7 What Does It Mean for a Computer to "Have" Emotions?

Rosalind W. Picard

There is a lot of talk about giving machines emotions, some of it fluff. Recently at a large technical meeting, a researcher stood up and talked of how a Barney stuffed animal (the purple dinosaur for kids) "has emotions." He did not define what he meant by this, but after repeating it several times, it became apparent that children attributed emotions to Barney, and that Barney had deliberately expressive behaviors that would encourage the kids to think Barney had emotions. But kids have attributed emotions to dolls and stuffed animals for as long as we know; and most of my technical colleagues would agree that such toys have never had and still do not have emotions. What is different now that prompts a researcher to make such a claim? Is the computational plush an example of a computer that really does have emotions?

If not Barney, then what would be an example of a computational system that has emotions? I am not a philosopher, and this paper will not be a discussion of the meaning of this question in any philosophical sense. However, as an engineer I am interested in what capabilities I would require a machine to have before I would say that it "has emotions," if that is even possible.

Theorists still grapple with the problem of defining emotion, after many decades of discussion, and no clean definition looks likely to emerge. Even without a precise definition, one can still begin to say concrete things about certain components of emotion, at least based on what is known about human and animal emotions. Of course, much is still unknown about human emotions, so we are nowhere near being able to model them, much less duplicate all their functions in machines. Also, all scientific findings are subject to revision—history has certainly taught us humility, that what scientists believed to be true at one point has often been changed at a later date.

I wish to begin by mentioning four motivations for giving machines certain emotional abilities (and there are more). One goal is to build robots and synthetic characters that can emulate living humans and animals—for example, to build a humanoid robot. A

second goal is to make machines that are intelligent, even though it is also impossible to find a widely accepted definition of machine intelligence. A third goal is to try to understand human emotions by modeling them. Although I find these three goals intriguing, my main focus is on a fourth: making machines less frustrating to interact with. Toward this goal, my research assistants and I have begun to develop computers that can identify and recognize situations that frustrate the user, perceiving not only the user's behavior and expressions, but also what the system was doing at the time. Such signs of frustration can then be associated with potential causes for which the machine might be responsible or able to help, and the machine can then try to learn how to adjust its behavior to help reduce frustration. It may be as simple as the computer noticing that lots of fancy "smart" features are irritating to the user, and offering the user a way to remove all of them. Or, it may be that the computer's sensitive acknowledgment of and adaptation to user frustration simply leads to more productive and pleasing interactions. One of the key ideas is that the system could associate expressions of users, such as pleasure and displeasure, with its own behavior, as a kind of reward and punishment. In this age of adaptive, learning computer systems, such feedback happens to be easy and natural for users to provide.

Discussion

Picard: One of the things that is controversial with respect to agents is if they should show empathy to people. This is sort of strange, a computer saying "That feels pretty bad, and I am sorry to hear that you had such a bad experience," when a computer has no feelings. You would think this would just upset people. In fact, Reeves and Nass[1] found the same surprises in their studies out in Stanford. They tested their system with Stanford students who know that the machine does not have emotions.

Stern: Do you think people could possibly be thinking, well, the person who wrote this program has empathy?

Picard: This is one of the key factors that Reeves and Nass talk about and that we all address: Do they attribute any expression of feelings to the designer of the software? And if so, then the computer should not be saying "I," "my," "me," or whatever, it should be saying "the makers of the software." And they ran experiments investigating that, too. They have concluded that it's the machine. Even though people know better, they act as if it's the machine and not the maker of the machine.

Elliott: One of our experiences is similar, and I would say that to the extent that there is complexity in the understanding, the feeling of sincerity goes up. If you say "Here is everything I know, it's not much, but I do know this," people seem to accept that, and to the extent that there is more complexity in there and it feels like if there is more understanding, they accept it even more. It does not matter if you say it's not real; it's a bit like flattery not being real, and still ...

Picard: That's interesting. Yes.

Elliott: Flattery wears off, when you get it two or three times in a row, it's like: "Well, I heard this before." But if the complexity is there, it doesn't seem to run off in the same way: "I know it's not real, but you seem to understand quite a few things about how I feel, and that satisfies me." If it's just "I am sorry, but I don't know why" instead of "I am sorry because I believe that you really wanted this thing, and you did not get it, and you are embarrassed that you did not get it."

Picard: This reminds me of the strategy I use with my two-year-old: if he tries to do something, and I don't like it and say "no," then he says "why?" If I give him a short explanation, he asks "why" again. If I give another short explanation, he says "why" again. If I give him a really long complex explanation, he gets bored and forgets about it. He is training me in a sense.

Bellman: I guess I would want to see more experimentation about the implicit people behind the artifacts, because I think we have some information from our experiences that say that people, even though they suspend disbelief, are actually very aware of the authorship by other human beings. And in fact, in our virtual world studies, we often get users who come up to the author and say: "I really enjoyed your robot. He is so great, he is so lovable, you know!"

Elliott: I think that is reflection after the fact, though.

Ball: But in these experiments, people know. There is no question about misunderstanding this computer. It's just that they are still affected.

Elliott: They are inherently engaged, and this satisfies this feeling of empathy.

Sloman: We are biologically programmed to respond to this kind of behavior. If this behavior comes from a computer, we will still respond.

Ortony: But we may at the same time also praise the author as we praise the parent of a child. We have enjoyed an interaction, we don't attribute the behavior of the child to the interaction alone, but we see, as Clark said, a sign of the parent in the child. So, we say: "I really liked your kid."

Picard: But people do say: "Can't you control your child?"

Ortony: Well, that's the other side.

Picard: Well, we are not going to be able to control, so to speak, these agents at some point. I think this is a responsibility decision to make as designers, while we are in control.

My first goal thus involves sensing and recognizing patterns of emotional information—dynamic expressive spatiotemporal forms that influence the face, voice, posture, and ways the person moves—as well as sensing and reasoning about other situational variables, such as if the person retyped the same word many times and is now using negative language. All of this is what I refer to in shorthand as "recognizing emotion," although I should be clear that it means the first sentence of this paragraph, and not that a computer can know your innermost emotions, which involve thoughts and feelings that no person besides you can sense. But once a computer has recognized emotion, what should it do? Here lies my second main goal: giving the computer the ability to adapt to the emotional feedback in a way that does not further frustrate the user. Although "having emotion" may help with the first goal, I can imagine how to achieve the first goal without this ability. However, the second goal involves intricacies in regulating and managing ongoing perceptual information, attention, decision making, and learning. All of these functions in humans apparently involve emotion. This does not mean that we could not possibly implement them in machines without emotion. At the same time, it appears to be the case that all living intelligent systems have emotion in some form, and that humans have the most sophisticated emotion systems of all, as evinced not just by a greater development of limbic and cortical structures, but also by greater

facial musculature, a hairless face, and the use of artistic expression, including music, for expressing emotions beyond verbal articulation.

Part of me would love to give a computer the ability to recognize and deal with frustration as well as a person can, without giving it emotions. I have no longing to make a computer into a companion; I am quite content with it as a tool. However, it has become a very complex adaptive tool that frustrates so many people that I think it's time to look at how it can do a better job of adapting to people. I think emotion will play a key role in this. Let's look more closely at four components of emotion that people have, and how these might or might not become a part of a machine.

7.1 Components of Emotion

I find it useful to identify at least four components when talking about emotions in the context of what one might want to try to implement in machines (figure 7.1). Some of these components already exist in some computational systems. The components are (1) emotional appearance, (2) multiple levels of emotion generation, (3) emotional experience, and (4) (a large category of) mind-body interactions. These four components are not intended to be self-evident from their short names, nor are they intended to be mutually exclusive or collectively exhaustive. Let me say what I mean by each, and why all four are important to consider.

