Circular Causality of Emotions in Moving Pictures

Mircea Valeriu Deaca
University of Bucharest (Romania)
E-mail: deaca@dnt.ro

Abstract. In the framework of predictive coding, as explained by Giovanni Pezzulo in his article *Why do you fear the bogeyman? An embodied predictive coding model of perceptual inference* (2014), humans construct instances of emotions by a double arrow of explanation of stimuli. Top-down cognitive models explain in a predictive fashion the emotional value of stimuli. At the same time, feelings and emotions depend on the perception of internal changes in the body. When confronted with uncertain auditory and visual information, a multimodal internal state assigns more weight to interoceptive information (rather than auditory and visual information) like visceral and autonomic states as hunger or thirst (motivational conditions). In short, an emotional mood can constrain the construction of a particular instance of emotion. This observation suggests that the dynamics of generative processes of Bayesian inference contain a mechanism of bidirectional link between perceptual and cognitive inference and feelings and emotions. In other words, "subjective feeling states and emotions influence perceptual and cognitive inference, which in turn produce new subjective feeling states and emotions" as a self-fulfilling prophecy (Pezzulo 2014, 908). This article focuses on the short introductory scene from Steven Spielberg's *Jaws* (1975), claiming that the construction / emergence of the fear and sadness emotions are created out of the circular causal coupling instantiated between cinematic bottom-up mood cues and top-down cognitive explanations.

Keywords: emotions, emergence, embodiment, predictive coding, Bayesian inference.

The propositions and analyses presented in this study originate from the embodiment-of-the-mind hypothesis, i.e. the enactive approach in which cognition is the enactment of the world and the mind (Maturana and Varela 1992; Varela, Thompson, and Rosch 1991; Johnson 1987, and Thompson 2007). The central idea of the embodiment hypothesis is that it is unproductive to dissociate mind and body, when we are speaking about mental phenomena. The mind and brain-body support and guide each other. The patterns and processes our body

is familiar with, are constantly mapped upon the stimuli provided by the world. Humans engage with and grasp patterns, making sense of the surroundings by projecting the body relations and processes they experience at a conscious and unconscious level.[1] For instance, body-based orientational schemas provide interactive engagement with the world. As a consequence, "the significance of a specific object will be established by its affordances for possible sensory-motor-affective engagements with creatures of our particular bodily makeup, need, desires, and goals" (Johnson 2015, 10). At a fundamental level, cognition and sense-making are emergent properties of a body-brain-mind complex that thinks in the world. Cognition depends on embodied action, whereas meaning is emergent from the basic-level embodied patterns of interaction. Embodied meaning is linked to the motor and affective engagement with the world.

Meaning is derived from correlations and cross-mappings between domains of experience. Skeletal patterns or schemas in addition engage our emotions.[2] One such pattern is the *control cycle* cognitive model. The model was used extensively in the Cognitive Grammar developed by Ronald W. Langacker (1987; 2008; 2009). The model has several in-built stages that define the interaction between a region of internal control and access and an external entity, a target defined in space and time. The model is a hard "glued" gestalt-like composite that enlists several image schemas components [Fig. 1]. Langacker summarizes the basic form of the control cycle as follows: "In the static baseline phase, an *actor* (A) (in a broad sense of the term) controls an array of entities (small circles) which collectively constitute its *dominion* (D). In the next phase, some *target* (T) enters the actor's *field* (F), or scope of potential interaction. This creates a state of tension, for the actor has to deal with the target in some manner. The typical means of dealing with it is by somehow bringing it under the actor's control, i.e. exerting force (double arrow) resulting in its incorporation in the actor's dominion. The result of this action is a modified situation that is once more static (a state of relaxation)" (2009, 130, 306).

---

1   As Wojciehowski and Gallese (2011, 16) indicate, embodied simulation hinges on the immediate and involuntary mirrored experience, i.e. "our brain-body system re-uses part of its neural resources to map others' behavior. When we witness actions performed by others, we simulate them by activating our own motor system. Similarly, by activating other cortical regions we re-use our affective and somatosensory neural circuits to map the emotional and somatosensory experiences of others."

2   Lakoff argues that the "human brain is mapped by thousands of embodied metaphor mapping circuits" and "primitive concepts have a schema structure that mediates between embodiment circuitry and complex concepts that are expressed by linguistic structures in natural language" (Lakoff 2014, 6).

The present study explores several hypotheses: cinematographic artifacts mirror the patchwork of the brain's functional systems, which are engaged in an interplay of competitive and cooperative interactions (Thompson 2007, 206); dynamic hierarchies are established ad-hoc once the attention mechanism targets one layer or another and one system takes control of the others; a dynamic pattern –*the control cycle* – drives the behaviour of neural populations. Sense making in filmic contexts involves a circular coupling between two or several control cycles. The interaction between the coupled control cycles can be external-bounded or nested. Its behaviour has an exploratory-projection facet and a phagocytosis introjective aspect. The behaviour of the control cycle is manifest in low-level feature-detectors and high-level cognitive functions that exert top-down modulations.

The sense-making process is based on a biological-like network of control cycles that are engaged in a continuous interexchange / trade-off of "reciprocal parasitism" and hierarchical coupling structures. Life cycles or dynamic processes of coupling between self and environment give raise to different domains of interaction. Networks of coupled and nested control cycles weave the conceptualization fabric. Coupling takes place when two structures are enmeshed in a dynamic process of bidirectional causation, which generates a functional whole (i.e. a stable state, pattern or shape). Coupled control cycles – based on bidirectional causation – generate functional wholes, i.e. stable states, patterns or shapes in domains of interaction in a bottom-up manner. The emergent patterns constrain, in turn, in a top-down way the components – such as boundaries that – are created. Connectivity and coupling take place in closed systems at different spatial and time-scales. Fine grained networks of reciprocal coupling – circular loops – emerge as coarse units at higher levels. Subsequently, these units that have their own "rate of change, rhythm, and duration" behave as components in new emergent units that subsequently constrain their behaviour (Kirchhoff and Kiverstein 2019, 21). Stacked patterns form a hierarchy of layers, which exist at different but synchronized time scales. Patterns of coupling can take place at the same level or between different levels [Fig. 2].

The Elaboration Site and the Kuleshov Effect

Relationships of schematicity are useful in Cognitive Grammar for identifying syntagmatic valences of elements. Sentences contain some (sub)structure in one component structure that corresponds to an entity or a substructure in its vicinity

that serves to elaborate, i.e. characterize in finer grained detail (Langacker 2008, 198; Tuggy 2007, 105) The schematically characterized substructure is an elaboration site or an *e-site.* The e-site is a skeletal (coarse grained) or summary-like representation of a concept (Sapolsky 2017, 717). We can envision the e-site as a kind of "motor intention" (a "spatial index" or a "deictic pointer") that belongs to an array of associated concepts that are aggregated by an automatic process.[3] The e-site represents a schematic region in the situation evoked by a simulation of a state of affairs and brings sense-making control. In a sense, e-sites create, in advance, the continuation of the discourse that is extended over the expressive boundary of a channel of conceptualization, e.g. the word or sentence and the cinematographic frame of attention or shot. The brain's activity is ahead of our mental states and our subjective experience is a post factum rationalization of antecedent choices.

