13

BELIEF, DESIRE, ACTION, AND OTHER STUFF: THEORY OF MIND IN MOVIES

Daniel T. Levin, Alicia M. Hymel, & Lewis Baker

AT THE BASE of our understanding of almost any film are the beliefs, desires, and goals of the characters. In fact, thoughts about these kinds of things are, according to many researchers, a primary component of our understanding of almost all social situations, and so it should not be surprising that an art form that tells stories about human conflict, desire, aspiration, and emotion should draw heavily on the everyday skills necessary to understand the mental representations inside the heads of human agents. These thoughts are usually referred to as intentional, and the specific causal principles organizing them constitutes a theory of mind (TOM), which, according to many researchers, constitutes a basic domain of thought subserved by a specific neural system. Over the last 30 years there has been an explosion of research exploring this system, and by now a range of well-articulated theories are available to help us understand its parts. One of the most interesting things about this research for present purposes is that TOM seems to have strong foundations in visual experience. As infants look to their surroundings, they see other agents who not only look at important things but also have thoughts about them. These early experiences with the simple actions that agents engage in eventually flower into an understanding of the beliefs, desires, and goals that constitute the cognitive backstory to the external visual reality of living, thinking agents. Accordingly, we would argue that many of the perceptual and cognitive skills necessary to understand both the basic visual events in edited films and the broader stories they tell are derived from the everyday mentalizing skills that constitute TOM.

Given the clear applicability of TOM to film, it seems important to consider specific ways in which this cognitive system might help viewers comprehend the visual narratives characteristic of film. In this chapter, we first review the perceptual and cognitive subprocesses that make up a fully functioning TOM system, and while doing so make explicit links between this system and the comprehension of movies. This review will roughly conform with previous distinctions between initial TOM processes, including gaze following and true belief tracking, and later more effortful processes necessary to understand ways in which beliefs diverge from reality or constitute multiple levels of subjective analysis of events and stories (e.g., Baron-Cohen, 1997; Leslie, Friedman, & German, 2004). Accordingly, we first review research and theory on the core processes that support TOM by allowing people to understand basic facts about how people's beliefs explain what they do. Then, we explore the higher level processes that help people understand the complexities that occur when beliefs are incorrect or when different people's beliefs about the same thing diverge.

While reviewing these simple and complex TOM processes, we focus both on how understanding TOM can explain specific aspects of film and on how filmmaking practice can inform our understanding of TOM. However, one of the drawbacks of a broad multiprocess model of the interaction of TOM and film is that it tends to become an ad hoc exercise in matching practices in film with specific cognitive processes that can hardly fail, and it may not be particularly incisive. Therefore, to sharpen our analysis, we will focus the final section of this chapter on a discussion of the extent of TOM in film by asking what happens in situations where TOM is *clearly not* the only kind of cognition necessary to understand action. We hope this analysis will bring the special role of TOM into relief and perhaps allow us to generate interesting hypotheses about the cognitions underlying film, and about cognition more generally.

Perceiving Social Actors and Understanding the Things They Do: The Front End of TOM

One of the most basic elements of an effectively edited film is a correctly done cutaway in which a shot of a character looking off-screen is followed by a shot of the thing he or she is looking at. For example, in the film *Blue Velvet*, the character Jeffrey is shown in a long shot walking across a field and then lingering to throw some stones (Figure 13.1A). When he looks to the ground to pick up a second stone he pauses, intently looking at something on the ground (Figure 13.1B). This look is followed by a close-up of the thing he is looking at—a severed ear (Figure 13.1C). This is a classic example of a cutaway, and it is important to note that the initial cutaway of the ear shows a view that is very different from the long shot that preceded it. (However, the director clearly matched Jeffrey's hands on the first cutaway. It is interesting to note that he did not feel the need for this match on the subsequent cutaway;

FIGURE 13.1 Stills showing a series of cutaway shots from the film *Blue Velvet* (1986, De Laurentiis Entertainment Group).

Figure 13.1D–F.) So, one might ask how on earth the viewer is supposed to understand that the ear is even in the same part of town as Jeffrey, much less the thing he's looking at? This is not a trivial question, and early filmmakers (and their backers) were deeply concerned that a succession of visual scenes that viewers did not choose for themselves would be baffling. After all, the cutaway was inflicted upon our visual system. It was not something the film viewer chose to look at, so many of the internal cues that would otherwise be associated with an internal reorienting of attention are absent. The answer is probably that we understand where the ear is, and that it should be linked to Jeffrey's stare, because many other kinds of information derived from our basic understanding of how people look at things and think about them are sufficient to make up for the absence of internal information about gaze reorienting. From everyday experience, we know that people look at things that interest them, and that we might usefully follow their gaze to observe important objects. Therefore, the cognitive principle organizing our understanding of this event is our understanding that he is looking at the ear *because he's thinking about it.* This insight seems simple, but it is based on a whole series of cognitive and perceptual skills: One has to perceive Jeffrey's face and eyes, note where they are pointed, somehow represent the location they are pointing to, and, most critically, attribute a thought process to Jeffrey that has led him to look at the thing.

One of the most interesting things about the link between visual storytelling and cognitive science is the strong theoretical link between early visual experience and the development of theory of mind (Flavell, 2004; Flavell, Green, & Flavell, 1990; Gopnik, Slaughter, & Meltzoff, 1994). Consistent with this, severely visually impaired children have been found to have delayed performance on tasks testing theory of mind (McAlpine & Moore, 1995; Minter, Hobson, & Bishop, 1998). In this part of the chapter, we review these skills, discussing research that documents how infants perceive faces, gaze, and basic events. We also discuss research documenting how these basic skills are initially built into an understanding of belief and how these understandings may interact with the online perception of events.

The ability to visually monitor others' behaviors and shifts of attention is instrumental in developing the skills necessary to understand their intentions. The foundations of these abilities can be observed in newborn infants. Immediately after birth, infants focus their attention on facelike objects (Morton & Johnson, 1991), and shortly thereafter, they can imitate facial expressions (Meltzoff & Moore, 1977) and preferentially fixate on targets cued by the eye movements of another agent (Farroni, Massaccesi, Pividori, Simion, & Johnson, 2004). Infants elaborate these simple skills into the ability to follow the gaze of another person, until by the time they are 18 months old, they can look at an object that another person looks at, even if the object was initially out of the infant's view or behind a barrier (Butterworth & Jarrett, 1991; Hood, Willen, & Driver, 1998; Moll & Tomasello, 2004).

However, as a number of authors have pointed out (see, e.g., Woodward, 2005), none of this demonstrates that young infants have any insight into others' mental states. Knowing to look where someone else looks and knowing that living things behave differently from nonliving things do not require much of an understanding of the beliefs, desires, and goals that drive these actions. The interesting thing about these early skills is that they may be sufficient to gain a foothold on perceiving cutaways—the basic habit of linking a looker with a gazed-at object seems sufficient to overcome any confusion about the spatial relation between the two. Furthermore, infants who can follow gaze to occluded objects or to objects out of their own current view would likely have little difficulty with similarly nonoverlapping views in cutaways, so long as the timing of these edits is not too different from real-world timings. Accordingly, it is possible to hypothesize that at least some of the skills involved in perceiving shot-to-shot relationships require only precognitive TOM-related skills as opposed to a full-scale cognitive analysis of thinking.

Of course, soon after their second year of life (and some would argue even earlier), children do begin to understand the beliefs that drive looking. According to Tomasello (1995), the foundations for this achievement are laid by several kinds of joint-attentional interactions of increasing complexity. The first of these to develop is simultaneous looking, in which infants orient the same location as another agent but do not necessarily attend to the same properties of the location. In this instance, the agent's attentional focus acts as a cueing mechanism to guide the child's attention. However, the child does not yet have insight into the agent's intentions and is unable to determine the precise target of the agent's attention. After 9 months, children begin to understand that other people are independent agents with their own unique mental states. With this realization comes the ability to both follow the focus of an agent's attention and attempt to change that focus with their own behavior (such as pointing). In this more sophisticated joint attention, when following another's gaze, not only is attention directed to the location of that gaze, but also the infant understands that he or she is sharing attention to the gazed-at location in conjunction with the other viewer. The ability to engage in true joint attention is believed to be one of the first and most basic indicators that infants understand that other humans are intentional agents with their own mental states (Tomasello, Kruger, & Ratner, 1993).

