxonomic hierarchies, consisting of subordinate, basic level, and superordinate categories, where basic level categories have a privileged and cognitively salient status similar to those in object taxonomies. Empirical findings also suggest that participants, for example, tend to name simple stories using basic-level names (Morris and Murphy 1990; cf. Zacks and Tversky 2001: 6).

In some respects, then, objects and events both possess partonomic and taxonomic features, although it should be emphasized that events are necessarily transitory, whereas objects are temporally relatively stable. Part boundaries of
objects are spatial in nature, while event boundaries occur chiefly in time. Still, also events may be perceived as partonomic structures, and the structures and the segmentation of events are partly influenced by the physical features of activities, but also driven by top-down processing, such as people’s experiences and expectations (cf. Zacks and Tversky 2001: 9). Repeated exposures to events and actions, either directly or through descriptions or narrative mediations, lead to the creation of schemata and anticipations concerning ongoing activities, in an interplay of bottom-up and top-bottom processing. Moreover, the various parts of actions are also identified and interpreted in relation to assumed goals of agents. However, the time scales within which people notice behavior parts are usually limited; actions related to minor subgoals as well as those related to larger ongoing goals are usually not preceived as salient. For example, the breathing of the postman cycling on the Champs-Élysées as well as his general endeavor for enduring economic security by carrying out his profession would hardly be noted, while delivering mail would more likely be recognized as an immediate and salient goal-related activity. In general, event schemata as here described provide frameworks for understanding ongoing activities and relating them to goals, and action units which are more immediately relevant within these schemata are also easier remembered and recognizable than irrelevant ones (cf. Zacks and Tversky 2001: 7; 13–15).

3 A case example: the Annunciation

As we have seen in the previous sections, schemata and their constituents as here outlined seem to play a crucial role for text and story comprehension, and understanding human behavior in general. Texts can hardly contain all details required to be fully understood, and readers’ various types of schemata compensate for inevitable information gaps. Quite obviously, this applies, and even more so, to pictorial renderings and narratives, which due to their media-specific properties demand considerable interpretative work. Thus beholders inevitably have to interpret, e.g., pictorial works of art according to acquired category systems, schemata, and habits to which the work has been adjusted. Moreover, the production of these works is in turn also guided by the concerns and needs of a certain public. The artist responds to these demands and offers opportunities for the beholder to apply his background experience of his “way of life” and of widespread narratives, as well as of current artistic-stylistic conventions. The recognition of familiar items or themes, the experience of something as typical in some sense, may even give the beholder an enhanced feeling of satisfaction or enjoyment (see also Section 4).
Indeed, the production of visual works of art or pictures highly interacts with the requirements of a certain public. Both storyteller and recipient share numerous unstated lifeworld assumptions concerning causal relationships, concepts of class, gender, age, social roles, general cultural knowledge, acquaintance with various action scripts, and so on. Moreover, narratives also contain numerous normative assumptions or claims (e.g., ethical, political, or epistemic ones). The beholder interprets a work of art (within given historical, social, institutional, ideological, and/or religious contexts) in accordance with this background knowledge which the work has been adapted to, and completes or “fills in” narrative gaps with unstated, but presupposed connecting and sense-making details. The generation of pictorial narrativity as such an interactive relationship could perhaps be schematized as follows (Figure 5).

Similar ideas have also been suggested in Michael Baxandall’s acclaimed work “Painting and Experience in Fifteenth-Century Italy” (first published in 1972). As he states, numerous paintings during that period had to represent biblical stories with the purpose of deepening religious beliefs and sentiments. With pictures as external scaffoldings, beholders were given the opportunity to meditate on particular themes or stories. Thus paintings functioned as some kind of condensed reminders of religious stories which were expanded and “filled in” by an individual’s private and more particularized meditation (or something like preconceived mental images). Hence, paintings tended to represent, despite their seeming richness of details, rather general types of situations, places, and people which were supplied with even further details by the beholders’ imagination (Baxandall 1972: 45–48). Instructions for such practices were provided by, for example, public sermons and prayer books, such as the “Zardoine de Oration” (Garden of Prayer), written for young female readers in 1454. Here the worshipers