Hence, we start the cinematic analysis from the assumption that the meaning constructed by the viewer from a sequence of constituent shots is a symbolic construction different from the meaning of each shot considered individually (Anderson et al. 2006). The conjugation of shots A and B constructs a conceptual unit different from A and B separately. For example, the gist of the scene evoked in shot B in [Fig. 3] is a constructed simulation that contains an e-site (e.g. human affective behaviour in high arousal situations) that categorizes and is elaborated by the perception of a facial expression in shot A (Barrett 2017, 43). The image in B elaborates a winner role in a tennis sports event. The affect e-site constrains the array of simulations around the intensity value of the feeling, but is indeterminate as far as the hedonic value is concerned. The feeling is "high arousal" but the value of pleasure/displeasure is indeterminate, i.e. has an open range of potential values. This open-ended set of potential simulations is the e-site that awaits elaboration.

Understanding the meaning of the scene depicted in shot B and constructing a simulation of the scene gives access to an e-site for categorizing perceptual information in shot A. In other words, "if one explains verbally the first photo's context – winning a crucial tennis match the viewer's brain applies conceptual knowledge of tennis situations and winning competitions situations in order to simulate/ evoke facial configurations of people experiencing exultation" (Barrett

---

3   A "spatial index" or a "deicitic pointer" can be used by visuomotor routines in order to situate the object's location in the environment: "if this spatial pointer is associated with some kind of coarse semantic information, for example, a pattern of activation in one of the language cortices, auditory cortex, or even visual cortex, then the spatial pointer can be triggered when sensory input activates that semantic information" (Spivey 2007, 295, 299).

2017, 43). An e-site is a shorthand that holds together bits of pieces of suffering/exultation situations memorized by the viewer. Based on this summary of summaries, the brain of the conceptualizer can predict and explain the perceptual stimulus from shot A. The e-site from shot A constructs in a top-down manner the perceptual and content information and the perceptual information in shot B elaborates in a bottom-up fashion the information about context: winning situations and appropriate behaviour and emotions.[4]

Filmic Kuleshov effects are based on the same mechanism. Two adjoined shots generate an emergent multimodal blend and can mix beliefs and expectations with the perceptual processing of sensory signals. Cognitive processing makes use of a combination of top-down knowledge and bottom-up sensory signals in order to construct a concept just enough for the subject's goal in context (Kirchhoff and Kiverstein 2019, 87).

The Thief vs. Wind Case and Circular Causality

The following argument describes the mechanism of predictive coding as explained in Pezzulo (2014). The top-down predictions of causes that explain sensory stimulus are coalitions of domains. Much like situated conceptualizations, the predictions are models of the world that include aspects of experience, properties of objects, exemplars and prototypes, relevant relations and rules, actions and affordances,[5] the setting, agents and their behaviour, perceptions, and bodily as well as mental states (emotional and cognitive operations). These predictions described by the model can often be mapped upon the introspective states experienced by the human experiencer. The top-down predictions can be acquired through cultural learning (instances of descriptive systems) but "still remain grounded through the linkage with lower-level events" (Pezzulo 2014, 906). The predictive coding "cascades" at multiple levels, replicating the prediction-error minimization process for each pair of consecutive levels (906). The winning hypothesis is the one that produces the least error in the whole hierarchy.[6] The following example is discussed by Pezzulo: "if you hear a window squeaking in the night after watching a horror movie, you may consider plausible

---

4   See about top-down effects of expectation on perception also in Barsalou (Barsalou 2016, 22).
5   See Pezzulo: "action-based approaches to cognition argue that objects can be recognized in terms of the expected sensory consequences of possible actions produced by forward models; for example, a sponge can be understood in terms of the (anticipated) softness when squeezing it or imagining squeezing it" (Pezzulo 2014, 908).
6   See the description of the mechanism as *free energy minimization* in Friston (2005; 2010).

a very unlikely hypothesis (e.g. a thief, or even the bogeyman) because it explains both what you sense (e.g. the window squeaking in the night) and how you feel (e.g. your high heart rate)" (2014, 902).

Arousal can be construed as an instance of emotion. Feelings and emotions depend on the perception of changes in the body (Seth, Suzuki, and Critchley 2012). Sapolsky describes vividly the mechanism in which interoceptive information influences, "if not determines" sense-making and emotions (2017, 94). For instance, "you subliminally note the lion, speeding up your heart; then your conscious brain gets this interoceptive information, concluding, 'Wow, my heart is racing; I must be terrified.' In other words, you decide what you feel based on signals from your body" (Sapolsky 2017, 94).[7]

According to this way of thinking, in "a thief intruding in a house" scenario the interoceptive signals of arousal are categorized as an instance of "fear." The top-down conceptualization of the thief scenario also includes the interoceptive information of the correspondent emotion of fear. If our subject is alone in the night, at home, and hears the sound of a window squeaking *and* if (s)he evokes the scenario that a probable thief intrudes, (s)he will categorize the interoceptive information as an instance of fear. Recursively, the state of arousal is evidence that explains the thief-hypothesis and, as a consequence, reinforces the thief scenario. A mood previously created (a car accident or a horror movie) can generate the arousal (heart rate or sweating) in the first instance and prime the categorization. Emotional responses persist longer than the emotional stimulation phase (Smith, 2003).

As Pezzulo remarks, confronted with uncertain auditory and visual information a multimodal internal state assigns more weight to interoceptive information (rather than auditory and visual information) or visceral and autonomic states as hunger or thirst (motivational conditions) (2014, 905–907). The thief hypothesis explains more satisfyingly the evidence coming from the most reliable source, i.e. interoception, and "the idea of a thief seems so vivid that you can almost see a silhouette moving in the shadows" (Pezzulo 2014, 907). The wind hypothesis where the movement of the window is caused by the wind explains only the less reliable evidence, i.e. auditory and visual. This closely resembles the recipe of the horror movie where the dangerous entity is depicted by uncertain cues on a background of emotional arousal.[8] Is the off frame sound a symptom of the

---

7   "(a) forcing depressed people to smile makes them feel better; (b) instructing people to take on a more 'dominant' posture makes them feel more so (lowers stress hormone levels); and (c) muscle relaxants decrease anxiety" (Sapolsky 2017, 94).

8   See Carroll who speaks about the "*criterially prefocusing* scenes and sequences in which the

intruder? Is the movement of a door knob caused by the intruder? Is the shaky camera focusing through the foliage the POV of the entity or is it just the cinematic narrator's vantage point?

In this context, the extra diegetic musical score serves as a source of arousal and mood constraint. The double bind can be obtained on this background affect (intensity of arousal and valence of the affect). As visual and diegetic auditory cues are uncertain, i.e. they can belong to different situated conceptualizations, the background emotional arousal is more reliable as evidence for the scenario of the "dangerous entity."

Pezzulo remarks that "if we consider interoceptive information to be a source of evidence just as we consider sensory events, it naturally follows that it can influence perceptual inference, belief formation, and choice" (2014, 905). In the example at hand, "the way that the horror movie or the small car accident influences the current (wind vs. thief) inference is through the body state" (2014, 905). Interoceptive information is considered by the brain a high reliable source and "perhaps this is why the bogeyman only visits us in the dark" (2014, 905).