Beyond inferring mental states from another's gaze (or predicting another's gaze from beliefs about his or her mental state), we use our Judgments about an agent's beliefs, desires, and goals to understand the agent's actions as he or she moves throughout the world. This understanding can be seen both when making simple predictions such as the completion of a grasping motion, and when tracking the full sequence of complex events, such as the steps required when making a cup of coffee, or the necessary tasks a person must complete before leaving for work in the morning. As in gaze perception, the developmental foundations of the emergence of a link between pre-TOM skills and action perception can be seen during the first year of life. For example, infants can effectively distinguish the basic movements of living things from nonliving things (Kuhlmeier, Bloom, & Wynn, 2005; Spelke, Phillips, & Woodward, 1995) and seem to know that living things have something akin to basic object-directed goals (Gergeley, Nadasdy, Csibra, & Biro, 1995; Woodward, 1998).

The strong influence of TOM on basic event perception can also be seen in adults. Heider and Simmel's classic study exposed people's willingness to ascribe mental states and relationship dynamics to simple moving shapes. Participants who watched two triangles and a circle move around and within a rectangular enclosure described the animation as if the shapes were intentional agents (Heider & Simmel, 1944). Even though the stimulus was composed only of moving shapes, participants overwhelmingly interpreted the movie in terms of a series of events with a causal structure, much like one would if watching videos of humans. Other recent research has demonstrated that basic parameters of events can affect perceptions of agents. For example, subjects are more likely to attribute mental states to agents when the agents move at a humanlike pace (Morewedge, Preston, & Wegner, 2007), and particular clearly defined patterns of goal-oriented behavior exhibited by agents, such as chasing, can also have a profound effect on how we perceive and interpret their actions (Gao, Newman,

& Scholl, 2009). Other research has shown that manipulations of kinematic properties of simple animations, such as velocity and flow, can predictably mediate viewers' perception of genre and narrative (Visch & Tan, 2009). Results such as these suggest that adults have a tendency to see a wide range of events as evidence of agency, implying that TOM is broadly invoked to understand simple events (see, e.g., Barrett & Lanman, 2008).

Although these broad attributions of agency might be seen as evidence that all of the deep consideration of mental processes associated with TOM occurs by default when perceiving events, a range of findings suggest this is not the case. For example, on some measures even infants do not attribute agency to simple moving shapes. Research has shown that 12-month-olds make predictive eye movements guided by goal states. When watching a movie of an adult moving balls into a bucket, the infants' gaze arrived at the bucket before the adult's hand. However, in a condition with self-propelling balls (and no human agent), this pattern of predictive eye movements was not observed, and infants' gaze arrived at the goal after the ball (Falck-Ytter, Gredebäck, & von Hofsten, 2006). Similarly, there are many situations where adults effectively discount surface features of mechanical agents (such as the anthropomorphic features of robots) when making basic predictions about their actions (Levin, Killingsworth, & Saylor, 2008; Levin, Saylor, Killingsworth, Gordon, & Kawamura, in review), and so do not conclude that robots are intentional, in stark contrast to findings in other situations in which adults seem very willing to attribute agency to intelligent machines (Nass & Moon, 2000).

These findings suggest that there is substantial variability in the depth to which TOM is applied when interpreting events. One interesting consequence of this variability is that online event perception may sometimes fail to include all of the basic processes that TOM might afford. Recent research in our lab suggests that this relatively minimal default has particularly interesting implications for online event perception. One key function of a sophisticated analysis of online events is to predict upcoming actions based on an actor's beliefs, desires, and goals. If there were a strong default to predict the sequence of intentional events, then one would expect that misordered events would be easy to detect. However, our ability to perceive order may not be as foolproof as once thought. We have recently demonstrated that it can be difficult to perceive if an event is out of order (Hymel & Levin, 2011). Participants were shown a series of videos of people performing common activities. For example, one video depicted a woman making a cup of coffee. The video began with an establishing shot of the actor reaching toward a coffee pot. Over a series of cuts, the actor pours a cup of coffee, adds creamer and sugar, stirs the coffee with a stirrer, and then drinks from the cup. However, in some of the videos, the actor removed the stirrer from the cup after she used it to stir the coffee. Even when they were allowed to focus their full attention on finding reversals, participants frequently missed them, and participants almost never saw reversals when they were performing a simultaneous interference task, or when they were not warned that reversals might occur. These results suggest that online event perception does not necessarily involve an intensive process of constant prediction and error checking.

Not only do these findings demonstrate that the online analysis of events in films may be minimally predictive, but this; inability to consistently track order across multiple short scenes may also provide an opportunity for filmmakers to manipulate the order of actions in subtle ways while operating outside of their audience's awareness. The use of artistic variations in event order to evoke mood was suggested by Münsterberg (1915/2001), who argued that the order of scenes in a movie need not necessarily reflect the order in which the events might realistically unfold. One particular instance Münsterberg discusses is altering the order of actions to reinforce a character's excited state, as if to reflect his or her quick and disrupted thoughts. Although filmmakers are rarely so bold as to completely reverse events, a similar effect can be seen in the film *The Conversation*. In the climatic scene, the protagonist Henry Caul has rented a hotel room adjoining another room where he believes a murder takes place. The beginning of the scene is edited in a relatively conventional manner, in strong continuity and very little ellipses as we see Caul listening in dread to the events next door. In the middle of the scene Caul wanders out to the balcony of his own room and is shocked to witness the murder on the balcony next door. He retreats in terror to his hotel room, closes the curtains, turns on the television, readjusts the curtains to ensure they are completely closed, and then retreats to his bed across a series of shots. The shots depicting the events immediately following the murder mirror his terrified and fragmented state of mind (this fragmentation partly stems from the main theme of the movie, which is that Caul is constantly violating his own basic don't-get-involved philosophy). Not only are most of the shots joined by jump cuts in which Caul suddenly changes location in the hotel room (in one shot he is suddenly standing while in the last frame of the preceding shot he was sitting), but also some of the things he does in one shot appear to be reversed in the next shot. For example, immediately upon returning to the room, Caul closes the curtain to the balcony completely, but in the first frame of the next shot, we can see that the curtain is clearly open, and in the first frame of the shot following that one, it is again closed.

The degree to which an audience is aware of these misorderings is not clear, but our research suggests that Münsterberg (2001) may have been correct, and that misorderings may be useful as a subtle nudge rather than a blatant attention-attracting device. More generally, failures to detect misorderings should not be taken as evidence that TOM is not important in understanding events. Rather, the point of this research is that perceiving and understanding events does not require online predictions, and that other kinds of action processing may be much deeper. However, the more general point that there is variability in the depth to which beliefs, desires, and goals are considered might be consistent with models of TOM that make a distinction between basic automatic TOM subprocesses, such as gaze detection and joint attention, and more cognitive processes responsible for belief tracking that might not be automatic. In the next section of this chapter, we turn our attention to processes that more explicitly track true and false beliefs. This kind of belief tracking is important not only for researchers who study people's understanding that beliefs are not mere copies of the world, but also for filmmakers who want to explore the consequences of beliefs that diverge from reality in a wide variety of ways.

The Role of TOM in Conceptualizing Narratives

Tracking gaze, understanding it as a link between objects and thinking, and using it to share attention are important elements of most models of TOM. However, none of these processes reflects a deep understanding of the beliefs that underlie behavior. In order to get beyond the simple inference that people do things because they have specific beliefs, it is necessary to understand how beliefs are related to, but not exactly the same as, their real-world referents. After all, beliefs are not internal representations that completely and veridically recapitulate the current state of reality. Instead, they are intentional representations used by agents who may or may not have had the opportunity to observe any particular detail inherent to the current state of a rapidly changing and dynamic world. Therefore, much developmental research has explored children's ability to predict another person's behaviors by understanding how

that person has out-of-date false beliefs. Children's ability to do this is taken as evidence that they understand how the beliefs that drive behavior must be taken as a level of analysis separate from the actual state of the world. For example, in the well-known Sally-Anne task (Baron-Cohen, Leslie, & Frith, 1985), children are introduced to two dolls—Sally and Anne. The children watch as Sally picks up a marble, places it in a covered basket, and then "goes for a walk." While Sally is absent, Anne takes the marble and places it in another basket. The children are then asked where Sally will go to find her marble. Children under 4 years old will often mistakenly claim that Sally will look for the marble in the new hiding place. This error demonstrates that children have difficulty tracking a mental representation (Sally's false belief about where the marble is) that real-world events have become outdated and therefore predict that Sally will look in the wrong place. In contrast, 4-years-olds who correctly predict that Sally will look in the old hiding place have successfully understood that she is laboring under a false impression and have been able to use that understanding to understand her incorrect actions.