A computer that "has emotions," in the sense that a person does, will be capable of:

1. **Emotional appearance**
2. **Multilevel emotion generation**
3. **Emotional experience**
4. **Mind-body interactions**

Figure 7.1

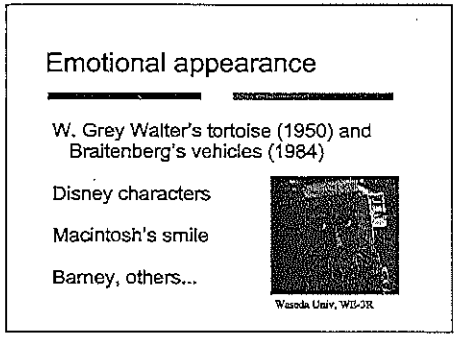

Figure 7.2

Emotional Appearance

Barney the stuffed animal sometimes *sounds as if* he is happy. Like a 3-D animated cartoon, he has expressions and behaviors that were designed to communicate certain emotions. "Emotional appearance" includes behavior or expressions that *give the appearance that* the system has emotions (figure 7.2).

This component is the weakest of the four, in the sense that it is the easiest of the four to produce, at least at a superficial level. However, I include it because this quality is all that an outside observer (nondesigner of the system, who cannot access or decipher its inward functions) has at his or her disposal in order to judge the emotional nature of the system. By and large, it is what the crew in the film *2001: A Space Odyssey* did not perceive about the computer HAL until the end of the film, otherwise they might have obtained earlier clues about HAL's increasingly harmful emotional state, which at the end of the film is illuminated when HAL finally says, "I'm afraid, Dave, I'm afraid." This component is also the most commonly implemented in machines today—primarily in agents and robots that display emotional behaviors in order to "look natural" or to "look believable." (Note: the following discussion occurred when the original slide had this component labeled as "emotional behavior.")

Discussion

Ortony: I think emotional behavior is not really interesting. Acting is emotional behavior—it's all imitation and mimicry. The Mac's smile is not emotional behavior, unless it is actually initiated by and related to an emotion.

Picard: Actually, I could map these behaviors to your slide with the emotional response tendencies (figure 7.2).

Ortony: Well, no, because they are actually in response to an emotion, while the Mac's smile isn't in response to anything. It's just a curvy line.

Picard: No, no! It holds a response to an internal state that reads as "satisfactory, good . . . ," which I would not call an emotion, but some people would say it gives rise to this feeling that universally is recognized as "all is going well."

Ortony: If you say "emotional behavior" related to "multilevel emotion generation" [in figure 7.1], then I am perfectly happy. What I am not happy with is mimicry, acting, and all those other things that are "as if" and intended to pretend. My point is actually a causal connection between 2 and 1. These things are not independent—that's all I am saying.

Picard: I have never said that these things are independent. In fact, I talked about the explicit interconnections between these!

Ball: The behavior is irrelevant. I don't think it's necessary to show behaviors.

Picard: I am not saying that if you don't see emotional behavior, there is no emotion.

Ortony: I understand that. It's just the examples you give of emotional behavior that aren't actually caused by 2, 3, and therefore they are not examples of emotional behavior, they are "as if" behaviors! That's all.

Picard: And I am just talking about what they are received as. You know, to an observer it may not make a difference.

Ortony: We are talking about what they are capable of, not how they are interpreted.

Picard: I see your distinction. I think we could work it out as the semantics of what we are talking about.

Ortony: No, no. We have to work it out with a causal model as opposed to a set of causally unrelated independent things.

Picard: Ok. What I would say is that a person who has emotions is capable of emotional behavior.

Ortony: Yes, absolutely right.

Because emotional appearance results largely from emotional behavior, and because I include the making of facial, vocal, and other expressions as kinds of behavior, I have previously referred to this component as *emotional behavior.* I am here changing my two-word description because a couple colleagues at the Vienna workshop argued that it was confusing; however, I am not changing what it refers to, which remains the *emotional appearance of the system's behavior.*

Examples of systems with behaviors that appear to be emotional include the tortoises of W. Gray Walter (1950) and Braitenberg's Vehicles (Braitenberg 1984). When one of Braitenberg's little vehicles approached a light or backed rapidly away from it, observers described the behavior as "liking lights" or as "acting afraid of lights," both of which involve emotional attribution, despite the fact that the vehicles had no deliberately designed internal mechanisms of emotion. Today there are a number of efforts to give computers facial expressions; the Macintosh has been displaying a smile at people for years, and there is a growing tendency to build animated agents and other synthetic characters and avatars that would have emotional expressions. These expressive behaviors may result in people saying the system is "happy" or otherwise, because it appears that way.

I think all of us would agree that the examples just given do not have internal feelings, and their behavior is not generated by emotions in the same sense that human or animal behavior is. However, the boundary is quickly blurred: Contrast a machine like the Apple Macintosh, which shows a smile because it is hardwired to do that in a particular machine state, and a new "emotional robot," which shows a smile (Johnstone 1999) because it has appraised its present state as good and its present situation as one where smiling can communicate something useful. The Mac's expression signals that the boot-up has succeeded and the machine is in a satisfactory state for the user to proceed. However, most of us would *not* say that the Mac is happy. More might say that the robot is happy, in a rudimentary kind of way. But, if the robot's happy facial expression were driven by a simple internal state labeled "satisfaction," then it would really be no different than the Mac's display of a smile. As the generation mechanisms become more complex and adapted for many such states and expressions, then the argument that the expression or behavior really arose from an emotion becomes more compelling. The more complex the system, and the higher the user's expectations, the harder it also becomes for the system's designer to craft the appearance of natural, believable emotions. Nonetheless, we should not let mere complexity fool us into thinking emotions are there.

If a system really has emotions, then we expect to see those emotions influence and give rise to behavior on many levels. There are the obvious expressions and other observable emotional behaviors, like saying "Humph," and turning abruptly away from the speaker; however, emotions also modulate nonemotional behaviors: The way you pick up a pen (a neutral behavior) is different when you are seething with anger versus when you are bubbling with delight. True emotions influence a number of internal functions, which are generally not apparent to anyone but the designer of the system (and in part to the system, to the extent that it is given a kind of "conscious awareness" of such). Some of emotion's most important functions are those that are unseen, or at least very hard to see. The body-mind mechanisms for signaling and linking the many seen and unseen functions are primarily captured by the fourth component, which I'll describe shortly.

Multiple Levels of Emotion Generation

Animals and people have fast subconscious brain mechanisms that perform high-priority survival-related functions, such as the response of fear in the face of danger or threat. LeDoux (1996) has described the subcortical pathway of fear's "quick and dirty" mechanism, which precedes cortical involvement. This level of preconscious, largely innate, but not highly accurate emotion generation appears to be critical for survival in living systems. One can imagine giving robots and machines sensors that operate at a similar level—in a relatively hardwired way, detecting when the system's critical parameters are in a danger zone, and triggering rapid protective responses, which can shortly thereafter be modified by slower, more accurate mechanisms.

The level of emotions just described stands in contrast with slightly slower (although still very fast) emotion generation that tends to involve higher cortical functions and may or may not involve conscious appraisals (figure 7.3). If you jump out of the way of a snake, and suddenly realize it was only a stick, then that was probably an instance of the fast subconscious fear-generation mechanism. In contrast, if you hear that a convicted killer has

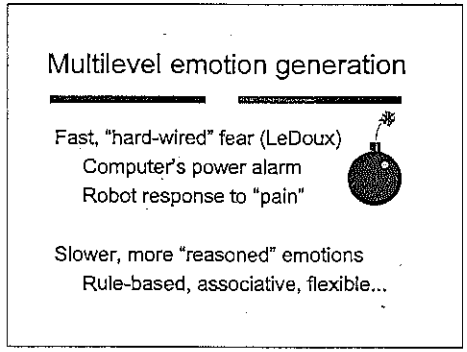

Figure 7.3

escaped a nearby prison, and consequently decide that you don't want to leave the house, then it is likely that your thoughts generated a form of a learned fear response, which subsequently influenced your decision. You may have never seen a convicted killer, but you cognitively know that such a person could be dangerous, and you associate with it a response that you learned from a similar but real experience. This learned fear response engages some of the same parts of the brain as the lower-level quick version of fear, but it additionally involves reasoning and cortical appraisal of an emotional situation.