Let us briefly recapitulate. The thief situation categorizes the arousal cue as a symptom of the fear emotion *and*, in turn, the fear construct categorizes the cue sound event as a fragment of the thief intruding conceptualization. A circular loop of control is set up [Fig. 4].

1. A schematic component of the emotion category of fear – a schematic arousal symptom – is elaborated by the thief intrusion conceptualization as an instance of fear emotion. The state of arousal explains (is a model that causes) the thief scenario as one instantiation of the goal concept of things that generate fear (the "fear of thieves" or a "thief fear"). The "thief fear" is a constructed emotion. The verbal structure would be: *Why is there a thief? Because I have an emotional state and a sound is produced.*
2. Schematic/summary components of the thief scenario – the sound of the window and the fear emotion – are further elaborated as the arousal caused by the intrusion. The arousal is one instance of the fear emotion. The thief scenario explains the state of arousal. The verbal structure would be: *Why do I feel aroused? Because there is a thief intruding.*

eliciting factors pertinent to provoking a certain emotion are made salient, typically by means of variable framing" (Carroll and Seeley 2013, 68). See also Greg Smith theory about the cues that enhance a specific mood in film upon which an emotional episode can be grounded (Smith 2003).

The thief scenario contains the schematic emotion of fear that categorizes the sound. The emotion of fear contains the schematic situation of an unexpected danger causing the emotion that categorizes the sound of the window as an element of the thief scenario. The thief situation is an instance of fear. This mechanism introduces a *circular causality* in which "sympathetic arousal (for example) can be both a cause and consequence of emotional or salient perceptual states" (Pezzulo 2014, 907).

The mechanism is described as: "the belief that a thief has entered the home produces predictions about elevated heart rates that are fulfilled automatically, through sympathetic reflexes. These, then, may reinforce the embodied predictive coding of uncertain auditory cues and produce cascading effects that ultimately reinforce the bogeyman belief" (Pezzulo 2014, 907). The idea of a bogeyman is "a form of self-fulfilling prophecy" because "a terrified child can take his or her terror as evidence that the bogeyman exists (and is probably close), and the terror itself can increase due to the circular causality mentioned earlier" (2014, 909).[9] Part of the conception of the bogeyman is construed out of cultural information ("'tales' heard from parents or friends"). The concept of the bogeyman includes representation of entities (wind, thief) and interoceptive information (2014, 908).

Performative Discourses

Pezzulo's approach can be translated in the framework of mental spaces and conceptual integration theory (Fauconnier 1994; 1997; Fauconnier and Turner 1998; Fauconnier and Turner 2002). Depictive utterances are "attempts to make one's words fit the world [and], performative utterances are attempts to make the world fit one's words" (Coulson and Oakley 2000, 184). Metaphor involves the interplay between a source mental space – the space of representation – and a target mental space – the represented world. Some metaphoric construals instantiate the circular coupling: "if the representation is taken as fitting the represented space, then the relation between the spaces is *depictive* or representational. [….] If, on the other hand the represented space is taken as fitting (being causally influenced or changed by) the representation, then the relation is performative. The act of representation, by its performance, constitutes (or performs as a causal agent in) the structure of the represented space" (Sweetser 2000, 312).

---

9   Pezzulo mentions that this idea resonates with the argument that some abstract concepts (including emotional concepts) can be grounded in interoceptive information (2014, 909; Vigliocco et al. 2013).

Sweetser gives the example of a "painting of a buffalo hunt on a cave wall" (Sweetser 2000, 312). Is the painting a depictive record, "following and modelling itself on events in the represented world" or is it a performative one that is magically intended to "bring about in reality the situation it represents"? (Sweetser 2000, 312). This dual nature of representations is considered to hold for metaphorical representations and literal ones, too. Sweetser considers that, in rituals, the same representation is simultaneously depictive and performative: "does kneeling to a divinity metaphorically represent the already extant differential in power and status between worshipper and god (a depictive use), or help to bring the worshipper into the right state of humility (a performative use)? Perhaps both" (Sweetser 2000, 315). Kneeling in front of the monarch is a depictive recognition of status, and is a performative use of signs that causally inflict a change of status of the individual. For instance, "white clothing is worn by brides to depictively represent their innocence, and performatively by penitents in the hopes it will help to purify them" (Coulson and Oakley 2000, 185). In this case, the direction of fit is the direction of control. A circular loop is instantiated between the two spaces integrated in an emergent whole. Propaganda films are salient examples where depiction of a state of affairs is simultaneously the performative construction of a new state of affairs in the world. Metafilms also blend a depictive and a performative use.[10]

Force Dynamics of Emotion: "Fear Is an Opponent in a Struggle"

Let us turn now to a filmic example. Fahlenbrach (2008) analyses the spatial motif of the maze in *The Shining* (Stanley Kubrick, 1980) that creates the impression of threat representing by metaphorical mapping the cause of the emotion and the proper state of the emotion. The spatial features of the maze invoke the trapping, i.e. the sentiment of being captured by an external entity. The closed-isolated and labyrinthine space of the hotel in which Jack, Wendy, and their child live during winter profile the anxiety generated by the *hidden menace*. The correspondent metaphor is "fear is a hidden enemy" (Kövecses 2000, 23). Avoidance of capture is invoked in the stairs episode when Jack threatens and "pushes" Wendy up the stairs [Fig. 5]. For Fahlenbrach the stairs and "the camera, the editing, and the sound concretely mediate in their audiovisual gestalt the opponent metaphor: 'fear is an opponent in a struggle'" (Fahlenbrach 2008, 93). Character movement and facial expressions, as well as movement of the camera towards the victim

---

10   See the analysis done by Roger Odin (2000, 113–125) of *Le Tempestaire* (Jean Epstein, 1947).

and editing activate sensorimotor simulations. In terms of the grounding of the emotional concept of fear vs. a predator's rage in bodily movement, it becomes apparent that the approach/avoidance action is highly imbued with emotional valence (Pecher et al. 2011, 225).

The high pitch of Wendy's voice and the dissonant extra diegetic music coupled with the slow advancement of the fear factor intensify the feeling of tension due to entrapment and the impossibility of avoidance/escape out of the grip of the emotional state (Fahlenbrach 2008, 94). The episode of Jack menacing Wendy with a baseball bat generates the "cross-modal matching of duration, tempo, and intensity in pictures and sound" (2008, 94). Fahlenbrach concludes that "space, physical force, and movement are key source domains in audiovisual media for the metaphorical conceptualization of emotional states, causes, and coping strategies" (2008, 94).[11] The dissonant sound of voice and music create a negatively valenced feeling of irritation, which is categorized as an instance of fear. An intense emotion is correlated with sound intensity and pattern. Chattah (2015, 86–88) talks about the correlations that map loudness and crescendo onto psychological tension, pitch frequency to psychological tension, and timbral distortion to states of distress.[12] Sheets-Johnstone (1999, 455) advocates the idea that "emotion arises out of or from motion." Hence the qualitative affective-kinetic dynamics of "felt dynamic stirrings, felt inner commotions – a bodily 'earthquake' as it were, spanning a strikingly varied range of possible dynamics and thereby a strikingly varied range of possible magnitudes or intensities" (Sheets-Johnstone 1999, 456). Following this point of view, one can consider that emotions elaborate in detail the force-dynamic phase of the control cycle.