If this all sounds like the basis for any of a thousand sitcoms, it will come as no surprise that many authors have argued that these higher level TOM skills not only are useful in understanding almost all narratives but also are deeply embedded in much of human culture. The skills necessary for this robust understanding of how beliefs drive behaviors are also known as "mindreading" or intersubjectivity, and some have hypothesized that theory of mind evolved in response to growing populations of early migrant humans (Dunbar, 2000; Mithen, 1996). These theories propose that TOM co-occurred with the development of language. The growing population coupled with the benefits gained through the exchange of ideas would have created an additional challenge: keeping track of in-groups, out-groups, and the flow of information among them. Knowing the reliability of information soon became just as important as the information itself.

Theory of mind has even been proposed as an evolutionary basis of skills that allow people to create and understand complex narratives (Abbott, 2008; Butte, 2004; Palmer, 2004; Zunshine, 2006, 2008). Just as early humans required TOM to track group dynamics through language, Zunshine (2008) argued that modern written narratives rely on particularly complex forms of TOM to track plots and characters. However, while TOM has been discussed at length in its applications to understanding written narratives, few have explicitly compared TOM in visual and nonvisual narratives. Film provides a visual context for mindreading that is similar to its everyday context, and in suggesting that a theatrical context for TOM facilitates deeper mindreading, Zunshine (2008) implies that this visual context may induce interesting changes into the kind of mentalizing characteristic of these different mediums. In the next section, we outline some of the different kinds of mentalizing necessary to understand narratives. We will then discuss the perceptual factors that occur simultaneously with TOM in film that might distinguish film from literature.

Higher Level TOM in Narrative

False beliefs proliferate throughout modern narratives. Sitcom characters are often victims of mistaken identity, red herrings draw detectives away from killers, and misconstrued circumstances tear romances apart. Several common narrative devices rely on an audience's ability to follow the stream of beliefs with relative ease. When successful, these narrative devices push audiences to infer, understand, empathize, and identify with a character's goals (Bordwell & Thompson, 2008).

Perhaps the most basic of these devices is dramatic irony (see Table 13.1), which occurs when the audience is privy to information that one or more characters do not know, building tension to an inevitable end (Fowler, 2004). An example can be found in *The Silence of the Lambs*. At the climax of the film, Agent Starling is conducting interviews with "Jack Gordon," who the audience knows to be Buffalo Bill, the serial killer. The relatively mundane visual and spoken narrative of the scene only becomes suspenseful when we understand that Gordon intends to kill Starling, who is unaware of the danger. The scene comes to a peak when Starling realizes Gordon's true identity.

Dramatic irony can be turned on its head, creating a false-false belief on the part of the audience. In many a horror film, we understand that the character has a false belief of safety—we know that a psychopath secretly lies behind a door, for example. As the character slowly opens the door we are driven to suspense as we expect the murderer to attack. Often, as in the *Scream* series, the murderer is not in that room, but another room. The audience has been tricked and the character stays oblivious and safe. Our belief about the character's false belief is false.

A second example of false belief propelling narrative is in the construction of mystery narratives (Fowler, 2004). Misdirection is central to the mystery narrative, in which the audience must constantly update a series of beliefs about the world. In Agatha Christie's *And Then There Were None* and the various film adaptations that have followed, each of the 10 characters suspects or trusts different characters throughout the narrative. Not only is the audience encouraged to draw, and then revise, its own conclusions as to the killer's identity, but also the various alliances, suspicions, and red herrings in the plot change the characters' beliefs. The different climaxes occur when a character seems to know a clue the audience does not immediately understand. Not only are we driven to know the character's beliefs, but we must question whether those beliefs are true. This tension is resolved when a character's finally complete knowledge is revealed to us. An excellent example occurs in *The Thin Man*. At the climax of the film, the protagonist Nick Charles deduces the identity of the "thin man" and his

TABLE 13.1

Documenting the Progression of Mental State Knowledge Along Narrative Timelines

	Exposition	*Climax*	*Resolution*
Dramatic irony	Audience knows everything.	Audience knows everything.	Everyone knows.
	Character A knows nothing.	Character A knows B.	
	Character B knows nothing.	B does not know A.	
Narrative hook	Audience knows nothing.	Audience only knows A.	Everyone knows.
	Character A knows B.	A knows B.	
	Character B knows A.	B knows A.	
Mystery	No one knows anything.	Audience knows nothing.	Everyone knows.
		A knows B.	
		B may know A.	

killer. He then calls a now-clichéd dinner party where he reveals the true killer among all the suspects. At the climax, the only narrative device driving the plot is the audience's unresolved understanding of Nick's knowledge. We are driven to suspense by not knowing.

As a final example of false beliefs in narrative structure, we look at the device used to "hook" audiences. The hook consists of a quick exposition whereby the audience is introduced to a set of characters (Fowler, 2004). One of the most common forms of narrative hooks is the "dramatic action," where a story unfolds with a series of unusual but unexplained actions (Fowler, 2004). The audience must quickly identify the beliefs and goals of the characters and frame them within the greater story (for examples, see Abbott, 2008). For instance, at the beginning of *Casablanca*, we see an as yet unnamed Ilsa walk into Rick's Café, where she asks Sam to play "As Time Goes By." Hearing the song, Rick marches over to the piano. As he approaches he sees Ilsa and his face changes. This brief yet important scene plays off our innate drive to describe the motivations of others. These actions lead us to ask the questions: "How does Ilsa know Sam?" "Why does she request 'As Time Goes By'?" "Why does Rick react angrily?" "And why does he appear shocked when he arrives at the piano?" One simply cannot explain the narrative of these visual events without delving deeply into the beliefs, desires, and goals of the characters. We can infer that because Ilsa appears to know Sam, and because Sam and Rick have discussed their long history together earlier in the film, Ilsa must also know Rick. From this same logic, we can infer that if Ilsa and Sam know "As Time Goes By," then Rick must also know the song. Rick's anger in response to the song implies some history between him and Sam involving the song, and thus possibly involving Ilsa. This theory is confirmed when Rick sees Ilsa and reacts with shock and abated anger. Thus, the authors have manipulated the human drive to understand other minds to entice audiences into wanting to learn more.

The examples we have discussed are fairly simple, often balancing one or two mental states at a time, but as we can see from *Casablanca*, intersubjective reasoning gets complicated quickly. Zunshine (2007) and Butte (2004) have argued that modern authors (beginning with Jane Austen) purposefully omit overt descriptions of a character's mental states so as to force readers to make inferences. While there are some advantages to attributing false belief reasoning to narrative, there may be a limit to our ability to process intersubjectivity that peaks around four levels of depth (Kinderman, Dunbar, & Bentall, 1998). To summarize Kinderman and colleagues, levels of depth work as such:

1. I *know*
2. I *know* she *knows*
4. I *know* she *knows* that he *knows* that I *know*
5. I *know* she *knows* that he *knows* that I *know* what she *did* when he *wasn't looking*

As we can see, the entire framework quickly degrades into absurdity past four levels of depth. This limitation could be indicative of something other than theory of mind processing. In order to keep track of multiple intersubjective abstractions, one must track physical states (the presence and location of characters) as well as mental states (awareness of events). In a written or verbal format, this abstraction relies heavily upon working memory, spatial reasoning, and long-term memory of each character's goals. For example, the Sally-Anne task tests second-order intentionality: The child knows (first) that Sally thinks (second) the marble is in the basket. Additionally, the task can indicate third-order intentionality whereby the child knows (first) that Anne thinks (second) that Sally thinks (third) that the marble is in the basket. To demonstrate a theory of mind, participants traditionally have to show a second level of abstraction, indicating a separation of their knowledge from the knowledge of others. What happens with additional levels of abstraction is not just a separation of knowledge, but an elaborate system of tracking and noting different beliefs. This becomes remarkably tricky in literature, where spatial relationships must be inferred and maintained in working memory.