Some of the most common methods of "implementing emotions" in computers involve constructing rules for appraising a situation, which then give rise to an emotion appropriate to that situation. An example is the OCC model (Ortony, Clore, and Collins 1988), which was not designed to synthesize emotions, but rather to reason about them, but works in part for either. Consider the generation of joy, which involves deciding that if an event happens, and that event is desirable, then it may result in joy for oneself or in happiness for another. A machine can use this rule-based reasoning either to try to infer another's emotion, or to synthesize an internal emotional state label for itself. All of this can happen in the machine in a cold and logical way, without anything that an outsider might observe as emotion. It can happen without any so-called conscious awareness or "feeling" of what the machine is doing. This kind of "emotion generation" does not need to give

rise to component one—emotional appearance—or to the other two components listed below, but it could potentially give rise to all of them. In a healthy human, such emotional appraisals are also influenced by one's feelings, via many levels of mechanisms.

People appear to be able to reason in a cold way about emotions, with minimal if any engaging of observable bodily responses. However, more often there seem to be bodily changes and *feelings* associated with having an emotion, especially if the emotion is intense. An exception arises in certain neurologically impaired patients (e.g., see accounts in Damasio 1994) who show minimal signs of such somatic concomitants of emotion. If you show these patients grotesque blood-and-guts mutilation scenes, which cause most people to have high skin conductivity levels and to have a feeling of horror and revulsion, these patients will report in a cool cognitive way that the scenes are horrible and revolting, but they will not have any such feelings, nor will they have any measurable skin conductivity change. Their emotional detachment is remarkable, and might seem a feature, if it were not for the serious problems that such lack of emotionality actually seems to be a part of in day-to-day rational functioning, rendering these otherwise intelligent people severely handicapped. What these patients have is similar to what machines that coldly appraise emotions can have—a level of emotion generation that involves appraisal, without any obvious level of bodily or somatic involvement.

It is not clear to what extent normal people can have emotions without having any associated bodily changes other than those of unfelt thought patterns in the brain; consequently, the levels of emotion generation described here may not typically exist in normal people without being accompanied by some of the mind-body linkages in the fourth component, described below. Nonetheless, multilevel generation of emotion is an important component because of its descriptive power for what is believed to happen in human emotion generation, and because some of these levels have already been implemented to a certain degree in machines. It is also relevant for certain neurologically atypical people, such as high-functioning autistics, who describe their ability to understand emotions as "like a computer—having to reason about what an emotion is" versus understanding it intuitively.

The two levels just described—quick and dirty subconsciously generated emotions and slightly slower, more reason-generated

emotions, are not the only possibilities. Nor does my choice of these two examples impose a belief that "reasoning" has to be conscious. My point is instead that here are examples of emotions occurring via different levels of mechanisms. I expect that neuroscientists will find unique patterns of activation (and deactivation) across cortical and subcortical regions for each kind of emotion—joy, fear, frustration, anger, and so forth, and possible unique patterns for significant variations in levels of these. I would also expect we would build multiple levels of activation of emotion-generation mechanisms in machines, varying in resources used and varying in timing and in influence, in accord with the specific roles of each emotion. Some would be quick and perhaps less accurate, while some would be more carefully deliberated. Some would be at a level that could be consciously attended, or at least attended by some "higher" mechanisms, while some would occur without any such monitoring or awareness. Some of the mechanisms would be easy to modify over time, while others would be fairly hardwired. Some of the emotion-generation mechanisms might be rule based, and easy to reason about—at least after the fact if not during—while others would be triggered by patterns of similarity that might not be easily explained. And many or even all of these mechanisms might be active at different levels contributing to background or mixed emotions, not just to a small set of discrete emotions. In summary, machines will have different combinations of mechanisms activating different emotions, a veritable orchestra for emotion generation.

Emotional Experience

We humans have the ability to perceive our personal emotional state and to experience a range of feelings, although many times we are not aware of or do not have the language to describe what we are feeling (figure 7.4). Our feelings involve sensing of physiological and biochemical changes particular to our human bodies (I include the brain and biochemical changes within it as part of the body). Even as machines acquire abilities to sense what their "bodies" are doing, the sensations remain different than those of human bodies, because the bodies are substantially different. In this *sense* machine feelings cannot duplicate human feelings. Nonetheless, machines need to be able to sense and monitor more

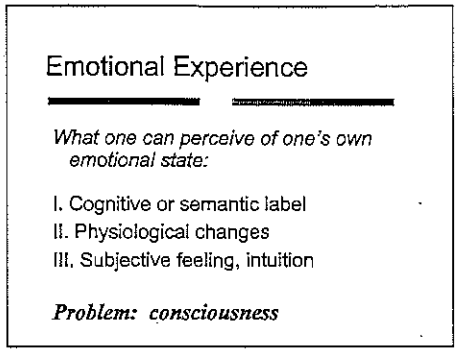

Figure 7.4

of what is going on within and around their systems if they are to do a better job of regulating and adapting their own behavior. They will likely need mechanisms that perform the functions performed by what we call consciousness, if only to better evaluate what they are doing and learn from it.

A great distinction exists between our experience and what machines might have. The *quality* of conscious awareness of our feelings and intuition currently defies mechanistic description, much less implementation in machines. Several of my colleagues think that it is just a matter of time and computational power before machines will "evolve" consciousness, and one of them tells me he's figured out how to implement consciousness, but I see no scientific nuggets that support such belief. But I also have no proof that it cannot be done. It won't be long before we can implement numerous *functions* of consciousness, such as awareness and monitoring of events in machines, but these functions should not be confused with the *experience of self* that we humans have. I do not yet see how we could computationally build even an approximation to the quality of emotional experience or experience of self that we have. Thus I remain a skeptic on whether machines will ever attain consciousness in the same way we humans think of that concept. Consciousness, and life, for that matter, involves qualities that I do not yet see humans as capable of creating, outside of procreation. Perhaps someday we will have such creative abilities; nonetheless, I do not see them arising as a natural progression of

past and present computational designs, not even with the advent of quantum computing.

Discussion

Riecken: I don't know what consciousness is.

Picard: It's more loaded than awareness. I prefer that we say (referring to figure 7.4): what we perceive of our own emotion, what something in us can perceive or become aware of. Because my list applied both to people and computers, I didn't want to put the word *self* in. I always said "one's," but, you know, that's not quite the same as "self." I think *self* is a more loaded word than just saying what this entity perceives of what's going on within this entity.

Sloman: A good operating system has a certain amount of self-awareness.

Ortony: It's a little tough for a machine to be aware of its physiological changes if it does not have a physiology.

Picard: A computer can sense physically. We recently hardwired the back of our monitor to sense surges in voltage, so it could sense the precise instant that it was displaying the image to subjects in one of our studies. The operating system did not give us the hooks to sense that. I think we need to build software *and* hardware that has better self-awareness.

Rolls: Well, can I say that I am really worried about saying that any machine has self-awareness in this sense?

Picard: Yes. It sounds just as dangerous as saying it has emotions.

Rolls: Whenever we use the word "awareness," it implies to me qualia of phenomenology. If you would replace that by "self-monitoring," we would not get into a problem.

If we can understand something, we can model it and build a computational model of it. Modeling is a form of imitation, not duplication. Thus I use the term "imitate" instead of "duplicate" with respect to implementing this component in machines. In fact, we should probably be more careful about using the phrase "imitating some of the known mechanisms of human emotion in machines" to describe much of the current research concerned with "giving machines emotion." For brevity and readability, the latter phrase is what I will continue to use, with hope that with this

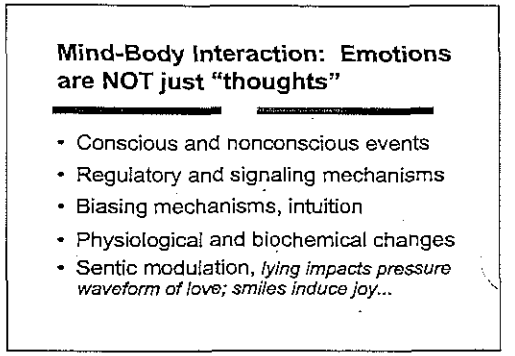

Figure 7.5

paper, we will begin to find some common understanding for what this shorter expression represents.