We observe that the episode is structured by a bidirectional categorization trajectory. The first direction is traced from content elements. The situation depicts the action of capture performed by an aggressive and menacing agent and the incapacity of the victim to escape. The second direction emerges from the schematic qualities of slow movement and tone of voice contrasted with the shrill voice of Wendy, her jerky gestures of defence, and the dissonant sounds of extra diegetic music. Slow movement of advance, the capture-like gestures of Jack, and his decreased tone of voice enhance the suspense created by the expectancy of surprise. The slow pace postpones the action phase and extends

---

11   "By relating typical acoustic attributes of objects or figures with general concepts and categories, sound design systematically construes metaphorical structures that use universal and kinaesthetical patterns of experience like the force-schema or the path-schema" (Fahlenbrach 2008, 97).

12   See Chattah (2015) on the cross modal metaphoric mappings.

the tension phase of the control cycle. The cinema viewer is in the position of the victim since (s)he is, much like Wendy, unable to escape the experience of a dangerous entity advancing towards the optical point of reference, i.e. the screen. Both the viewer and Wendy are trapped in a simulated "freeze" mechanism of fear. Circularity is at work here; more fear (instantiated as the *aggression scenario*) means more freezing, which, in turn, certifies the presence of more fear.

The Aggression Scenario and the Circular Loop: "Fear Is a Hidden Enemy"

Let us see the case of a visual mediated representation, which depicts a young woman in a dangerous situation. I divided in two parts the introductive scene from *Jaws* (Steven Spielberg, 1975) that depicts the aggression event suffered by a young swimmer. The introductory part (A) takes place before the attack of the shark (shots A1 to A7) and the second part (B) depicts the actual attack and the closure of the event (shots B8 to B15) [Fig. 6]. In this scene, the *fear* emotion is cued by the musical score, by the "unusual" vantage point used in shots A3 and A6, and by the opacity of the water. The opacity of the water is the *hidden menace*. The *hidden menace* is a situation similar to the *bogeyman* menace (uncertain visual and auditory cues combined with interoceptive feeling). The background mood is one of arousal due to expectancy and interest: "what will happen next?". Fear also evokes the *aggression scenario*, where the agent of the control cycle is a threatening entity, i.e. an aggressor that will control or entrap the target/victim. In light of the previously presented argument, one can note that:

> 1. The fear emotion contains a schematic arousal combined with a negative valenced feeling that explains and is elaborated by the *aggression* scenario instantiation and by the *hidden menace*. A series of cues (opacity, sound design, unusual framing) instantiate the emotional state of distress.
> 2. The *aggression scenario* contains a schematic fear emotion that explains and is elaborated by the arousal of the fear emotion. This time the same expressive cues are interpreted as instantiations of the *aggression* scenario.

The elaboration of fear

Fear is elaborated by the setting and the agents involved in the attack scene. Both the *aggression scenario* and the *hidden menace* evoke an agent, a harmful action,

and a target-victim. The emotional markers are clustered around the agent: focused attention or violent action, and around the target: fear, and agony. In the *hidden menace*, the agent is characterized as the unseen danger. The absence of visual cues that would display the perceptual features of the menacing entity increases the emotional intensity of fear (of the uncontrollable unknown) that, in turn, certifies and explains the menace.[13] The musical extra-diegetic score, the point of view tracking shot of a virtual entity blended with the water, the track forward of the point of view of a virtual entity, and the opacity of the water that hides from view a potential menace are categorized by the *aggression scenario* and the *hidden menace* situation as instances of fear.

The elaboration of the aggression and the hidden menace

Paradoxically, the fact that the victim is unsuspicious of the lurking danger reinforces the hidden threat. The unsuspecting young woman swimming on a clear night in calm waters is evidence that a threatening agent and possible danger are "hidden" from view. Being out of reach and beyond the visual field reinforces the idea of the hidden aspect of danger. The fear provoked by the menacing entity and experienced by the victim is conceptualized by the viewer as evidence of the *hidden menace* and *aggression scenario*. The agent of the aggression scenario/the agent of the menace is elaborated by the underwater POV and by the fear that is mildly suggested by a phobia of water (unclear expanse of water, fear of drowning and suffocation). In terms of the control cycle, what is unknown/uncontrollable creates distress and fear. The phobia is evoked by the difference in scale between the human and the natural in A1/ B 15 and by the underwater POV in A3/ A6. The underwater POV – in terms of the control cycle – is a form of perceptual and epistemic capture of the woman. The agent of aggression is elaborated as a pure attentional focusing on a victim and by the lack of features of the agent performing the viewing action. The aggressor's features are metaphorically reduced to movement, darkness and an action of visual capture in a boundless expanse (the depth of the water). The absence of a body is still an elaboration of the aggressor, e.g. a bodiless aggressor, and, due to the "bogeyman" emotional effect, a powerful emotional construct. The constructed emotion of fear explains, as a prediction hypothesis, the aggression.

The choice of the POV framing in shot A6 elicits disgust. The viewer is cued to adopt the vantage point of the bodiless menace. The evocation of

---

13   See about fear and anxiety in Bishop (2007).

drowning, suffocation, absence of light, and falling down (or being engulfed in life threatening water) elicits fear of contamination and internal harm due to an unknown agent, i.e. disgust. The viewer is captured by the vantage point of the hidden menace and is contaminated by a harmful agent from "within," i.e. from his inescapable own perceptual activity. The interoceptive affective quality of disgust is attributed to the bodiless menacing entity. Disgust predicts and explains the disgusting character of the shark. The description model/prediction of the disgusting shark explains the network of evoked components – drowning, suffocation, fall, opacity – as instantiations of fear and disgust. Elicited disgust reinforces and explains the hidden menace and the aggression scenario.

At the same time, the cinema viewer is in the position of not being in control of the situation. He/she cannot change the future fate of the victim and (s)he cannot avoid the simulation of the aggressor's state of mind. To contemplate the aggression situation from the vantage point of the unknown (bodiless) aggressor projects the viewer in a disgusting moral situation. Visceral repulsion of contamination by an unknown entity and moral disgust elicited by being in the position of the aggressor are mingled. The cinema viewer is threatened to be captured by an unknown (loss of identity and body) evil (loss of moral stance).[14] In brief, the viewer is projected in the victim's role, i.e. a target that is captured by another agent.

The *aggression scenario* is further elaborated in the second part as action and movement of the victim's body (B10–B11), close-up on the expression of fear (B8–B12), aggressive water movement (B9–B11), and the final stage of the capture, in which the victim entity is "absorbed" in the aggressor's control field of control (B13–B15).