The limitations of TOM are not necessarily strict limitation, but are rather seen by some as a challenge to constructing narrative (Butte, 2004; Zunshine 2006, 2008). In this way intersubjectivity is not just a tool to further the plot. Multiple layers promote cognitive involvement of the reader, a greater understanding of complex situations, and an empathy with characters caught in the intersubjective web. This kind of involvement can in turn provide the motivation to delve more deeply into the characters' minds, and if this motivation is not present, this multilevel intersubjectivity is likely to be ignored altogether (Zunshine, 2006).

Intersubjectivity in Film and Written Narrative

In written narrative, the reader must imagine adequate representations of each character's goals, motivations, and personal history, in addition to the character's immediate physical ability to acquire knowledge independent of other characters. Literary conventions can ease this task. In the novel (as well as other narrative forms), readers can utilize genre and character archetypes to assist with their judgments (Bordwell & Thomson, 2008; Campbell, 1949; Visch & Tan, 2008). For instance, we rarely question the motivations of Phillip Marlowe in *The Big Sleep* because the novel's genre (pulp detective fiction) and Marlowe's archetype (the surly but morally driven antihero) lead the audience to assume a limited possibility of mental states for the charming, if abrasive, protagonist. The author, Dashell Hammet, still manipulates Marlowe's intentions, but these variations are scaffolded by the audience's genre-specific expectations. Meanwhile, Humphrey Bogart's Marlowe in the 1946 movie of the same name is even less ambiguous. The genre and character conventions still apply, but the film uses several additional tools to facilitate TOM. Thus, we are better equipped to understand the elaborate triple-double-cross played by Marlowe because an effectively edited scene allows viewers to use gaze, spatial relationships, and temporal cues to support basic TOM skills (Berlinger & Cohen, 2011; Smith & Handerson, 2008). More generally, consider a character who suddenly shoots a steely gaze up at the bully who towers over him. In a simple behavior that occurs in a contested space, it is possible to use a rich set of visual cues to reinforce a complex intersubjective plot in which one character suddenly draws strength over a bully who is unaware that his momentary tactical advantage is about to be overcome.

It is therefore possible that visual narrative can rely on familiar visual events to free up the cognitive resources that would be used in an otherwise highly effortful process of TOM-based updating. However, it is also possible that this putative visual freeing-up of resources is only an illusion. Research on visual attention has demonstrated that people are strikingly overoptimistic about the degree to which they will be aware of visual changes (Levin, Momen, Drivdahl, & Simons, 2000) and unexpected visual stimuli (Levin & Angelone, 2008). This overoptimism led Varakin, Levin, and Fidler (2004) to warn visual interface designers about an "illusion of visual bandwidth" whereby they overestimate the degree to which visually rich

computer interfaces are effective in transmitting important information to users. There are a range of reasons to believe that an account of visual support for TOM might fall prey to this illusion as well, because the support provided to TOM by visual events comes with some marked costs. For example, a filmed scene will end regardless of the audience's ability to infer intentions. In contrast, readers can easily backtrack if they do not understand something. It is interesting, however, to note that real-world interactions probably vary considerably in the degree to which it is possible to interactively gain clarification about someone's thoughts, so it would seem that both film and written narratives impose realistic constraints on TOM reasoning. Accordingly, it is possible to contemplate a research program designed to test the relative costs and benefits to TOM-based reasoning afforded by visualization of intentions across different situations.

There is one last but rather large factor to be accounted for in narrative design: the authors and directors. Abbott (2008) argues that the author must take more levels of intersubjective depth into account than the reader, as the author must infer the reactions of the audience to the work. Abbott does not account for the role of audience as critic, though. It could be said, then, that the audience must account for what the author intended for an audience to infer from the characters. Whether or not this game of inference occurs, and if it matters at all, is a subject more fitting for literary theory. What is important for psychological research is how directorial style influences the inference of mental states. For instance, a scene shot in mostly close-ups may lead to different inferences than the same scene shot from a single "fly on the wall" perspective. In addition, an incompetently edited scene may fail to support TOM and therefore lessen the accuracy of inductions made by the audience. These might be the basis of a particularly interesting research project because it would allow us to understand the specific relationship between visual events and TOM. For example, it is possible via the timing of a cut to suggest that an actor has immediately produced an utterance or has paused before speaking. This clearly has the potential to communicate the nature of the actor's beliefs, but the same information might be transmitted via the lines the actor is actually producing. If this redundancy can be shown to increase memory for the actions or subtext of the scene, it would demonstrate that visual events and language combine to produce a final understanding of the scene.

Seeing TOM in Its Limits

The movie *High School Musical* may not be particularly deep, but it does contain at least one interesting moment of clarity. The central conflict in the film revolves around false choices between different groups of friends and the activities that go with them. At one point an infuriated character spits out that the entire social order in the school is coming apart because people are doing "other stuff"—the jocks are singing, the brains want to dance, and the skateboarders want to play the cello. The point is, of course, that the characters grow by rejecting social pigeonholes and by accepting their friends' true natures (and the delicious food that comes with a basketball player's obsession with baking). However, in straining against these social straight-jackets, narratives like this seem at least a little perverse because they rely on social stereotypes as a springboard for the story, bringing the categories they rail against into strong relief. The viewer needs to know the stereotype of a jock to understand

how shocking it is supposed to be that the guy wants to sing show tunes. In this section of the chapter, we attempt to bring the role of TOM in film into relief by considering what happens when it is considered in the context of other stuff (although we hope that doing so is not perverse). After all, TOM is not the only core system helping us to perceive and reason about events—knowledge about physical cause, number, space, and biology have all been implicated as autonomous core systems of reasoning, and even if one does not buy into a domain-specific view of cognition, it is clear that events may vary with respect to the breadth of the cognitions they invoke. In this section, we speculate about what might happen when TOM meets its limits and events must be understood in terms of multiple systems of reasoning. It is important to note that we did not use the word *speculate* lightly: Although there is a large amount of research exploring TOM and each of the other core domains, the vast majority of this work concentrates on one of these kinds of thought at a time. Much less work explores what happens when events must be understood using multiple systems.

Perhaps the most logical place to start this section is with the observation that TOM tends to be a promiscuous system in that people seem to apply it readily in a wide range of settings, based on relatively thin evidence. One line of evidence supporting this comes from research exploring people's concepts about God (Barrett & Keil, 1996). On the one hand, Barrett and Keil observed that most people readily agree in the abstract that God cannot be much like a human, that such an entity cannot share any of the cognitive or spatiotemporal limits that characterize people, and that God's beliefs, desires, and goals are not commensurate with those of humans. Even so, when participants read stories about God's actions, participants falsely recognize sentences implying all of these mundane limits. Accordingly, it appears as though people implicitly apply their basic TOM-derived assumptions about thinking to an agent that they know cannot be constrained by such limits. A similar phenomenon can be observed in a more perceptual setting: The famous Heider and Simmel (1944) experiments demonstrate that people require only the most minimal evidence to apply TOM to simple moving shape animations that only hint at having any cognitive capacity at all. These findings have been reinforced in a wide range of settings, as research has demonstrated that people readily anthropomorphize dogs, cats, cars, computers, and even alarm clocks (for review see Epley, Waytz, & Cacioppo, 2007). This research led Barrett (2004) to hypothesize a "Hyperactive Agency Detection Device," an automatic neural module that sets off an alarm any time the slightest sign of agency appears.

The ready application of TOM in a wide variety of settings clearly demonstrates the importance of this system, so it is not difficult to make the argument that it underlies a large proportion of the cognitive processing necessary to understand events involving sentient agents. However, it may be that the cognitive sciences have slightly overemphasized the importance of the TOM system and perhaps have missed opportunities to observe the full range of processes necessary for its real-world application. This limit derives primarily from the typical cognitive science TOM experiment that either explores TOM in isolation or does not focus much attention on other cognitive systems necessary for the correct interpretation of events. This is one thing that makes the Barrett research so interesting—it demonstrates how one kind of thinking (TOM) runs roughshod over another kind of thinking (specific religious concepts). Other research does explore the relationship between TOM and other kinds of thought, primarily as a means of demonstrating TOM's limits. For example, developmental research has explored how children reveal

TOM-based errors on false belief tasks that do not extend to other kinds of false representations, such as those that might be characterized by a photograph that has become "out of date" because the world has changed (Slaughter, 1998). More recently, we have explored how people limit their attributions of agency when reasoning about the actions of computers and robots (Levin et al., in press). In both cases, intentional attributions stop in the face of some specific barrier.