![Figure 5: Pictorial Narrativity as the Interaction between Producer and Recipient (modified adaptation of a communication model in Sonesson 1999: 96).](image-url)
are advised to mentally picture sequences of sacred moments, in order to build up feelings of piety and compassion:

Alone and solitary, excluding every external thought from your mind, start thinking of the beginning of the Passion, starting with how Jesus entered Jerusalem on the ass. Moving slowly from episode to episode, meditate on each one, dwelling on each single stage and step of the story. And if at any point you feel a sensation of piety, stop: do not pass on as long as that sweet and devout sentiment lasts. (Baxandall 1972: 46)

Exercises of religious contemplation as here described would then contribute to a more detailed comprehension and interpretation of paintings, onto which private mental visualizations were projected. Still, this nuanced ability to interpret pictures was not necessarily the result of a higher education or of having read written tutorials, but also became widespread and fostered through popular sermons performed by charismatic preachers. These orators, such as Fra Roberto Caracciolo da Lecce, repetitively “drilled their congregations in a set of interpretative skills right at the centre of the fifteenth century response to paintings” (Baxandall 1972: 48).

A concrete example, which Baxandall discusses at some length, concerns one particular event in the life of Christ, namely, the Annunciation of the Virgin Mary. In the Bible, the Annunciation is described in Luke 1:26–38, that is, the scene where the angel Gabriel announces to Mary that she will conceive and give birth to the son of God. This theme has become one of the most frequent subjects in Christian art, in particular during the Middle Ages and the Renaissance. Noteworthy for our current discussion, however, is how, and which of, the action constituents of this narrative have been pictorially rendered. According to Baxandall, beholders in Renaissance Italy would have been very sensitive to and aware of subtle differences in the illustrations of biblical stories, and certainly so compared to viewers today (Baxandall 1972: 36–37). Instructive in this respect were, for example, Fra Roberto’s sermons, where he discerns between three general “mysteries,” or basic action constituents, of the Annunciation: (i) the Angelic Mission; (ii) the Angelic Salutation; and (iii) the Angelic Colloquy. Especially the last part becomes a focus of interest in pictorial art, where, further, one may distinguish between “five successive spiritual and mental conditions or states attributable to Mary” (or, as phrased by Fra Roberto, “Laudable Conditions”): “disquiet” (Conturbatio), “reflection” (Cogitatio), “inquiry” (Interrogatio), “submission” (Humiliatio), and “merit” (Meritatio; Baxandall 1972: 49–55). Seen as an action sequence, however, it would rather be appropriate to describe these conditions as behavior episodes within an action script or narrative, which contemporary beholders were very well acquainted with, through direct Bible

1 For an interesting discussion of the Annunciation theme and Baxandall’s analysis from a feminist perspective, see Carvalho 2015.
readings and, probably more commonly, through public sermons as Fra Roberto’s as well as general “discursive formations” (as Foucault could have put it).

The annunciation, in itself being a crucial part, among others, of the lengthy narrative “Life of Christ,” can then be analyzed as part of an action sequence with a partonomic structure. We may then, as explicated by Fra Roberto, discern between these five partonomic segments of the Angelic Colloquy sequence.

3.1 The segment Conturbatio – disquiet

At this first stage, Gabriel arrives and speaks: “‘Hail, thou that art highly favored, the Lord is with thee: blessed art thou among women’ – she was troubled” (Bax-andall 1972: 51). It is not the appearance of the angel as such which astonishes Mary (who previously became acquainted with such apparitions), but rather the gravity and tone of the message. This event, identifiable by the hands with palms turned outward and the somewhat evading posture of Mary, has been one of the most depicted Annunciation scenes. Examples illustrating this stage would, for example, be Filippo Lippi’s “The Annunciation in Florence” (1440–60), Simone Martini’s “The Annunciation with Two Saints” (c. 1333), and Sandro Botticelli’s “The Cestello Annunciation” (1489; Figure 6).