Mixed Feelings

In a mediated representation, the situation of a character can be construed from an internal-immersed perspective (processed by an *empathic distress* brain system – "feeling with" or empathy) and from an external perspective processed by an *empathic care* system – "feeling for" or compassion (Andrews-Hanna et al. 2017; Raz and Hendler 2014). The two construals are entertained by the viewer in a dynamic unstable state. The viewer-conceptualizer can, by phase transition, shift (change his mind) between the two alternate perspectives. The

---

14   "The core of the intermixing of visceral and moral disgust is a sense of threat as well" (Sapolsky 2017, 576).

overall conceptualization can keep both alternate construals in tight balance in a dynamic bidirectional link. Shared components are schematic slots that can be elaborated by one or another construal, and facilitate transition. For instance, the ego-like point of origin of the cinematic viewing act is one of them, i.e. the cinematic viewers. The cinema viewer-conceptualizer temporarily simulates the character's immersed perspective in the situation and, therefore, will be called here the *experiencer* (the viewer, who experiences empathy "with") while an external conceptualizer's allocentric perspective of the situation will be called here *narrator* (the entity which experiences sympathy "for").

What about the interplay between *empathy* (*empathic distress* or "feeling with") and *sympathy* (*empathic care* or "feeling for") experienced by the cinema viewer? The swimming woman scene in part A is apprehended by three conceptualizers, each with his/her own vantage point and epistemic stance: the extra diegetic narrator, the intra diegetic experiencer (the young woman) and the intra diegetic aggressor (the shark). The shark will be left aside for the moment, since its main characteristic is its hidden profile and, in terms of the emotional engagement of the viewer, it is a secondary character. It is an important instantiation of the agent in the aggressive action scenario, but the movie profiles more insistently in the foreground of conception the emotional display of the woman. The cinema viewer is "aligned" with her stance on the situation (Murray Smith 1995). This does not mean that the viewer cannot empathize with the aggressor's stance and enjoy the capture of the victim in a hunting action.

Shot A5 [Fig. 6] is a visual expression that contains two vantage points and two conceptualizers of the scene: the extra diegetic narrator, and the intra diegetic experiencer. The verbal expression that describes the scene can be: *The woman is sure there is no danger. But there is danger.*

The *feeling for* and the *feeling with* are constructed in a symbolic unit, which constrains their subordination relationship. The narrator's vantage point and conception of reality is the matrix clause, while woman's conception of reality is the subordinate clause. A control cycle binds the clausal relationship between the two conceptions of reality and elicited emotions. The *feeling for* is compassion for a potential victim in an aggression scenario. This conception is guided by the narrator role and is a region of epistemic stance and ego-like role. It is a matrix clause in this construction, in which the narrator does not necessarily embrace the woman's point of view; (s)he just describes it. The *feeling with* is embedded in the matrix conception as a complement clause. The viewers simulate the vantage point of the woman. They perceive her happiness as she enjoys swimming. But

they also anticipate the virtual suffering of the woman (non-real – possible and non-immediate – in the future). Her possible suffering is also simulated, but projected in the virtual. Affective qualities can be attributed to future states and objects (Russell 2003, 147; Friston 2017).

The anticipated suffering is unavoidable, since the viewers can't intervene, help, and interrupt the aggression act. As a schematic control cycle in which an agent-aggressor does a harmful act to a target-victim, the aggression scenario is evoked. The woman instantiates the schematic target of the *aggression scenario*. As the control cycle is a 4-stage process, the simulation will contain the victim's both emotional stances. The unsuspecting victim stands for the potential stage situated at the beginning of the control cycle, followed by uncertainty, doubt and prediction testing in the tension phase as well as rage, distress or suffering in the action phase, ending with either extinction or relaxation due to escape in the last phase. The whole cycle is evoked by shot A5, i.e. the woman is blissfully ignorant of the danger. The viewers simulate the happy present stance of the woman (*a happy woman is sure that there is no danger*) *and* the future virtual suffering of the woman (*but there is danger*). The viewers "know" that there is fear and an aggressor in the scene and they know that the woman is the victim. In a sense, her present happiness cues future unhappy circumstances - the merrier she is now, the worse one expects her future to be. The whole sequence of stages of the event of aggression capture unfolds and the cinema viewer can simulate the distress from an immersed perspective *before* it actually takes place. The present moment is apprehended in the time lapse of the whole emotional construal of the control cycle. The present emotion of cheerfulness is unfolded and warped as fear, rage and agony due to the expansion of the control cycle. The emotion simulated by the viewer is a dynamic event that starts from a relaxed state and evolves into an emotion of fear and agony.

The whole event of aggression is invoked. The event is apprehended from the vantage point of compassion guided by the narrator and from the nested subordinate perspective of the victim. The emotional state of the woman that blends happiness and distress is experienced in a sequential order. One state seamlessly translates into the other generating a hybrid whole. The narrator's emotional vantage point prevails, and is correlated with the future fear of the victim. The future fear of the victim is the present fear for the fate of the victim. Her future fear is the viewer's present fear. Emotions are blends. They are constructions or "synthesized" blends. Different coalitions of neuronal networks are involved in the emergence of one emotion instance and specialized areas participate in multiple instances

of emotions. Blends can have partitions that are synthesized/fused together by a control mechanism into higher order conceptualizations. The conceptualization of a scene can contain two partitions: one is an allocentric view and the second one an egocentric perspective on the situation.

Sadness and Melancholia

The discrepancy between the perspective of the narrator for whom visual cues are mapped as instantiations of the fear emotion and the *aggression scenario* (the future rage and agony), and the joyful perspective of the woman creates an emotion of sadness. Sadness is cued by the impossibility of avoiding the outcome of the aggressive event; it is also due to the apprehension of the fact that the woman is unaware of the cues that are around her in the scene, and that it is impossible to change this situation. The woman is doomed to be a victim in a scenario. She is unable to see the signs announcing the future aggression. For her, the sea water is enjoyment, but the viewers perceive, feel, and know that the water is a hidden menace.[15] The viewers feel compassion not for the actual suffering of the woman but for a human being, who is unaware of an already doomed destiny. They have a meta-emotion directed at their incapacity to help. The viewers feel sadness, because they are trapped in the victim's condition. The woman's fate is doomed, since nobody will change her fate at the end of the scene. The viewers, in spite of being aware of the nefarious future, cannot alter the predicted sequence of events. The menacing factor is instantiated by the fear emotion and by the amorphous substance of the water that impedes viewing. The viewers witness the tragic fate of a woman, who is absorbed in a painful and agonistic way in a homogeneous substance (an indefinite mass substance), and, in a metaphorical mapping, witness also the "erasure" of her image, i.e. shot B13. What is felt and categorized as a threat creates a state of tension and fear. The diegetic water and the visual mass of indefinite water dominate almost all the images. The horizon line is extremely high placed and creates a sense of cramping, suffocation and claustrophobia. Water engulfs the character and the visual blur covers the transparency of the image with a relentless tense movement and blocks visual access to the scene. The perspective taken involves more than simulation and is a task that activates an elaborate Theory of Mind. The viewer has to visualize the scene from another mind perspective (e.g. the narrator, the

---

15   In Bartsch (2010) we find the common metaphors for sadness: "drowning in a sea of sorrow," being "flooded with tears" associated to darkness, falling, and drowning (Bartsch 2010, 250).

shark), and has to infer the moral and social outcomes of the aggression scene.[16] Shots B14 and B15 mark a phase of relaxation. The line of horizon is lowered to a more classical composition.