All of these findings suggest the importance of exploring how TOM is balanced with other kinds of reasoning. This idea has rarely been explored explicitly, and we suspect that a close look at the movies will provide guidance as we ask new questions about the interactions among different domains of thought. For example, what happens when an event involves both TOM and another core domain of thinking such as number? One can easily imagine a scene in a film in which two characters fight over some number of objects. So, both characters might see six cookies in front of them, but one character thinks that all of them have been left for her by a friend, while the other actually knows that the cookies are leftovers from a party. To understand the ensuing argument, one needs to consider both the conflicting beliefs, desires, and goals of the characters and the fact that there are six cookies, which could equitably be divided between two people who want them. This situation provokes whole series of interesting questions about the cognitive processes that might underlie the need to manage two fundamentally different kinds of thinking, but one interesting way of considering this question would be to explore how filmmakers have dealt with stories in which TOM must be understood in the context of, and balance with, reasoning in other non-TOM domains of thought.

One interesting example of the balance between TOM and other domains of thought plays itself out in the Michele Gondry film *Be Kind Rewind*. The film revolves around an old video rental store still stuck in the days of VHS. At the beginning of the film, the store's owner goes out of town, leaving the store in the care of his employee, Mike, who must fend off the bumbling Jerry, a local character who is obsessed with the brain-damaging microwaves emanating from the power substation near the trailer he lives in. Jerry develops a foolish scheme to destroy the transformers, and in the process of breaking in, he electrocutes himself in a dramatic lightning explosion. He emerges from the plant dazed and charred, but otherwise functional. Of course, he heads straight for the video store, and we gradually discover that he has been magnetized and as he walks around the store, his proximity to all of the videos erases them completely. The interesting thing about this event is that it requires knowledge both about intentional and nonintentional causality to understand. At the broadest level, the audience understands that Jerry has accidentally erased the videos, but that the accident clearly stems from Jerry's recklessness. In a previous scene Gondry has emphasized this by showing Jerry accidentally knocking some of the very same videos he will eventually erase off of their display shelves as he horses around the store with some kids. Another scene preceding the erasure visually emphasizes the danger Jerry poses to the tapes by strongly juxtaposing Jerry with the tapes (Figure 13.2A) as he works to convince Mike to join him in sabotaging the power station. It is interesting to note that during this scene Mike is trying to decode a message from the store's owner to "keep Jerry out" of the store. So, an intentional framework not only helps viewers understand the low level of purposefulness characteristic of the accident but also helps viewers understand Mike's gradual discovery of his instructions to protect the tapes from Jerry, and Mike's and Jerry's discovery that Jerry has erased the tapes. In both

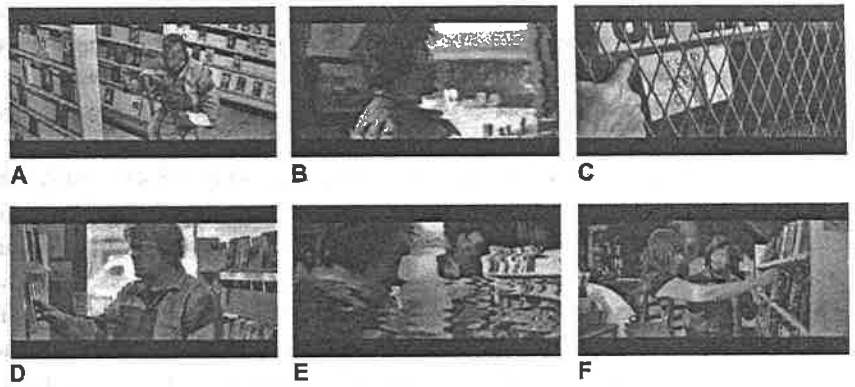

FIGURE 13.2 Scene from *Be Kind Rewind* (2008, New Line Cinema) combining intentional and nonintentional causality.

of the latter two cases, the audience is allowed to enjoy an act of recognition as the characters arrive at conclusions that the audience has already understood.

Clearly, a key part of this scene is the need for mentalizing to be combined with an understanding of distinctive physical forces. This is not unusual because intentions often have their effects through behaviors that operate in the physical world. When someone wants something, he or she grabs it, and so even in straightforward settings some combination of intentional cause and physical cause must be understood. However, there are easy-to-understand and difficult-to-understand versions of both forms of causality. As we have reviewed earlier, understanding people's false beliefs and ignorance requires some effort, although in simple cases of incorrect mental representations the effort is probably minimal. In the case of the tape erasure, however, the physical mechanism producing the impact of Jerry's carelessness is not straightforward. The audience has to understand that Jerry has been magnetized, and this requires at least some suspension of disbelief because Jerry is not made of metal. Viewers might also fail to reflexively understand that VHS tapes can be erased by magnets and may need reminding that magnetism operates most strongly over small distances but does not require physical contact to have an effect. Accordingly, viewers may require a fair bit of support to understand what is happening, and Gondry clearly takes pains to facilitate this.

Not only does Jerry's earlier bumbling and Mike's dawning awareness of the urgency of guarding the store against Jerry establish the vulnerability of the tapes, but also a whole series of cues and actions during the scene itself converge to suggest the erasure of the tapes. As Mike staggers into the store we hear an electric buzz accompanied by a picture distortion that might be familiar to viewers from the days when an analog TV would experience some kind of signal interference (see Figure 13.2B), and each instance of the buzzing and distortion after the first one is juxtaposed with the tapes themselves. For example, in Figure 13.2E, the distortion, and especially the accompanying buzzing, mirrors Jerry's movement past a shelving unit full of tapes and the dolly-induced movements of the tape shelves across the screen. To further establish the damage Jerry's magnetism is doing to the tapes, Gondry has him aggressively touch the tapes as he engages in a (relatively forced) argument with a customer, and at one point Jerry slams a tape against the shelves, physically knocking some

tapes onto the floor (the sound of the tapes hitting the floor mirrors the previous episode when Jerry had knocked tapes over). Finally, flashes of light, nominally from the sun reflecting off of passing cars, sporadically shoot into the store, and these serve as a visual link to the shock Jerry experienced at the power plant, superimposing it on the vulnerable tapes (Figure 13.2F).

The interesting thing about all of this support is that it was included despite the fact that the nonintentional action clearly resonates with the intentional action because Jerry's bumbling clearly reinforces the magnetic action at a distance that is not much different from simply touching the tapes to wreck them. In addition, the more general idea that someone can affect something at a distance is likely a precursor of intentional reasoning, so this seems like a case where the nonintentional elements of a scene are not very challenging both because they conflict only minimally with typical intentional events and because they are embedded in a rich intentional context. In addition, all of the basic actions and events in the scene are consistent with basic agency, and the scene is cut in strong continuity. There are only a few ellipses, and spatiotemporal continuity is strongly maintained as Jerry walks around the store. (There is a 180-degree violation in the scene, but it occurs in the context of a strong spatial axis from the inside to the outside of the store, which likely lessens the impact of the violation; Levin & Wang, 2009.) So, if this is a situation where intentional and nonintentional action are relatively consistent, can it be contrasted with another situation where they are more different, or where nonintentional action is more salient?