Mind-Body Interactions

The fourth component is a broad category including many signaling and regulatory mechanisms that emotion seems to provide in linking cognitive and other bodily activities (figure 7.5). Here we find that emotions often involve changes in bodily systems outside the brain, as well as inside the brain. There is evidence, for example, that emotions inhibit and activate different regions of the brain, facilitating some kinds of cognitive activity while inhibiting others. Researchers have shown numerous effects of emotion and mood biases on creative problem solving, perception, memory retrieval, learning, judgment, and more. (See Picard 1997a for a description of several such findings.) Not only do human emotions influence brain information processing, but they also influence the information processing that goes on in the gastrointestinal and immune systems (see Gershon 1998 for a description of information processing in the gut).

Emotions modulate our muscular activity, shaping the space-time trajectories of even very simple movements, such as the way we press on a surface when angry versus when joyful. I call the way in which emotions influence bodily activity *sentic modulation*, after Manfred Clynes's (1977) work in sentics, where he first attempted to quantify and measure a spatiotemporal form of

emotion. Clynes found that even simple finger pressure, applied to a nondescript firm surface, took on a characteristic pattern when people tried to express different emotions. Moreover, some of the emotions had implications for cognitive states such as lying or telling the truth. Subjects were asked to physically express either anger or love while lying or while telling the truth, and their physical expressions (finger pressure patterns, measured along two dimensions) were recorded and measured. When subjects were asked to express anger, the expressions were not significantly different during lying than when telling the truth. However, when subjects were asked to express love, the expressions differed significantly when lying versus when telling the truth. In other words, their bodily emotional expression was selectively interfered with by the cognitive state of lying, given that it was not obviously interfered with in any other way.

I expect that this particular love-lying interaction is one of many that remain to be characterized. The interaction between emotions and other physical and cognitive states is rich, and much work remains to be done to refine our understanding of which states inhibit and activate each other. As each interaction is functionally characterized in humans, so too might it be implemented in machines. Ultimately, if a machine is to duplicate human emotions, the level of duplication must include these many signaling, regulatory components of emotion, which weave interactive links among physical and mental states.

Consider building synthetic pain-sensing and signaling mechanisms. Some machines will probably need an ability outside of their modifiable control to essentially *feel bad* at certain times—to sense a kind of highest-priority unpleasant attention-refocusing signal in a situation of dire self-preservation (where the word *self* is not intended to personify, but only to refer back to the machine). This "feeling," however it is constructed, would be of the same incessantly nagging, attention-provoking nature that pain provides in humans. When people lose their sense of pain, they allow severe damage to their body, often as the accumulation of subtle small damages that go unnoticed. Brand and Yancey (1997) describe attempts to build automatic pain systems for people—in one case, a system that senses potentially damaging pressure patterns over time. The artificial pain sensors relay signs of pain to the patient via other negative attention-getting signals, such as an obnoxious sound in their ear. One of the ideas behind the artificial system is

to provide the advantages of pain—calling attention to danger—without the disadvantages—the bad feelings. The inputs approximate those of real pain inputs, and the outputs are symbolically the same: irritating and attention getting. Ironically, what people who use the artificial system do is either turn these annoying warnings off or ignore them, rationalizing that it can't be as bad as it sounds. Eventually the pain-impaired person gets seriously injured, although he or she doesn't really mind because it does not hurt. In short, the artificial pain system doesn't work; somehow it has to be "real" enough that you can't override or ignore it for long. Otherwise, injury accumulates, and the long-term prognosis is bad.

Whatever version of "pain" we give machines, if its goal is system preservation, then it must be such that it is equivalent to not being able to be turned off, except under greater goals, or except by the machine's designer. This is not simply to say that pain avoidance should always have the highest priority. Self-preservation goals may at some point be judged as less important than another goal, as suggested by Asimov's three laws of robotics, where human life is placed above robot "life," although this presumes that such assessments could be accurately made by the robot. Humans sometimes endure tremendous pain and loss of life for a greater goal. Similar trade-offs in behavior are likely to be desirable in certain kinds of machines.

Before concluding this section, let me restate that it is important to keep in mind that all computers will not need all components of all emotions. Just like simple animal forms do not need more than a few primary emotion mechanisms, not all computers will need all emotional abilities, and some will not need any emotional abilities. Humans are not distinguished from animals just by a higher ability to reason, but also by greater affective abilities. I expect more sophisticated computers to need correspondingly sophisticated emotion functions.

7.2 Discussion

What is my view on what it means for a computer to have emotion? Before closing this discussion, we should keep in mind that we are still learning the answer to this question for living systems, and the machine is not even alive. That said, I have tried to briefly describe four components of emotion that are present in people and to discuss how they might begin to be built into machines.

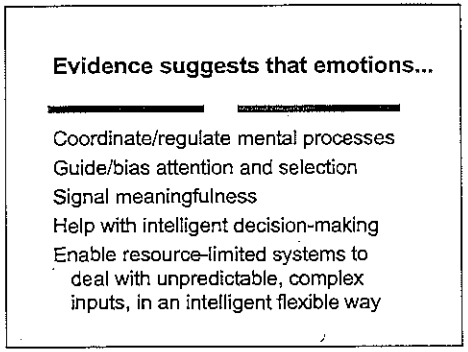

Figure 7.6

My claim, which opened this section, is that all four components of emotion occur in a healthy human. Each component in turn has many levels and nuances (figure 7.6). If we acknowledge, say, $N = 60$ such nuances, and implement them in a machine, then the machine can be said to have dozens of mechanisms of emotion that imitate, or possibly duplicate, those of the human emotional system. So, does the machine "have emotions" in the same sense that we do? There is a very basic problem with answering this: One could always argue that there are only N *known* human emotion mechanisms and more may become known; how many does the machine have to have before one will say that it has humanlike emotions? If we require all of them to be identified and implemented, then one can always argue that machines aren't there yet, because we can never be assured that we have understood and imitated everything there is to know. Consequently, one could never confidently say that machines have emotions in the sense that we do. The alternative is to agree on some value of N that suffices as a form of "critical mass." But that is also ultimately unsatisfactory. Furthermore, some machines will have or may benefit from aspects of emotion-like mechanisms that humans don't have. Animals doubtless already have different mechanisms of emotion than humans, and we are not troubled by the thought of someone saying that they have emotions.

Ultimately, we face the fact that a precise equality between human and machine emotion mechanisms cannot be assured because we simply do not have complete lists of what there is to compare, nor do we know how incomplete our lists are.

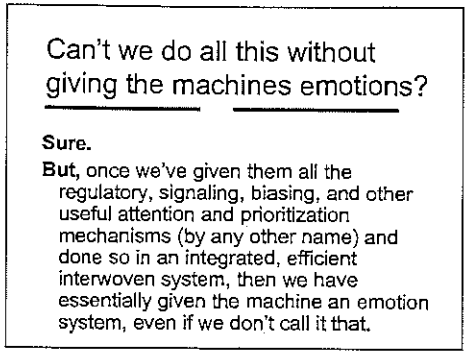

Figure 7.7

Machines are still not living organisms, despite the fact that we describe many living organisms as machines (figure 7.7). It has become the custom to associate machine behavior and human behavior without really thinking about the differences anymore. Despite the rhetoric, our man-made machines remain of a nature entirely different than living things. Does this mean they *cannot* have emotions? I think not, if we are clear that we are describing emotions as mechanisms with functional components like the four described here. Almost all of these have been implemented in machines at some level, and I can see a path toward implementing all of them. At the same time, it is prudent to acknowledge that one of the components, emotional experience, includes components of consciousness that have not yet been shown to be reducible to computational functions. Machines with all components but this one might be said to have emotion systems, but no real feelings.

As we make lists of functions and match them, let us not forget that the whole process of representing emotions as mechanisms and functions for implementation in machines is approximate. The process is inherently limited to that which we can observe, represent, and reproduce. It would be arrogant and presumptuous to not admit that our abilities in these areas are finite and small compared to all that is unknown, which may be infinite.