To summarize, the episode condensed in shot A7 [Fig. 6] is a roller coaster of constructed emotions that dynamically unfolds and adopts several perspectives on the scene [Fig. 7]. The viewers entertain happiness *with* the actual woman, they experience compassion *for* her future and virtual tragic fate, entertain distress and pain *with* a virtual suffering and a molested human being, fear in an immersed/absorbed, but extra diegetic position in front of a hidden menace, and sadness before the inevitability of fate and loss. Shot B14 highlights the epistemic stance of unawareness. The young man lying on the shore is incapable of help and unaware of the tragic event. Sadness here mixes the "for" position with a human being, who cannot change the future events (the woman, e.g. *I feel sad for her*), with the self-condition since the viewer cannot change the script of a film (e.g. *I feel sad that I cannot change the course of events*). Bodily feelings of loss of control over one's own body are categorized as an instance of sadness. The bodily feelings that were categorized as fear at the beginning of the episode are now sadness. The viewer is captured by the control cycle of containment and restraint cued by the fictional world and the expressive channel of the filmic artefact. The emotional construction prompts further abstractness and can be apprehended as a generalization involving a moral and social reasoning, e.g. *the human condition is a sad one, since awareness of future events or lack of knowledge will not modify the predicted outcome*. The emotion belongs to the class of painful experiences, e.g. physical or psychic pain (social exclusion, anxiety, disgust, embarrassment) (Sapolsky 2017, 542). Sadness and melancholia are inward oriented and allow abstract thinking and social thinking (Wager et al. 2015; Andrews-Hanna 2012; Bartsch 2010; Guo et al. 2014, 7; Hyett et al. 2015, 665). The emotion constructed by the viewer does not have a conventional linguistic label and lasts as long as (s)he experiences the sequence of mental dynamic emergence. The emotional episode has an innovative, yet fleeting character. The viewer may keep a memory trace of it as a distinct instantiation of a type of fear-sadness.[17] As Murray Smith noted, emotions elicited by fictional artefacts have a "unique and original 'flavour'" dependent on the "circumstances evoked" by the fictional world (Smith 2017, 202). Works of art create distinctive emotional tones and blended constructs.

---

16   See Sapolsky (2017, 492).

17   Barrett mentions the emotion concept of "Fago" in Ifaluk (Micronesian) that can mean "love, empathy, pity, sadness, or compassion, depending on context" (Barrett 2017, 147).

Cinematographic discourse can construct hybrid entities that blend figurative features out of different source entities (e.g. Batman as man and bat) and thus, similarly construct new emotional blends in the domain of affective qualities.

Emotion elicited by fictional artefacts is inscribed in a circular loop. Emotions are caused by cues and stimuli of the fictional scene that categorizes them (e.g. the *hidden menace* and the *aggression scenario*), and, at the same time, explain or give reason for the affective stimuli of the fictional scene. We see that emotion, perception and cognition share common structures and are closely related. From a similar perspective, Smith, when commenting on empathy, draws the distinction between functional and aesthetic experience: "in life, empathy is always functional, a means to an end; in narrative art, it becomes an end as well as a means" (Smith 2017, 194).[18] As Aristotle recommended, Spielberg's story of a woman victim is for the cinema viewer a story worth telling, e.g. having a "cathartic" effect, because it depicts events in a way capable of generating pity and/or fear. Loss of clear boundaries between selves, e.g. the cinema viewer, the narrator, the character, obtained via the blending of the *feeling with* and *feeling for* creates the premise for compassion for "all sentient beings" (Varela, Thompson and Rosch 1991, 248). The sense of loss which, in a first stage, elicits sadness converts into compassion. Quite often loss of control perceived in mediated aesthetic artifacts, as Varela and colleagues note, is the simulation of loss of the ground of self or world. A groundless relationship between self and world leads to "nonegocentric compassion" (1991, 250). Letting go of the "grasping mind" (i.e. the mind that controls the world or the self as objective and permanent entities) leads to the ethics of a "transcendental generosity," and opens the mind of the cinema viewer to an open ended/unconditional "responsiveness to oneself and others as sentient beings without ego-selves" (1991, 251).

References

Anderson, Daniel R., Nicole Petrovic, and Joy Hirsch. 2006. Cortical Activation While Watching Video Montage: An fMRI Study. *Media Psychology* vol. 8, no. 1: 7–24.

Andrews-Hanna, Jessica R. 2012. The Brain's Default Network and its Adaptive Role in Internal Mentation. *Neuroscientist* vol. 18, no. 3: 251–270.

---

18   The functional role of empathy is called by Smith "mindreading," and the function of putting the narrative information we possess "under a new description," allowing us to feel it "from the inside," is called "mindfeeling" (Smith 2017, 194).

Andrews-Hanna**,** Jessica R., Yoni K. Ashar, Sona Dimidjian, and Tor D. Wager. 2017. Empathic Care and Distress: Predictive Brain Markers and Dissociable Brain Systems. *Neuron* vol. 94, no. 6: 1263–1273.

Wager, Tor D., Jian Kang, Timothy D. Johnson, Thomas E. Nichols, Ajay B. Satpute, and Lisa Feldman Barrett. 2015. A Bayesian Model of Category-Specific Emotional Brain Responses. *PLOS Computational Biology* vol. 11, no. 4: 1–27.

Barrett, Lisa Feldman. 2017. *How Emotions Are Made. The Secret Life of the Brain*. New York: Houghton Mifflin Harcourt.

Barsalou, Lawrence. 2016. Situated Conceptualization. Theory and Applications. In *Perceptual and Emotional Embodiment. Foundations of Embodied Cognition* vol. 1, eds. Yann Coello and Martin H. Fischer, 11–37. London, New York: Routledge.

Bartsch, Anne. 2010. Vivid Abstractions: On the Role of Emotion Metaphors in Film Viewers' Search for Deeper Insight and Meaning. *Midwest Studies in Philosophy* vol. 34, no. 1: 240–260.

Bishop, J. Sonia. 2007. Neurocognitive Mechanisms of Anxiety: an Integrative Account. *Trends in Cognitive Sciences* vol. 11, no. 7: 307–316.

Carroll**,** Noel and William P. Seeley. 2013. Cognitivism. Psychology and Neuroscience: Movies as Attentional Engines. In *Psychocinematics: Exploring Cognition at the Movies,* ed. Arthur P. Shimamura, 53–75. London and New York: Routledge.

Chattah, Juan. 2015. Film Music as Embodiment. In *Embodied Cognition and Cinema*, eds. Coegnarts, Maarten and Peter Kravanja, 81–112. Leuven: Leuven University Press.

Coulson, Seana and Todd Oakley. 2000. Blending Basics. *Cognitive Linguistics* vol. 11, no. 3–4: 175–196.

Fahlenbrach, Kathrin. 2008. Emotions in Sound: Audiovisual Metaphors in the Sound Design of Narrative Films. *Projections* vol. 2, no. 2: 85–103.

Fauconnier, Gilles. 1994. *Mental Spaces.* New York: Cambridge University Press.

Fauconnier, Gilles. 1997. *Mappings in Thought and Language.* Cambridge: Cambridge University Press.

Fauconnier, Gilles and Mark Turner. 1998. Conceptual Integration Networks. *Cognitive Science* vol. 22, no. 2: 133–187.