Although almost all films focus on intentional action, there are interesting exceptions. Perhaps the purest example is the experimental film *The Way Things Go* by the artists Peter Fischli and David Weiss. This film depicts a 30-minute chain of completely mechanical events in which physical contact, fire, gravity, and simple motion form a causal chain much like a slow-acting Rube Goldberg machine. For example, at one point in the middle of the film, a tire rolls down a plane and under a ladder that gets tipped over when the tire hits a board tied to the ladder. The ladder then falls on the other end of the board the tire just rolled over, forcing the tire to continue on its way. The tire then hits a barrel, causing a small car on a board supported by the barrel to roll toward a fuse. When the car gets to the fuse, the lit candle on top of the car lights the fuse, which lights something that makes sparks spray forward. The sparks light a pool of flammable liquid on the floor, which in turn lights another fuse that causes a catapult to shoot a flaming puck at a flammable tether ball, which then unwraps, causing the next event in the chain. This goes on for 30 minutes, essentially uninterrupted (there are a series of subtle dissolves in the film, so the events were clearly not shot in a single take), and the whole thing is fascinating even though it shows nothing but these mechanical events. Even so, the events in the film are perfectly easy to understand, probably because they draw upon knowledge of very basic physical principles that can be interpreted in isolation of any other kind of reasoning. However, perhaps these events are so compelling because they violate at least some of our assumptions that intentional events are much more likely than mechanical events to chain together this way. In fact, a whole host of films emphasize the chaining of intentional events, especially in cases where they push the boundaries of intentionality by being accidental or by having unintended consequences. For example, films such as *Babel* (2006) and the entire TV series *Breaking Bad* are organized around these accidental or semiaccidental chains of intentional events that represent the foreseen and unforeseen consequences of decisions characters make based on beliefs, desires, and goals.

One important question is whether the causal chaining in *The Way Things Go* invokes concepts about agency even though these concepts are not strictly necessary to understand the film. This would certainly be consistent with the idea of a hyperactive agency device (Barrett, 2004). It is interesting to note that there are two distinct ways that agency could shape the cognitions associated with the scene. Viewers might anthropomorphize the objects in the film, or they could invoke representations of the agents who created the scene. The former possibility would be similar to the animacy attributed to the Heider and Simmel shapes, and the latter is at least suggested by commentators arguing that the events inevitably invoke the agents (the artists Fischli and Weiss) who created the chain (Danto, 2005). It is important to note that these two kinds of cognition might, or might not, be mutually exclusive. If viewers' cognitions about the scene are responsive to the logic of agency, thinking about the agents who created the chain would inevitably lessen the degree to which the objects themselves are anthropomorphized. However, this kind of logic does not necessarily determine the cognitive response to the film. It is possible that viewers *both* anthropomorphize the objects and invoke the agents. This kind of broad attribution of agency might even be functional if these cognitions would be effective in preparing viewers either to consider an expanded concept about the agency of objects that might have novel capabilities or to consider their hidden creators. After all, it might make sense to do this in a novel situation where we are uncertain about the causes of events and must therefore prepare a range of provisional hypotheses about what has been perceived.

If *Be Kind Rewind* represents a case where nonintentional causality is essentially subsumed in and consistent with an intentional narrative, there are other cases where nonintentional actions, forces, and ideas are salient as counterweights to human agency or are even portrayed as entirely incompatible with basic human narratives. One obvious example is the long tradition in the visual arts and literature of portraying nature as a justice-serving agent, as in Turner's slave ship–sinking storms. Alternatively, nature can be portrayed as a larger force with little concern for human struggles, as in Stephen Crane's transcendent natural calm surrounding the civil war (and for the converse of this, there is the movie *Cold Mountain* in which the warriors are chastised for blaming the civil war on natural forces that are outside their control when in fact blame for the war rests squarely on their shoulders). The orthogonal thrust of nature and human agency can, perhaps, be most clearly seen in films where the purest form of (almost always) men's agency confronts the cold, unyielding face of a mountain either succumbing to it (*The North Face*) or conquering it (*The Eiger Sanction*). In fact, *The North Face* can be seen as a modern version of justice-serving nature as the men who die on the Eiger are pre–World War II Germans (some Nazis and some more sympathetic non-Nazis).

Clearly, however, in these stories nonintentional forces are really only backdrops to highlight the power or limits of human agency, and even in cases where the story is not so overwrought, filmmakers have a hard time simply depicting natural events on their own. For example, documentaries depicting natural events such as volcanoes or hurricanes often take great pains to contextualize them with a human story, usually about the scientists studying the phenomena who constantly repeat the mantra that studying volcanoes, hurricanes, and tornadoes will help us predict them and save lives. Other films (usually ones where the scientists repeat the mantra that we should avoid ruining nature) put more focus on natural events in their own right, but when these involve animals, there is a strong tendency to

anthropomorphize them, ascribing beliefs, desires, and goals about long journeys, the search for food, and even the need for leisure time. There may be some exceptions to this Rule. For example, highly visual nature films designed to highlight high-definition technology such as the Discovery channel Planet Earth films show; natural events and the actions of animals appear with relatively little human context.

Although the background and justice-serving roles for nonintentional forces exemplified in nature films are common, there may be cases where nonintentional concepts more strongly conflict with intentional narratives. One key example of this comes from situations where nonintentional concepts become too complex for humans to effectively assimilate despite the character's powerful determination to do so. For example, while we are skeptical that perceivers do a lot of online prediction, most cause-and-effect reasoning about intentions does depend on some sense of forward-moving, nonreversible time, and when this is violated the result is often a mental breakdown. This is a common consequence for characters who develop time-travel technology, as in the film *Primer,* where time travel results in confusion and depression, or the film *12 Monkeys,* where time travel results in outright babbling insanity (at least initially). Another particularly interesting case where nonintentional thought leads to mental breakdown is obsession about number. In the film *Pi* the protagonist is driven to understand numerical patterns, and the contest between intentional thinking and nonintentional thought seems like a death match in which one mode of thought cannot coexist with another mode. In fact, with regard to number, there is good reason to believe that intentional and nonintentional thought are at least complementary, if not mutually exclusive. Research exploring the cognitive basis of autistic spectrum disorders (ASDs) suggests that systematizing reasoning about rule-governed analytical causality and number is fundamentally different from empathizing (essentially TOM-based reasoning), and that ASDs and associated developmental disorders such as Asperger's syndrome are caused by a predominance of systematizing relative to empathizing (Baron-Cohen, 2005). On this view, the predominance of the two kinds of reasoning might, in some cases, even be negatively correlated across individuals, although the more likely hypothesis is that they are generally independent and minimally correlated. However, it remains possible that these disorders reflect extreme cases where nonintentional reasoning can actually interfere with intentional reasoning.

A key question one could ask about all of these cases is whether the complexity of reasoning supported by TOM can also be present in reasoning about more nonintentional events and stories. In other words, can non-TOM systems of reasoning support reasoning as complex as the multilayer intersubjectivity inherent to the Jane Austen novels and films (Zunshine, 2007)? This is an important question, and we point out that relatively little research has directly explored the role of TOM in supporting reasoning, so approaching this question would require a relatively broad research project that characterizes the support that people can derive from both TOM and non-TOM knowledge. However, based on existing theory, it is possible to make some initial observations and develop some interesting hypotheses. First, research on problem solving and expertise makes clear the surprising degree to which knowledge supports reasoning. For example, participants are much less likely to make the simple error of affirming the consequent if the premises are familiar to them (Griggs, 1984). This is reinforced by a large cross-section of the expertise literature that documents ways in which knowledge supports both reasoning and perception (for review, see Ceci, 1990). So, on one view, the primary determinant of the complexity of non-TOM reasoning will be domain-specific expertise, especially in domains where there is a large body of systematic knowledge such as number. The alternatives are that more general skills and experience will support relatively complex reasoning, or that reasoning can be supported by more basic perceptual intelligence (e.g., spatial intelligence).

Another key set of research hypotheses might be derived from our discussion of the potential conflict between different cognitive systems. For example, when considering the relationships between TOM and numerical cognition, it is important not only to understand the cognitive processes necessary for mediating these potential interactions but also to understand how different individuals may handle these interactions and resolve between-system conflicts. This latter question is, for example, particularly important for researchers who aim to help children learn math skills. One current idea in this literature is the idea that good math learners "spontaneously focus on number" (Hannula & Lehtinen, 2005). That is, it would seem helpful to focus on number in a wide range of everyday settings because this would afford crucial math-learning practice. However, it is possible that effective math skills are more a question of effective strategic balance between number systems and other systems such as TOM than one of simple focus on number. This could be true for a number of reasons. For example, although focus on number will help one count, perhaps a focus on TOM-related concepts such as ownership leads one to actually operate on numbers by adding and subtracting. It's one thing to know that there are eight cookies on the table, but if one combines this understanding with the idea that three people in the room would each believe that two of the cookies are theirs, then one would need to subtract to realize that only two of those cookies are remain unclaimed. It is particularly interesting to note that edited narrative films are ideal settings in which to test these hypotheses because they present temporally reproducible events viewers respond to with excellent consistency. This might allow for between-condition differences in people's approaches to agency and number to be seen in strong contrast to a relatively small amount of measurement noise.