Figure 6: Sandro Botticelli: “The Cestello Annunciation” (1489), Uffizi, Florence. (Photograph: https://commons.wikimedia.org/wiki/File:Botticelli,_annunciazione_di_cestello_02.jpg).
3.2 The segment *Cogitatio* – reflection

At the second stage, likewise identifiable by Mary’s distinctive gestures (one hand on the breast, the other open) and the face seemingly deep in thought, she is wondering what kind of greeting this might be. This scene can, e.g., be seen in Joos van Cleve’s “Annunciation” (c. 1525) or Fra Carnevale’s “The Annunciation” (c. 1448; Figure 7).

3.3 The segment *Interrogatio* – inquiry

After the initial salutation, Gabriel announces to Mary that she will give birth to God’s son, after which she inquires how that, she being a virgin, could be possible. Also here characteristic gestures, such as Mary’s hand raised as if querying and Gabriel appearing to be in listening mode, help to identify the segment. Illustrations of this (less frequently occurring) scene are Filippo Lippi’s “Annunciation”

*Figure 7:* Fra Carnevale: “The Annunciation” (c. 1448), National Gallery of Art, Washington. (Photograph: https://commons.wikimedia.org/wiki/File:Fra_Carnevale_-_The_Annunciation_-_WGA04251.jpg).
3.4 The segment *Humiliatio* – submission

The angel then explains that the Holy Spirit will come on her, so that she will conceive the son of God. The following depicted stage shows Mary’s humble acceptance of the situation, and she speaks. “Behold the handmaid of the Lord … Be it unto me according to thy word” (Baxandall 1972: 51–55). Also here Mary’s gestures, crossing her hands across her breast and showing humility, help the beholder to discern the particular episode of the script. As examples we might mention a panel from Hubert and Jan van Eyck’s “Ghent Altarpiece” (1430–32), (1437–39), Vittore Carpaccio’s “Annunciation” (1504), or Alesso Baldovinetti’s “Annunciation” (1447; Figure 8).
Gerard David’s “Annunciation” (1506), or Fra Angelico’s “Annunciation” (c. 1437–46; Figure 9).

3.5 The segment *Meritatio* – merit

The last stage, then, shows Mary on her own, apparently in a state of deep thought, after the departure of Gabriel, and having conceived God’s son. Strictly speaking, this is not part of the biblical narrative, but emerged as some kind of pictorial appendix, also called *Annunziata* (i.e., depicting Mary after the “proper” Annunciation sequence). Examples would be Fra Angelico “Virgin Annunciate” (c. 1450–55) or Antonello da Messina’s “Virgin Annunciate” (c. 1473; Figure 10).

**Figure 9:** Fra Angelico: “Annunciation (cell 3)” (c. 1437–46), Museum of San Marco, Florence. (Photograph: https://commons.wikimedia.org/wiki/File:Fra_Angelico_-_Annunciation_ (Cell_3).-_WGA00538.jpg).
Depictions of the Annunciation as here presented were thus common during the Italian renaissance. Sometimes painters tended to focus upon and stick to particular stages of the Annunciation script (e.g., Fra Angelico almost exclusively to *Humiliatio*, and Botticelli to *Conturbatio*). One should also keep in mind that these episodes were not always as clearly segmented, as here described; sometimes distinct stages were combined or somewhat blurred. Moreover, the popularity of certain stages also varied; during the fifteenth century, for example, images of *disquiet* and *submission* were more prevalent, while images of *reflection* and *inquiry* occurred more frequently in the fourteenth century (Baxandall 1972: 55). Baxandall’s study mainly concerns pictorial representation used during a specific period, where after such a careful segmentation of the Annunciation subsequently decreased. However, these images, being remarkably popular at that time, may exemplify how prevailing, although fictitious, action schemas could enable a profound understanding of pictorial narratives and directed the production of those. Moreover, beholders were also able to differentiate between a variety of facial expressions and gestures, due to their acquaintance with the practices of orators and preachers, that is, “skilled visual performers with a codified range of gesticulation not special to Italy. An Italian preacher could tour northern Europe successively preaching even in places like Brittany …” (Baxandall 1972: 64).
4 Concluding remarks: schema deviations and tellability