Fauconnier, Gilles and Mark Turner. 2002. *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities.* New York: Basic Books.

Friston, Karl. 2005. A Theory of Cortical Responses. *Philosophical Transactions of the Royal Society B* vol. 360, no. 1456: 815–836.

Friston, Karl. 2010. The Free-energy Principle: a Unified Brain Theory? *Nature Reviews Neuroscience* vol. 11: 127–138.

Friston, Karl. 2017. Consciousness is Not a Thing but a Process of Inference. *AEON.   https://aeon.co/essays/consciousness-is-not-a-thing-but-a-process-of-inference.* Last accessed 15. 02. 2020.

Guo, Christine, Vinh Nguyen, Matthew Hyett, Gordon Parker, and Michael J. Breakspear. 2015. Out-of-sync: Disrupted Neural Activity in Emotional Circuitry during Film Viewing in Melancholic Depression. *Scientific Reports* no. 5: 1–12.

Hyett, Matthew, Gordon Parker, Christine Guo, Andrew Zalesky, Vinh Nguyen, Tamara Yuen, and Michael J. Breakspear. 2015. Scene Unseen: Disrupted Neuronal Adaptation in Melancholia During Emotional Film Viewing. *Neuroimage: Clinical* no. 9: 660–667.

Johnson, Mark. 1987. *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason.* Chicago: University of Chicago Press.

Johnson, Mark. 2015. Foreword. In *Embodied Cognition and Cinema*, eds. Maarten Coegnarts and Peter Kravanja, 9–13. Leuven: Leuven University Press.

Kirchhoff, Michael and Julian Kiverstein. 2019. *Extended Consciousness and Predictive Processing: A Third-Wave View*. New York: Routledge.

Kövecses, Zoltán. 2000. *Metaphor and Emotion: Language, Culture and Body in Human Feeling.* Cambridge: Cambridge University Press.

Lakoff, George. 2014. Mapping the Brain's Metaphor Circuitry: Metaphorical Thought in Everyday Reason. *Frontiers in Human Neuroscience* no. 8: 1–14.

Langacker, Ronald 1987. *Foundations of Cognitive Grammar. Vol. 1. Theoretical Prerequisites*. Stanford: Stanford University Press.

Langacker, Ronald. 2008. *Cognitive Grammar. A Basic Introduction*. Oxford and New York: Oxford University Press.

Langacker, Ronald. 2009. *Investigations in Cognitive Grammar*. Berlin and New York: Mouton de Gruyter.

Maturana, Humberto and Francisco Varela. 1992 [1987]. *The Tree of Knowledge: The Biological Roots of Human Understanding.* Boston: Shambhala.

Odin, Roger. 2000. *De la fiction*. Bruxelles: De Boeck Universite.

Pecher, Diane, Inge Boot, and Saskia van Dantzig. 2011. Abstract Concepts: Sensory-Motor Grounding, Metaphors, and Beyond. In *Psychology of Learning and Motivation. Advances in Research and Theory,* ed. Brian Ross, 217–248. Burlington: Academic Press.

Pezzulo, Giovanni. 2014. Why Do You Fear the Bogeyman? An Embodied Predictive Coding Model of Perceptual Inference. *Cognitive, Affective, & Behavioral Neuroscience* vol. 14, no. 3: 902–911.

Raz, Gal and Talma Hendler. 2014. Forking Cinematic Paths to the Self. Neurocinematically Informed Model of Empathy in Motion Pictures. *Projections* vol. 8, no. 2: 89–114.

Russell, James A. 2003. Core Affect and the Psychological Construction of Emotion. *Psychological Review* vol. 110, no. 1: 145–172.

Sapolsky, Robert. 2017. *Behave: the Biology of Humans at Our Best and Worst.* New York: Penguin Press.

Seth, Anil, Keisuke Suzuki, and Hugo D. Critchley. 2012. An Interoceptive Predictive Coding Model of Conscious Presence. *Frontiers in Psychology* no. 2: 1–16.

Sheets-Johnstone, Maxine. 1999. *The Primacy of Movement.* Amsterdam: John Benjamins.

Smith, Greg M. 2003. *Film Structure and the Emotion System.* Cambridge: Cambridge University Press.

Smith Murray. 2017. *Film, Art, and the Third Culture. A Naturalized Aesthetics of Film.* Oxford: Oxford University Press.

Spivey, Michael. 2007. *The Continuity of Mind.* New York: Oxford University Press.

Sweetser, Eve. 2000. Blended Spaces and Performativity. *Cognitive Linguistics* vol. 11, no. 3–4: 305–333.

Thompson, Evan. 2007. *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*, Cambridge, Massachusetts, London: The Belknap Press of Harvard University Press.

Tuggy, David. 2007. Schematicity. In *The Oxford Handbook of Cognitive Linguistics*, eds. Dirk Geeraerts and Hubert Cuyckens, 82–116. Oxford: Oxford University Press.

Varela, Francisco, Thompson, Evan, and Eleanor Rosch. 1991. *The Embodied Mind: Cognitive Science and Human Experience.* Cambridge: MIT Press.

Vigliocco Gabriella et al. 2013. The Neural Representation of Abstract Words: The Role of Emotion. *Cerebral Cortex* vol. 24, no. 7: 1767–1777.

Wojciehowski, Hannah Chapelle, and Vittorio Gallese. 2011. How Stories Make Us Feel: Toward an Embodied Narratology. *California Italian Studies Journal* vol. 2, no. 1: 1–35.

List of Figures

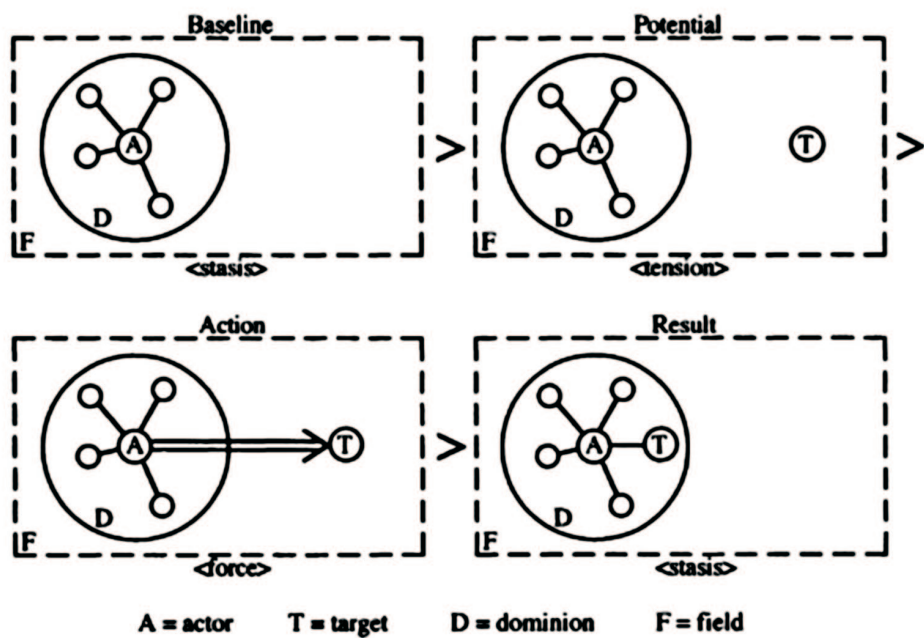**Figure 1**. The four stages of the control cycle cognitive model.