Intentionality and Emotion

It is important to note that our review thus far has left out one very important element in the interpretation of film: emotion. Perhaps one reason we have been circumspect about the topic is the simple historical fact that research on emotion and research on theory of mind are each vast enterprises, making the prospect of taking on both somewhat intimidating. However, there is a considerable amount of cognitive development research exploring the emerging link between TOM and emotion, and in the present context considering emotion seems necessary not only because emotional processes are clearly closely related to TOM but also because they have been extensively discussed in the context of film (and as an example of the relationship between cognitive science and film; M. Smith, 2004). In the present context it is a bit awkward to discuss emotion because it is rarely mentioned as a "core domain" of thought, perhaps because it seems difficult to conceptualize as a principled domain of knowledge as opposed to a motivational system that moderates thinking across domains. However, people's *understanding* of emotion has been explored in cognitive development research comparing this understanding with TOM, and it is clearly an interesting question

in this context for at least two reasons. First, there are interesting ways in which emotions are both intentional and strongly nonintentional. Second, the intersection of intentional and nonintentional events is often the spark that creates a strong emotional response, ranging from depression to a complete emotional and mental breakdown.

At the simplest level, emotions can be seen as the core response to frustrated or achieved intentions. More abstractly, emotions can be seen as valences associated with characteristic intentional representations. For example, in order to truly understand a person's smile at seeing a nice warm slice of pizza, one has to do all of the TOM mentalizing necessary to understand the person's desire for good food and the person's belief that the pizza is his or hers to eat. On this view, emotions are based on intentional representations (so, the smile is for pizza and will not be produced for a triangular piece of wood, a red and white bit of cloth, or a warm shoe), and they are closely related (for review see Wellman, Phillips, & Rodriguez, 2000). However, practically since the beginning of research on emotion, attributional theories of emotion have had strongly nonintentional components. According to these theories (and to a lesser degree their more modern appraisal-based variants), many emotions are initiated based on simple nonintentional interpretations of environmental stimuli. For example, if you are leaning on your desk, and the jackhammering next door makes you nervous, it is not because your emotion represents true knowledge about the world in which a representation in your head is closely and richly connected with some specific event that ought to make you nervous. Rather, it's because the shaking from the jackhammering has produced a bodily state (trembling) that produces the internal state, and even mental representations, of nervousness. Thus, this initial stage of emotional cognition almost exactly fits the definition of a nonintentional representation. Just as a thermostat will "think" it's hot whether the voltage entering it comes from a heat sensor or a humidity sensor (Dennett, 1989), any shaking of the correct frequency from any source will be sufficient to make someone "think" he or she is nervous, if only for a moment (until the person can appraise the source of the feelings he or she is having). Accordingly, an interesting question for ongoing research is the degree to which there is a close or distant relationship between the ability to apprehend emotional and intersubjective complexity in film—it is possible that one trades off for the other, but it is also possible that these reasoning systems are to a degree separate, and therefore do not invoke costs upon each other.

Another means by which TOM supports emotional experience during film viewing may be immersion (Gross & Levinson, 1995; Visch, Tan, & Molenaar, 2010). During the viewing process, audiences are thought to engage with characters on an emotional level, experiencing the plot with the characters. Aristotle (1997) believed this *mimesis*—the perfect imitation of reality—begins in art and carries through into human emotion. At the apex of *mimesis* is *catharsis*, when the audience, mirroring the characters, achieves an emotional release at the end of the plot. It is a simple transition to place Aristotle within our modern understanding of TOM and perspective taking. As we better identify with a character's mental states, we better channel the intended emotions of the portrayed events.

However, just as the false belief task and complex intersubjectivity rely on forms of TOM that require considerable modification of basic inferences about beliefs, emotional film narratives are also highly dependent on interpretive processes. Although Aristotle looked down upon comedy as an art form, this genera should be of particular interest to modern psychologists. For instance, take a scene in which the protagonist is repeatedly humiliated, spurned by his family, ejected from his house, and finally arrested. On one view of Aristotle's approach to

emotion we should suffer alongside this protagonist. However, in the film *Meet the Parents*, the audience achieves the paradoxical state of identifying with the protagonist's suffering but realizing the comedy of the situation. Emotional states in film therefore cannot be simple mimicry, but are instead the result of perspective taking and contextual evaluation where the intent of the movie must be implicitly analyzed alongside the intent of the characters. This view would be consistent with modern cognitive appraisal approaches to emotion, and it is interesting to note that an understanding of film may inform these basic science theories. For example, genre identification may comprise an important context that could drive appraisal, facilitating the desired emotional response to the stimuli (Visch & Tan, 2009). It would follow that increased immersion would result in heightened awareness of emotions and emotional context (genre) within a film, a finding supported by Visch and colleagues (2010). Moreover, Tan's work suggests that the interplay between emotion and perspective taking may fuel our desire to watch films in the first place (Tan, 2008).

Conclusions

In the chapter, we have explored the links between theory of mind and film. Clearly, understanding TOM can help us understand why film is the way it is. However, it also appears as though there may be ways in which research on TOM and action perception might suggest new avenues for filmmakers to explore, and conversely that current practice in film can help inform the development of research on TOM. More generally, it appears as though a careful consideration of the nature of film, as opposed to written narrative, can help develop interesting research hypotheses about the relative costs and benefits of visual support for intentional narratives. Finally, we would argue that understanding the role of TOM in film can be sharpened by considering the action of this cognitive system in the context of story elements that are nonintentional. In doing all of this, we believe that the intersection of film and cognitive sciences is particular productive, as both fields are characterized by a set of well-developed principles that often serve as independent verification of the importance of specific psychological principles necessary to perceive and understand real-world events.

Acknowledgment

This material is based on work supported by the National Science Foundation under Grant No. 0826701 to DTL.

References

Abbott, H. P. (2008). Unreadable minds and the captive viewer. *Style, 42*(4), 448–467.

Aristotle. (1997). *Poetics.* (M. Heath, Trans.). New York, NY: Penguin Classics

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a 'theory of mind'? *Cognition, 21*, 37–46.

Baron-Cohen, S. (1997). *Mindblindness: An essay on autism and theory of mind.* Cambridge, MA: MIT Press.

Baron-Cohen, S. (2005). The empathizing system: A revision of the 1994 model of the mindreading system. In B. Ellis & D. Bjorklund (Eds.), *Origins of the social mind* (pp. 468–492). New York, NY: Guilford Publications.

Barrett, J. L., & Keil, F. C. (1996). Conceptualizing a non-natural entity: Anthropomorphism in God concepts. *Cognitive Psychology, 31,* 219–247.

Barrett, J. L. (2004). *Why would anyone believe in God?* Walnut Creek, CA: Altamira Press.

Barrett, J. L., & Lanman, J. A. (2008). The science of religious belief. *Religion, 38,* 109–124.

Berlinger, T. & Cohen, D. J. (2011). The illusion of continuity: Active perception and the classical editing system. *Journal of Film and Video, 63*(1), 44–63.

Bordwell, D. & Thompson, K. (2008). *Film art: An introduction* (8th ed.). Boston, MA: McGraw-Hill.

Butte, G. (2004). *I know that you know that I know: Narrating subjects from Moll Flanders to Marnie*. Columbus, OH: Ohio University Press.

Butterworth, G., & Jarrett, N. (1991). What minds have in common is space: Spatial mechanisms serving joint visual attention in infancy. *British Journal of Developmental Psychology, 9,* 55–72.

Campbell, J. (1949). *The hero with a thousand faces*. Princeton, NJ: Princeton University Press.

Ceci, S. J. (1990). *On intelligence—more or less: A bio-ecological treatise on intellectual development*. Englewood Cliffs, NJ: Prentice Hall.

Danto, A. C. (2005). The artist as prime mover: Thoughts on Peter Fischli and David Weiss' *The Way Things Go*. In R. Fleck, B. Sontgen, & A. C. Danto (Eds.), *Peter Fischli David Weiss* (pp. 90–105). London, England: Phaidon Press Ltd.