Remember that I began this presentation asking whether or not it was necessary to give machines emotions if all we are interested in is giving them the ability to recognize and respond appropriately to

a user's emotion. Suppose we just want the computer to see it has annoyed someone, and to change its behavior so as not to do that again; why bother giving it emotions? Well, we still may not have to bother, if we can give it all the functions that deal with complex unpredictable inputs in an intelligent and flexible way, carefully managing the limited resources, dynamically shifting them to what is most important, judging importance and salience, juggling priorities and attention, signaling the useful biases and action-readiness potentials that might lead to intelligent decisions and action, and so forth. Each of these functions, and others that might someday be added to this list, may possibly be implemented by means other than "emotions." However, it is the case that these functions, in humans, all seem to involve the emotional system. In particular, these functions involve the second, third, and fourth components of emotion that I've described, which sometimes give rise to the first component. We may find once we have implemented all of these useful functions, *and* integrated them in an efficient, flexible, and robust system, that we have essentially given the machine an emotion system, even if we don't call it that.

Machines already have some mechanisms that implement (in part) the functions implemented by the human emotional system. Computers are acquiring computational functions of emotion systems whether or not one uses the "e" word. But computers do not have humanlike emotions in any rich or experiential natural sense. They may sense and label certain physical events as categories of "sensations," but they do not experience feelings like we do. They may have signals that perform many and possibly all of the functions performed by our feelings, but this does not establish equivalence between their emotion systems and ours. Computers may have mechanisms that imitate some of ours, but this is only in part, especially because our bodies differ and because so little is known about human emotions.

It is science's methodology to try to reduce complex phenomena like emotions to a list of functional requirements, and it is the challenge of many in computer science to try to duplicate these in computers to different degrees, depending on the motivations of the research. But we must not be glib in presenting this challenge to the public, who thinks of emotion as the final frontier of what separates man from machine. When a scientist tells the public that a machine "has emotion" then the public concludes that not only

could Deep Blue beat a grand master, but also Deep Blue could *feel* the joy of victory. The public is expecting that science will catch up with science fiction, that we will build HAL and other machines that have true feelings, and that emotions will consequently be reduced to a program available as a plug-in or free download if you click on the right ad. I think we do a disservice when we talk in such a way to the public, and that it is our role to clarify what aspects of emotion we are really implementing.

Emotions are not the system that separates man and machine; the distinction probably lies with a less popular concept—the soul—an entity that currently remains ineffable, but is something more than a conscious living self. I don't have much to say about this, except that we should be clear to the public that giving a machine emotion does not imply giving it a soul. As I have described, the component of emotional experience is closely intertwined with a living self and I remain uncertain about the possibility of reducing this component to computational elements. However, I think that the other three components of emotion are largely capable of machine implementation.

If the day comes that scientists think that human emotion and its interactions with the mind and body are precisely described and understood, then it will not be many days afterward that such functions will be implemented in a machine, to the closest approximation as possible to the human system. In that time, most people would probably say that the machine has emotions, and few scientists will focus on what this means. Instead, we will focus on how machines, made in our image, can be guided toward good and away from evil, while remaining free to do what we designed them to do.

Discussion: Emotional Maturity

Elliott: Socially intelligent systems or emotionally intelligent systems require emotions. And that is what we want to build, or what we want to study.

Picard: I don't actually think that we can say that emotionally intelligent systems require emotions. That's the question. I am trying to build emotionally intelligent systems, and I will see how far I can go without giving them emotions. Which components of emotions? I guess I can invent a mechanism for each of them, and I

don't need emotions to do that. But by the time we have built these special mechanisms that perform the decision making this way, maybe we discover that it would be more efficient to go back and just build an emotional system, because that one emotional system could maybe do it all.

Ortony: I thought you were going to say something else. I thought you were going to say that emotion did somehow fall out as a by-product of having built an integrated system.

Picard: Actually, I could say that as well, too. By the time you have done the integrated system without ever invoking the word "emotion," it is an emotional system. I have always thought this was obvious. I never bothered to say it....

Ortony: It's not totally obvious, you know.

Picard: I used to design computer architectures for a living. You can play out all the goals of what you want to achieve on the table, and then you figure out how you combine it all in an efficient architecture. So, to me, this was just obvious. We have been focusing on systems that help people communicate emotions, that help them express emotion; or the machine might express it—and might try to recognize emotion. Is that going to make a really emotionally intelligent system, if we do that? I don't think so. These are exactly the capabilities autistics have. Autistics have emotions, they can express emotions, and they can sometimes pattern-recognize other people's emotions, and yet they are really difficult to interact with as human beings. We cannot consider them as emotionally intelligent.

Ortony: Let me offer a new word, which is *emotional maturity*. I think it is a much more felicitous term than *emotional intelligence*. I think *emotional maturity* is what it really is! I mean, it's a much more natural way to think about it. Emotions do in fact develop in humans. In the normal course of development, they mature, and people become emotionally sophisticated and capable of doing all these things that go under the rubric of "emotional intelligence."

Or maybe there is a better word yet.

Sloman: *Competent. Emotionally competent!*

Bellman: Emotional competency in infants at a certain stage for certain things.

Ortony: Well, the reason I said "mature" is that it implies age-appropriate competence.

Picard: I know *intelligence* is a loaded word for a lot of people. And so I just list the set of learnable skills, as opposed to implying some innate capabilities.

Trappl: Maturity implies a genetic aspect.

Picard: Yes, it's rather hard to make a sure divide. But there are a lot of arguments for the case that you can teach people how to improve this set of skills, so to some degree, they are learnable, although to some degree they are probably also genetic.

Ortony: You can teach people to improve their posture, but that does not mean that development of posture is not a sort of natural maturation. There is a natural development of posture, and we can still correct it.

Note

1. B. Reeves, and C. Nass, *The Media Equation* (Cambridge University Press, Cambridge, 1996).

References

Braitenberg, V. (1984): *Vehicles: Experiments in Synthetic Psychology*. MIT Press, Cambridge.

Brand, P. W., and Yancey, P. (1997): *The Gift of Pain*. Zondervan Publishing House, Grand Rapids, Michigan.

Clynes, M., Jurisevic, S., and Rynn, M. (1990): Inherent Cognitive Substrates of Specific Emotions: Love is Blocked by Lying but not Anger. *Percept. Motor Skills* 70: 195–206.

Clynes, M. (1977): *Sentics: The Touch of Emotions*. Anchor Press/Doubleday, New York.

Damasio, A. R. (1994): *Descartes' Error; Emotion, Reason, and the Human Brain*. Putnam, New York.

Gershon, M. D. (1998): *The Second Brain: The Scientific Basis of Gut Instinct and a Groundbreaking New Understanding of Nervous Disorders of the Stomach and Intestines*. HarperCollins, New York.

Johnstone, B. (1999): Japan's Friendly Robots. *Technol. Rev.* May/June 63–69.

LeDoux, J. E. (1996): *The Emotional Brain*. Simon and Schuster, New York.

Ortony, A., Clore, G. L., and Collins, A. (1988): *The Cognitive Structure of Emotions*. Cambridge University Press, Cambridge.

Picard, R. W. (1997a): *Affective Computing*. MIT Press, Cambridge.

Picard, R. W. (1997b): Does HAL Cry Digital Tears? Emotion and Computers. In D. G. Stork, ed., *HAL's Legacy: 2001's Computer as Dream and Reality*. MIT Press, Cambridge.

Roseman, I. J., Antoniou, A. A., and Jose, P. E. (1996): Appraisal Determinants of Emotions: Constructing a More Accurate and Comprehensive Theory. *Cogn. Emotion* 10: 241–277.

Walter, W. G. (1950): An Imitation of Life. *Sci. Am.* 182 (5): 42–45.