The Annunciation series as here described is peculiar to a specific region and period in time; still, it may well illustrate how in general scene and script schemas, among others, influence the production and reception of narrative imagery. Fact-and-concept, person, self, role, context, procedure, strategy, and emotion schemas mentioned earlier are likewise crucial aspects, which not only direct our interpretations of pictures, but also further interactions with them (such as in our case here, steered by devote or pious Christian believers’ self and role schemas). Apart from this case, it would certainly be possible to find equivalent pictorial examples not only referring to other Biblical stories, but also to the lives of Buddha, ancient Assyrian, Egypt, Greek, Roman, and Indian gods, heroes, and rulers, just to mention some examples from antiquity (and more contemporary ones are easily conceivable). The comprehension of such narrative representations may vary more or less, depending on how fine-grained the action knowledge of the producers and beholders might be. As to the Annunciation, beholders at that time might have had a very detailed knowledge regarding the meaning of gestures, not only Mary’s. We may find some clues for interpretation through, for example, preserved instructions for Benedictine monks using sign language during periods of silence or the use of gestures by preachers. Thus grief could be shown by “pressing the breast with the palm of the hand” or shame by “covering the eyes with the fingers” (Baxandall 1972: 61–71). And slight deviations from an expected use of gestures displayed by images in certain contexts would have been noticed by contemporary beholders, with curiosity or astonishment (though not necessarily with approval).

This leads us to another aspect of schemata and their relevance for comprehending pictorial narratives, or more specifically the issue of “tellability,” namely, what makes some stories more worth telling or relevant than others. Based on studies of conversational storytelling, e.g., by William Labov (1972), the concept of tellability has become discussed with regard to a variety of narrative media. Some presuppositions for the tellability of a narrative are notably features such as eventfulness, or changes of state, and the deviation of event or action sequences from pre-established expectations. Alternative notions – such as suspense, curiosity, and surprise, unusualness, switches and contrasts, violations of certain orders (political, social, or moral), breaks with canonical scripts or schemas, and sudden plot switches – have been proposed by various narratologists to pin down the features which contribute to the tellability of stories (cf. Baroni 2014; Ranta 2019). If most events in a pictorial story are too schema-congruent, they tend to be untellable and/or less noteworthy, but if events deviate too much from prevailing
schemata, the pictures may not be be interpreted as narratives at all, or even experienced as disagreeable.

Schemata give rise to expectations, and if deviations occur, active cognitive processing becomes necessary. Referring to Piaget’s notions of accommodation and assimilation, the psychologist George Mandler, for example, has argued that such adaptations create affective experiences, or, more exactly, arousal changes in the autonomic nervous system. Events or stimuli which are extremely congruent with prevailing expectations or schemata are easily identifiable, but may also result in a relatively low affective value. Various degrees of discrepancy, however, may lead to positive or negative emotional experiences. In the case of slight incongruity, only demanding assimilative processing, the emotional experience is intensified and positively evaluated, as well as in some cases of striking incongruity, where, however, the stimuli have been successfully accommodated. Unsuccessful attempts to accommodate new information will, though, result in a negative valence. Events which, after cognitive processing, can be adapted to alternative schematizations, that is, with somewhat delayed congruity, tend generally to be experienced as agreeable (G. Mandler 1982, 1984; cf. also Ranta 2000: 239–240).

These considerations may very well be relevant for understanding pictorial narrativity and the affective relevance of schemata. Still, while schema deviations have been approved of in certain cultural contexts, in others, such as ancient Egypt or the Middle Ages, schema confirmation or striving for the “pleasure of recognition” seems to have dominated. Such cultural variations must of course be acknowledged, although some basic preference mechanisms may be cross-culturally valid; even in schema-conservative societies at least slight or subtle deviations might have been appreciated.

Further investigations, not least from art history, might fruitfully demonstrate these issues in more detail and with additional examples, that is, by exploring the relationship between various kinds of schema knowledge and the production as well as reception of pictorial narratives. Such research would certainly also be worthwhile including cross-cultural comparisons, regarding diverse historical periods with their specific preferences for schema deviations or confirmations.

References