Dennett, D. (1989). *The intentional stance*. Cambridge, MA: MIT Press.

Dunbar, R. (2000). On the origin of the human mind. In P. carruthers & A. Chamberlain (Eds.), *Evolution and the human mind: modularity, language, and meta-cognition* (pp. 238–253). New York, NY: Cambridge University Press.

Epley, N. Waytz, A. & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review, 114,* 864–886.

Falck-Ytter, T., Gredebäck, G., & von Hofsten, C. (2006). Infants predict other people's action goals. *Nature Neuroscience, 9*(7), 878–879.

Farroni, T., Massaccesi, S., Pividori, D., Simion, F., & Johnson, M. H. (2004). Gaze following in newborns. *Infancy, 5*(1), 59–60.

Flavell, J. H. (2004). Development of knowledge about vision. In D. T. Levin (Ed.), *Thinking and seeing: Visual metacognition in adults and children* (pp. 13–36). Cambridge, MA: MIT Press.

Flavell, J. H., Green, F. L., & Flavell, E. R. (1990). Developmental changes in young children's knowledge about the mind. *Cognitive Development, 5*(1), 1–27.

Fowler, H. (2004). *Fowler's modern English usage* (3rd revised ed.). Oxford, England, and New York, NY: Oxford University Press.

Gao, T., Newman, G. E., & Scholl, B. J. (2009). The psychophysics of chasing: A case study in the perception of animacy. *Cognitive Psychology, 59*(2), 154–179.

Gergeley, G., Nadasdy, Z., Csibra, G., & Biro, S. (1995). Taking the intentional stance at 12 months of age. *Cognition, 56,* 165–193.

Gopnik, A., Slaughter, V., & Meltzoff, A. (1994). Changing your views: How understanding visual perception can lead to a new theory of the mind. In C. Lewis & P. Mitchell (Eds.), *Origins of an understanding of mind* (pp. 157–181). Hillsdale, NJ: Erlbaum.

Griggs, R. (1984). Memory cueing and instructional effects on Wason's selection task. *Current Psychology, 31*(4), 3–10.

Gross, J. J., & Levinson, R. W. (1995). Emotion elicitation using films. *Cognition and Emotion, 9*(1), 87–108.

Hannula, M. M., & Lehtinen, E. (2005). Spontaneous focusing on numerosity and mathematical skills of young children. *Learning and Instruction, 15,* 237–256.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *American Journal of Psychology, 57*(2), 243–259.

Hood, B. M., Willen, J. D., & Driver, J. (1998). Adult eyes trigger shifts of visual attention in human infants. *Psychological Science, 9,* 131–134.

Hymel, A. M., Levin, D. T. (2011). Detecting sequence disruptions within events is not automatic. Poster presented at the annual meeting of the Vision Sciences Society, Naples, FL.

Kinderman, P., Dunbar, R., & Bentall, R. (1998). Theory-of-mind deficits and causal attribution. *British Journal of Psychology, 89,* 191–204.

Kuhlmeier, V., Bloom, P., & Wynn, K. (2005). Do 5-month-old infants see humans as material objects? *Cognition, 94,* 95–103.

Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in "theory of mind." *Trends in Cognitive Sciences, 8,* 528–533.

Levin, D. T., & Angelone, B. L. (2008). The visual metacognition questionnaire: A measure of intuitions about vision. *American Journal of Psychology, 121,* 451–472.

Levin, D. T., Killingsworth, S. S., & Saylor, M. M. (2008). Concepts about the capabilities of computers and robots: A test of the scope of adults' theory of mind. *Proceedings of the 3rd Annual IEEE International Workshop on Human and Robot Interaction, 3,* 57–64.

Levin, D. T., Momen, N., Drivdahl, S. B., & Simons, D. J. (2000). Change blindness blindness: The metacognitive error of overestimating change-detection ability. *Visual Cognition, 7,* 397–412.

Levin, D. T., Saylor, M. M., Killingsworth, S. S., Gordon, S., & Kawamura, K. (in press). Tests of concepts about different kinds of minds: Predictions about the behavior of computers, robots, and people. *Human-Computer Interaction.*

Levin, D. T., & Wang, C. (2009). Spatial representation in film. *Projections: The Journal for Movies and Mind, 3,* 24–52.

McAlpine, L. M., & Moore, C. L. (1995). The development of social understanding in children with visual impairments. *Journal of Visual Impairment and Blindness, 89,* 349–358.

Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science, 198,* 75–78.

Minter, M., Hobson, R. P., & Bishop, M. (1998). Congenital visual impairment and "theory of mind." *British Journal of Developmental Psychology, 16,* 183–196.

Mithen, S. (1996). The prehistory of the mind. In P. Carruthers & A. Chamberlain (Eds.), *Evolution and the human mind: modularity, language, and meta-cognition* (pp. 238–253). New York, NY: Cambridge University Press.

Moll, H., & Tomasello, M. (2004). 12- and 18-month-old infants follow gaze to spaces behind barriers. *Developmental Science, 7,* F1–F9.

Morewedge, C. K., Preston, J., & Wegner, D. M. (2007). Timescale bias in the attribution of mind. *Journal of Personality and Social Psychology, 93*(1), 1–11.

orton, J., & Johnson, M. H. (1991). CONSPEC and CONLEARN: A two-process theory of infant face recognition. *Psychological Review, 98,* 164–181.

Münsterberg, H. (2001). Why we go to the movies. In A. Langdale (Ed.), *Hugo Munsterberg in film: The photoplay—a psychological study and other writings* (pp. 171–182). New York, NY: Taylor and Francis Group.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues, 56,* 81–103.

Palmer, A. (2004). *Fictional minds.* Lincoln, NE: University of Nebraska Press.

Slaughter, V. (1998). Children's understanding of pictorial and mental representations. *Child Development, 69,* 321–332.

Smith, M. (2004). Who's afraid of Charles Darwin? The art of film in the age of evolution. In M. Brutsch and V. Hediger (Eds.), *Emotion and film* (pp. 289–312). Zurich, Germany: University of Zurich.

Spelke, E. S., Phillips, A. T., Woodward, A. L. (1995). Infants' knowledge of object motion and human action. In D. Sperber, D. Premack, & A. Premack (Eds.), *Causal cognition: A multidisciplinary debate.* New York, NY: Oxford University Press.

Tan, E. S.-H. (2008). Entertainment is emotion: The functional architecture of the entertainment experience. *Media Psychology, 11*(1), 28–51.

Tomasello, M. (1995). Joint attention as social cognition. In C. Moore & P. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 103–130). Hillsdale, NJ: Lawrence Erlbaum.

Tomasello, M., Kruger, A., & Ratner, H. (1993). Cultural learning. *Behavioral and Brain Sciences, 16,* 495–552.

Varakin, D. A., Levin, D. T., & Fidler, R. (2004). Unseen and unaware: Applications of recent research on failures of visual awareness for human-computer interface design. *Human-Computer Interaction, 19,* 389–421.

Visch, V. T., & Tan, E. S. H. (2008). Narrative versus style: Effect of genre-typical events versus genre-typical filmic realizations on film viewers≠ genre recognition. *Poetics, 36*(4), 301–315.

Visch, V. T., & Tan, E. S. (2009). Categorizing moving objects into film genres: The effect of animacy attribution, emotional response, and the deviation from non-fiction. *Cognition, 110*(2), 265–272.

Visch, V. T., Tan, E. S., & Molenaar, D. (2010). The emotional and cognitive effect of immersion in film viewing. *Cognition & Emotion, 24*(8), 1439–1445.

Wellman, H. M., Phillips, A. T., & Rodriguez, T. (2000). Young children's understanding of perception, desire, and emotion. *Child Development, 71,* 895–912.

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition, 69,* 1–34.

Woodward, A. L. (2005). Infants' understanding of the actions involved in joint attention. In R. V. Kail (Ed.), *Advances in child development and behavior* (pp. 229–262). Oxford, England: Elsevier.

Zunshine, L. (2007). Why Jane Austen was different and why we may need cognitive science to see it. *Style, 41,* 275–299.

4

Driving Emotions With Movies