Emotional Agents for User Entertainment: Discussing the Underlying Assumptions

Paola Rizzo *

IP-CNR – Institute of Psychology
National Research Council of Italy
Viale Marx 15, I-00137 Rome, Italy
paola@ip.rm.cnr.it
http://pscs2.irmkant.rm.cnr.it/users/paola/html/home.html

Abstract. Emotional agents can play a very important role in all those computer applications where the user does not (only) want to perform a task, but (also) to be entertained or to have a more engaging experience than with traditional systems. In fact, it is commonly assumed that the agents' ability to process and display affective states, and to show emotional reactions, is crucial for causing a more enjoyable interaction between agent and user. This work proposes to analyze such assumption from a critical viewpoint, identifying several open issues that are worth debating in the community and studying through empirical methods.

User interfaces are becoming more and more friendly, adaptive, and sensitive to the user's needs, preferences and wants. Along with this trend, computers are starting to comply with the user's wish to be entertained while or besides performing some "serious" task. For example, instructional software systems are being enhanced by lively anthropomorphic or animal-like tutors [7, 15]; multimedia presentation systems might look more likeable if they present material through expressive animations [1]; PC desktops may be more exciting if populated by "virtual pets" that can interact with the user and with one another in interesting ways [5].

The issue of user's entertainment is to be "taken seriously", since it can deeply influence the effect of software applications, and even make the difference between a successful and an unsuccessful interface. In fact, computer-based entertainment can not only be an interesting purpose in itself (as in the case of PC games, multi-user virtual environments, or software toys), but can also increase the user's satisfaction, boost her motivation to interact with an application (which, for example, is very important in the case of "edutainment"), help relieve the tensions that are usually related to the performance of difficult tasks, and give the opportunity to have a break from hard work. Such properties of entertainment come from its basically emotional nature and impact.

* The author is currently supported by a scholarship from the Committee 12 (Information Science and Technology) of the National Research Council of Italy. She is indebted to Maria Miceli and Amedeo Cesta for their useful comments, and to Daniela Petrelli and Elisabeth André for suggesting some relevant references.

Interfaces can be made more enjoyable by means of "believable agents" (also known as "lifelike computer characters", "synthetic agents", "virtual actors", and "animate characters"): these computational systems are built with the purpose of provoking in the user the "illusion of life" or, in other words, the impression of interacting with a "living" virtual being that has its own feelings, desires, and beliefs. [1] A strong argument in favor of using believable agents in interfaces is that they make interacting with computers much easier and nicer, enable (especially naive) users to adopt communicative styles similar to those typical of human-human communication [1], and can increase the level of interactivity and socio-emotional engagement produced by traditional applications. There are also some arguments criticizing the usability of agents in interfaces: in fact, agents might compromise the user's feeling of control over the tasks being performed, or might generate in the user wrong expectations about their own capabilities [11, 18]. Nevertheless, among the advocates of lifelike computer characters, it is also commonly believed that the agents' ability to process and display affective states, and to show emotional reactions, is crucial for improving the agents' believability, for eliciting emotions in the user, and consequently for causing a more entertaining interaction between agent and user [12, 2]. Therefore, several research projects are aimed at realizing *emotional agents* (e.g. [4, 14, 8]).

These agents are a promising direction of research on entertaining interfaces, but their realization and use seem to rely on assumptions that have not yet been made explicit and studied in great depth; this work proposes to outline and analyse them. In our view, such an analysis can be useful in two ways: on the one hand, it can help understand what are the main problems to address when realizing emotional agents; on the other hand, it can drive the empirical validation of the most common hypotheses regarding the relationships between entertainment, emotions, and believability, in order to build more effective agents.

Emotional agents are usually inspired by characters realized in traditional media like drama, cinema, cartoons, and seem to be founded on the following hypotheses:

Hypothesis 1: entertainment is a function of the character's ability to cause the "illusion of life" (believability);

Hypothesis 2: entertainment is a function of the character's ability to induce emotions.

These hypotheses, although difficult to verify empirically, are hardly questionable; they can actually be considered as basic assumptions. In fact, we have many examples of their validity from traditional media: the more believable a character is, the more likeable it is, and the emotions it can induce greatly contribute to the spectator's entertainment.

[1] This notion of believability is different from the perceived "credibility" of computer applications, which refers to the delivery of trustworthy information. [19]

In our view, the assumptions above are motivated by more specific and usually implicit hypotheses, that generally drive the actual realization of emotional agents:

Hypothesis 3: believability crucially depends on the agent's ability to show emotional reactions;
Hypothesis 4: believability also depends on the manifestation of a marked personality, which in its turn is based on an individual style of emotional behaviors;
Hypothesis 5: the user's emotional responses to an agent depend on the latter's emotional behavior;
Hypothesis 6: the agent's ability to show emotional reactions depends on its ability to represent and process emotional states.

Such hypotheses appear to be related to each other according to an instrumental hierarchy, as shown in figure 1: from bottom up, agent designers usually assume that the ability to represent and process affective states, which can be brought about by events affecting an agent's goals, preferences, or values, enables the agent to display emotional reactions; these in turn are deemed essential for enhancing the agent's believability, for characterizing its unique personality, which also contributes to believability, and for arousing emotions in the observer, possibly through an empathic process; finally, the agent's believability, and its ability to induce affective reactions in the user, are expected to cause the latter to be entertained. The user's entertainment can be an end in itself, or it can be functional to other affective states, such as satisfaction, ease, motivation, and the like, that can influence the effectiveness of several kinds of computer applications.

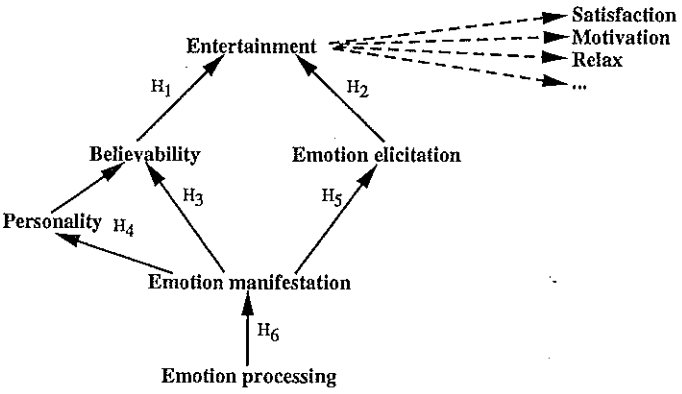

Fig. 1. The basic hypotheses underlying emotional agents.

Although these hypotheses might intuitively sound correct, several counterexamples can be found that seem to contradict or at least weaken them. For

instance, with regard to hypotheses 3 and 4, a pedagogical agent whose task is to help students learn new procedures might not have any particular personality nor show emotions, but it might still look believable to its users and therefore increase their level of satisfaction towards the system and their engagement in the learning process [15]. Concerning hypotheses 5 and 6, agents like ELIZA [20] or JULIA [9] neither "have" nor display any emotion, but are very successful in arousing affective reactions in the users. Along this line of reasoning, many questions related to the social attribution of human-like features to agents (see for instance [13]) should be taken into account in order to better understand the value of the hypotheses. For example, it is worth wondering how much agents which are not able to show emotions, but have goal-based personalities [16, 17], could be attributed affective states and induce emotional reactions in the users. These arguments point out that the hypotheses above should not be taken for granted, but rather should be empirically tested.

Some experimental investigations about the effectiveness of animate characters (e.g. [10, 6]; for a review see [3]) have focused mainly on two aspects: (a) some objective measure of the users' performance (like the number of solved problems, or the percentage of remembered items after their presentation), in order to see whether the performance is improved by the presence and/or help of an agent, and (b) several subjective measures of the characteristics (like intelligence, likeability, utility) that can be attributed by the users to the agent, and that are supposed to mediate the agent's effectiveness. Unfortunately, these studies are difficult to compare with one another because of their different experimental settings, dependent variables, etc., and consequently it is also hard to derive from them some useful generalizations. But, more importantly, the empirical works carried out so far do not seem to be based on a causal model that could explain the possible interrelationships among variables (i.e. among the subjective and the objective measures). In our view, the hierarchy of hypotheses proposed in this work can be considered as a model, focused on the issue of user entertainment, where the cause-effect links between variables are explicitly represented; as such, it could help devise new experiments aimed, on the one hand, at testing those links one by one and, on the other hand, at verifying whether the model should be modified by the introduction of other variables and related links/hypotheses.