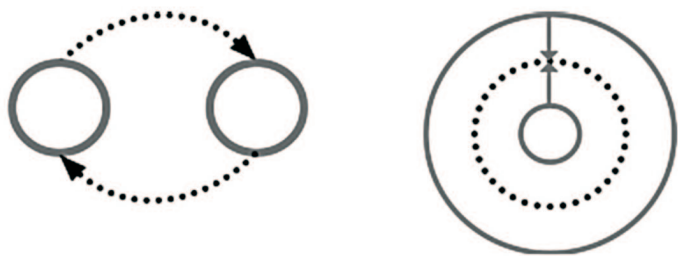**Figure 2**. Patterns of circular coupling.

Figure 3. Expression of emotion in a two-shot construction. (Illustration from Barrett et al. 2011, 287.)
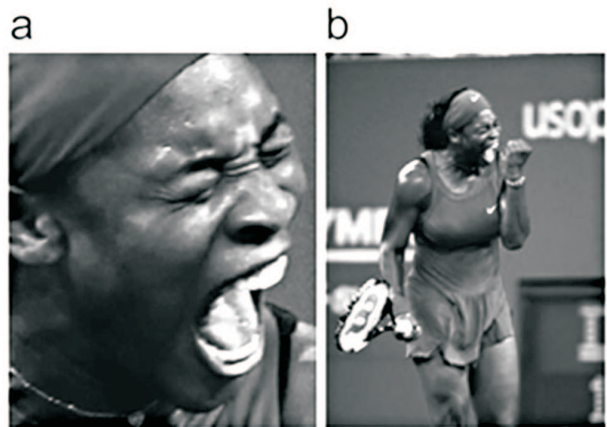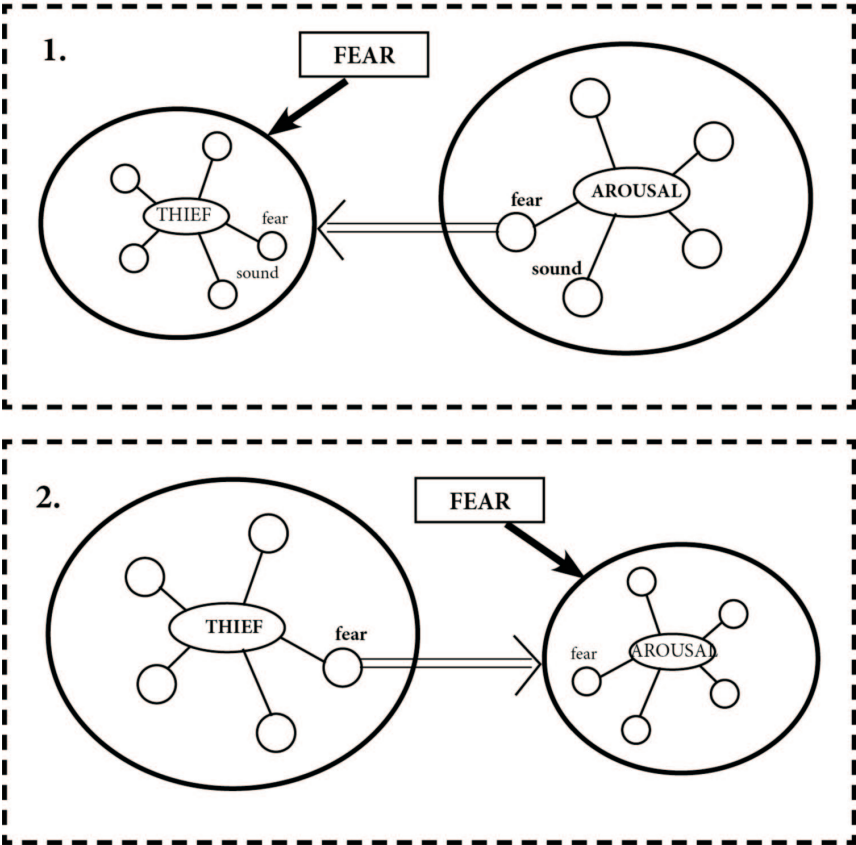

Figure 4.

Figure 5. *The Shining* (Stanley Kubrick, 1980).

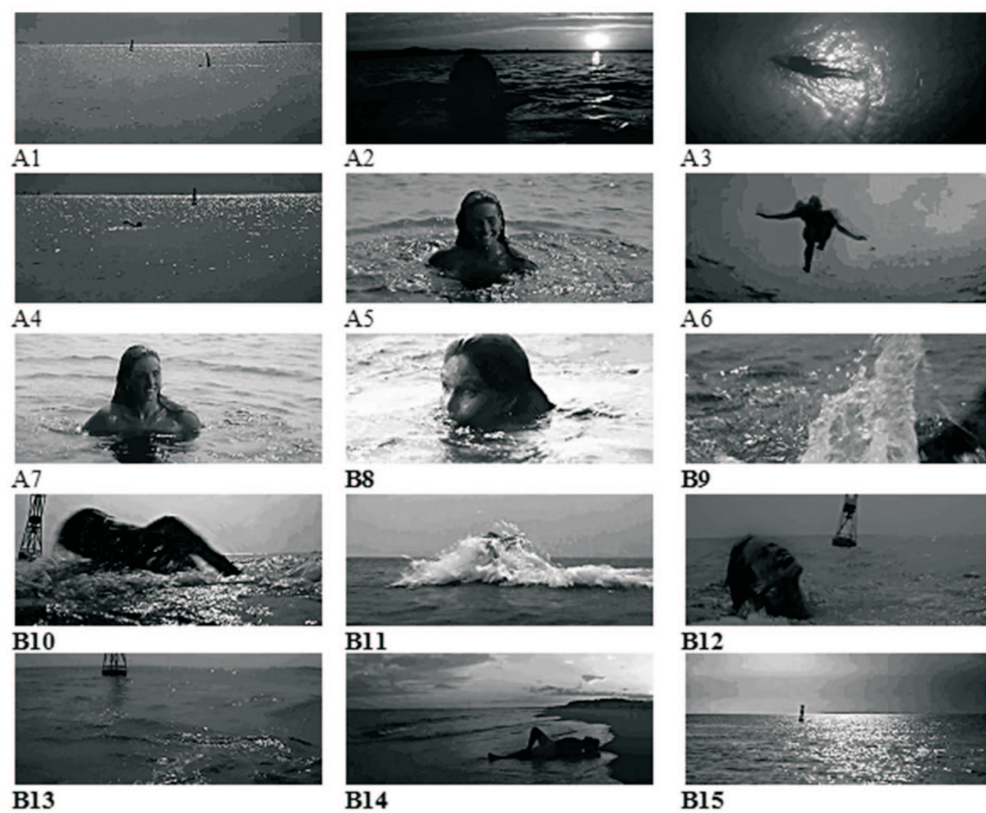**Figure 6**. *Jaws* (Steven Spielberg, 1975).

Figure 7. Summary of the perspectives adopted and the emotions unfolding in the episode condensed in shot A5 above.