References

1. E. André, T. Rist, and J. Müller. Integrating reactive and scripted behaviors in a life-like presentation agent. In *Proceedings of the Second International Conference on Autonomous Agents*, pages 261–268. ACM Press, New York (NY), 1998.
2. J. Bates. The role of emotion in believable agents. *Communications of the ACM*, 37(7):122–125, 1994.
3. D.M. Dehn and S. van Mulken. The impact of animated interface agents: A review of empirical research. *International Journal of Human-Computer Studies*, 1999. To appear.

4. C. Elliott. *The Affective Reasoner: A Process Model of Emotions in a Multi-Agent System.* PhD thesis, Northwestern University, Evanston (IL), May 1992. Technical Report no. 32.
5. A. Frank, A. Stern, and B. Resner. Socially intelligent virtual petz. In *Proceedings of the AAAI'97 Spring Symposium on "Socially Intelligent Agents". AAAI Technical Report FS-97-02*, pages 43–45. AAAI Press, Menlo Park (CA), 1997.
6. J. C. Lester, , S. A. Converse, S. E. Kahler, S. T. Barlow, B. A. Stone, and R. S. Bhogal. The persona effect: Affective impact of animated pedagogical agents. In S. Pemberton, editor, *Human factors in computing systems: CHI'97 Conference Proceedings*, pages 359–366. ACM Press, New York (NY), 1997.
7. J. C. Lester and B. A. Stone. Increasing believability in animated pedagogical agents. In W. L. Johnson, editor, *Proceedings of the First International Conference on Autonomous Agents*, pages 16–21. ACM Press, New York (NY), 1997.
8. C. Martinho and A. Paiva. Pathematic agents: Rapid development of believable emotional agents in intelligent virtual environments. In *Proceedings of the Third International Conference on Autonomous Agents*, pages 1–8. ACM Press, New York (NY), 1999.
9. M. L. Mauldlin. Chatterbots, tinymuds, and the Turing test entering the Loebner prize competition. In *Proceedings of the 12th National Conference on Artificial Intelligence (AAAI'94)*. AAAI Press, Menlo Park (CA), 1994.
10. S. Van Mulken, E. André, and J. Müller. The Persona Effect: How Substantial Is It. In *Proceedings of HCI'98*, pages 53–66, Sheffield, UK, 1998.
11. D. A. Norman. How might people interact with agents. *Communications of the ACM*, 37(7):68–71, 1994.
12. R. Picard. *Affective Computing.* The MIT Press, Cambridge (MA), 1997.
13. B. Reeves and C. Nass. *The media equation: How people treat computers, television and new media like real people and places.* Cambridge University Press, Cambridge (MA), 1996.
14. W. S. Reilly. *Believable Social and Emotional Agents.* PhD thesis, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, May 1996. Technical Report CMU-CS-96-138.
15. J. Rickel and W. L. Johnson. Integrating pedagogical capabilities in a virtual environment agent. In W. L. Johnson, editor, *Proceedings of the First International Conference on Autonomous Agents*, pages 30–38. ACM Press, New York (NY), 1997.
16. P. Rizzo. *Personalities in believable agents: A goal-based model and its realization with an integrated planning architecture.* PhD thesis, Center of Cognitive Science, University of Turin, Turin, Italy, 1998.
17. P. Rizzo, M. M. Veloso, M. Miceli, and A. Cesta. Goal-based personalities and social behaviors in believable agents. *Applied Artificial Intelligence*, 13(3):239–271, 1999. Special Issue on "Socially Intelligent Agents", edited by Kerstin Dautenhahn and Chisato Numaoka.
18. B. Schneiderman and P. Maes. Direct manipulations vs. interface agents: Excerpts from debates at IUI'97 and CHI'97. *Interactions*, 4(6):97–124, 1997.
19. S. Tseng and B. J. Fogg. Credibility and computing technology. *Communications of the ACM*, 42(5):39–44, 1999.
20. J. Weizenbaum. ELIZA–A computer program for the study of natural language communication between man and machine. *Communications of the Association for Computing Machinery*, 9(1):36–45, 1965.

Keep Cool: The Value Of Affective Computer Interfaces In A Rational World

Erik Hollnagel
Graduate School for Human-Machine Interaction
University of Linköping, Sweden

1 Introduction

The standard model for human-machine interaction is based on the communication paradigm formulated by Claude Shannon in the early 1940s (Shannon & Weaver, 1969). This paradigm describes the exchange of messages between a sender and a receiver, with emphasis on the capacity of the information channel and how messages can be distorted by noise. The paradigm was eagerly adapted by a science of psychology trying to break lose from the confines of behaviourism (Attneave, 1959; Miller, 1967), and provided the foundation for the psychological models that became part and parcel of information processing psychology, cognitive psychology and cognitive engineering (Lindsay & Norman, 1977; Newell & Simon, 1972; Wickens, 1984). It also corresponded well to the endeavour to describe human actions as the result of a rational choice, exemplified by the models for human decision making (Lee, 1971; Edwards, 1954).

When the study of human-machine interaction was devoured by the swelling interest for human-computer interaction in the 1980s, the basic paradigm remained. Human-machine interaction was seen as the exchange of messages (signals and control actions) across an interface, and the information processing view reigned supreme. The processes, cognitive or otherwise, that provided the basis for the communication and interaction were all "cold" rather than "hot", i.e., excluding emotions and affect (Abelson, 1963). Information processing psychology several times tried to develop theories that included emotions and affect, but with very limited success (Mandler, 1975; Simon, 1967).

2 The Efficiency Of Human-Human Communication

As the use of computers has spread in an almost epidemic fashion to all areas of life and work, enormous efforts have been was put into making human-machine interaction as easy and efficient as possible. One example of that is the campaign to develop systems that are user-friendly and require little or no learning. Another is the desire to use multi-media to enhance the communication and comprehension. A third is the effort to build adaptive systems and interfaces that "automatically" provide the user with the right information, in the right form, at the right time.

In all of this human-human communication has often been used as a paragon. The smoothness, versatility, and efficiency of communication between people is a distant goal for human-machine interaction, and it is therefore tempting to look for features of natural communication that are missing in artificial communication. One obvious candidate is emotion or affect, i.e., the fact that people have affects and also are uncannily effective in recognising the affects and emotions of others. Since affects clearly are missing from the information processing paradigm, it is tempting to consider whether affective interfaces would provide the coveted quantum leap in human-machine interaction.

2.1 Emotions And Control

It is not impossible that affective interfaces may serve a therapeutic purpose, in the sense that they make the user feel better. The main reason for using an affective interface should nevertheless be that it improves the efficiency of the communication, hence the ability to retain control. Generally speaking, the purpose of communication is to control the behaviour of the receiver, whether it is a human or a machine, and anything that enhances the ability to control is therefore worth considering. One effective way of describing human (and machine) performance is by means of a cyclical, rather than a linear, model (Hollnagel, 1998; Neisser, 1976). In Figure 1 this principle is used to show the communication between two agents, in this case two humans.

The model describes how actions are determined by the current understanding, which in turn depends on the evaluation (or interpretation) of the feedback. In communication, the message of one agent becomes the response of another, and *vice versa*. The model emphasises that performance is both feedback and feedforward controlled. As pointed out already by Conant & Ashby (1970, p. 92), feedback (error-controlled) regulation is inferior to feedforward (cause-controlled) regulation. Despite that observation, practically all information processing models of operators exemplify error-controlled regulation.

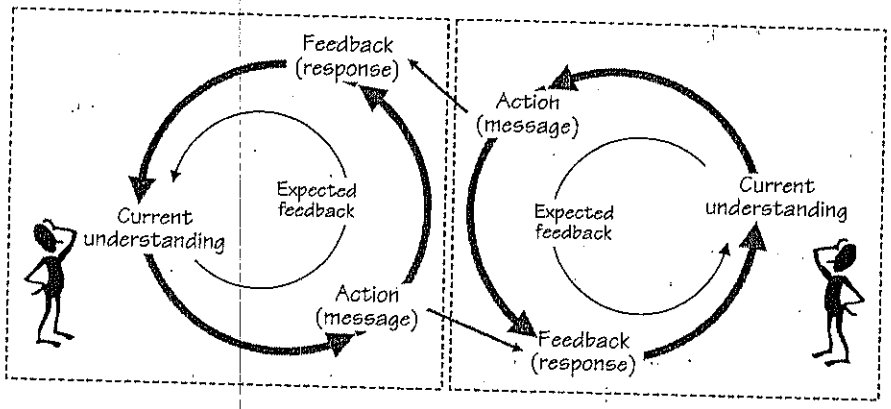

Figure 1: The cyclical communication paradigm.

2.2 Understanding The Mood Of The Other

The model in Figure 1 can easily be extended to address the role of affects by noting that one aspect of the current understanding is the assessment of the mood or emotional state of the communication partner. The emotional state of a person plays a major role in determining the reaction. "Hot" cognition may narrow the scope of interpretations and conclusions may be based on what was expected rather than what actually happened. Decisions may become coloured by highly subjective and temporary criteria, and otherwise reasonable alternatives may be disregarded. (This happens not just during affective states, but also during high stress and discomfort.) It is therefore important to take this into consideration when formulating and expressing the message. In normal life, and specifically during work, affective-emotional human-human interaction is usually not encouraged. On the contrary, we often try to counter the affective state of the other person to avoid the situation from boiling over or to prevent getting into a heated argument. The reason is simply that reciprocal affects may bring the system out of control. On the other hand, if the affective state of the other is correctly recognised, control can be regained and stability ensured.

3 Affective HMI – Dream Or Nightmare?

The model and the principles can be applied to human-machine interaction (HMI) as well as to human-human interaction. The purpose of HMI is not primarily to make the user feel good, but to enable the joint system to maintain control of a situation. The value of affective interfaces must therefor be considered in relation to this overall goal. One possibility is that the interface (or rather, the machine) is able to recognise the mood or affective state of the user. If this can be done in a reliable fashion, and if it further is possible to adjust the

functioning of the HMI accordingly, mood recognition may conceivably have a beneficial effect. It is in a way a logical extension of the principles to design interfaces for everyone (Newell, 1993), although extending the scope of normal and disabled users to include calm and agitated users as well. Another possibility is that the interface also can show or represent emotions, i.e., be an affective interface. This could be seen as an extension, although a significant one, of adaptive interfaces. It would therefore also comprise the risks of adaptive interfaces, and possibly amplify them. The main risk is that two systems that reciprocally adapt to each other very easily run the risk of going into an unstable region. If the response to a recognised affect is inappropriate, which in control terms means that the damping of the system is inadequate, then this may all too easily lead to an amplification of deviations, rather than the neutralisation and homeostasis that is really wanted (Maruyama, 1963).

For this reason alone, it is highly uncertain whether there is anything gained by having an affective interface, although there may be a potential advantage in having one that can recognise affects. The question to be pondered by researchers in HMI (and HCI) is therefore not whether an affective interface, in either sense, is possible or technologically feasible, but whether it is required. Once affective interfaces are available, and this will undoubtedly happen at some time, they may join the many other technological solutions that are in search of a problem. It would be nice if, for once, a genuine need was established first. At the moment, it is my belief that the virtue of human-human communication are exaggerated, and that the risks of affective interfaces need to be understood better. This will require a sorely needed revision of the dominating paradigms for HMI.

4 References

Abelson, R. P. (1963). Computer simulation of "hot" cognition. In S. S. Tomkins & S Messick (Eds.), *Computer simulation of personality*. New York: Wiley.

Attneave, F. (1959). *Applications of information theory to psychology: A summary of basic concepts, methods, and results*. New York: Holt, Rinehart & Winston.

Conant, R. C. & Ashby, W. R. (1970). Every good regulator of a system must be a model of that system. *International Journal of Systems Science, 1*(2), 89-97.

Edwards, W. (1954). The theory of decision making. *Psychological Bulletin, 51*, 380-417.

Motivations behind modeling emotional agents: Whose emotion does your robot have?

Thomas Wehrle

FPSE, University of Geneva, Switzerland
9, route de Drize
CH-1227 Carouge-Genève
wehrle@acm.org

Abstract

A communality between research in *Artificial Intelligence* and *Synthetic Emotion* is that it seems in both cases to be rather difficult to give an acceptable definition of the naturally occurring counterpart. One could speculate whether this is due to the multiplicity of the nature of both phenomena or due to a categorical misconception. In this paper I try to briefly outline a number of different motivations for modeling emotions, and to relate those motivations to two different principal design approaches for computational models of emotion. From these two aspects, together with our current assumptions about mechanisms underlying human emotions, I conclude with some speculations about *adaptation* in affective systems, and some implications of the notion of *grounding* emotions in adaptive systems.

Introduction

> It seems certain that, as we understand more about cognition, we will need to explore autonomous systems with limited resources that nevertheless cope successfully with multiple goals, uncertainty about environment, and coordination with other agents. In mammals, these cognitive design problems seem to have been solved, at least in part, by the processes underlying emotions. (Oatley 1987)

This is an example of why one might want to model an emotional agent. I will try to show that there are other motives, and that there are also different principal approaches to model emotions. I will briefly sketch some theory on emotion from a psychological point of view to illustrate our current assumptions about the nature of emotions, and about the structurally and functionally different subsystems that seem to be involved in emotions. Based on these three considerations (motives, modeling approaches, and mechanisms) I would like to make some speculations about possible principles of adaptation in computational models of emotion. I will also try to argue that the notion of *grounding* emotions in adaptive systems raises some interesting questions, and is dependent upon at least the same three aspects.

Different motives for modeling emotional agents

Science (Psychology, Neuroscience, Cognitive Science, Biology, etc.)

Modeling an emotional system can be an attempt to instantiate parts of a theory about a natural phenomenon with a computer program or a robot. The researcher hopes that such a system helps to improve the formalization, operationalisation, and internal consistency of theoretical postulates. The system is further expected to allow an examination of the required number and types of criteria needed for successful theoretical prediction (allowing also to compare theories differing in this respect) and to improve intuitive understanding of these parameters. Such systems can also be very useful in teaching and visualization. This synthetic approach complements the traditional approach of the analysis of behavioral data (and this data can be used to measure the quality of the model).
Criteria: description, explanation, and prediction
Main motivation: Improve our knowledge about the nature of emotion and its implications

Engineering

It seems likely that here the motive behind modeling an emotional agent is an indirect one. The engineer is primarily interested in constructing a useful artifact. In adopting some real or hypothesized natural principles the engineer hopes to increase the system performance in terms of task achievement and costs. The extent to which the principles that inspired the system eventually get into the system or the form and adequacy of the translation is of no significance. Emotion theories may serve as heuristics in finding a good solution to a problem, or to metaphorically describe the state or behavior of a system.
Criterion: Performance
Main motivation: Building good artifacts for a concrete task

Human Computer Interaction

Modeling emotions in a system that interacts with humans is a special case of engineering where human behavior and affectivity plays a significant role. In this case theoretical knowledge about emotions can be applied. In adopting some real or hypothesized natural principles the engineer hopes to increase the system performance in terms of acceptance and usability with respect to the user. Again, it seems likely that such a system does not necessarily need to represent emotion constructs in any form to generate the desired behavior.

Criteria: Performance, acceptance, and usability

Main motivation: Improve human computer interaction

Technology would in this line of reasoning be the attempt to extract general principles from the engineering work.

Two kinds of modeling approaches

Almost since the beginning of the computer age there have been exciting attempts to deal with emotions in one way or another. For examples of good overviews on existing systems I refer to Pfeifer (1988), Picard (1997) and the web page [3] of Hudlicka and Fellous. Given the different concerns of computational models of emotion and affective behavior it is no surprise that we find a whole variety of different modeling approaches. Wehrle and Scherer (1995) have argued that it might be useful to distinguish two classes of computational models of emotion: black box models and process models. Although the two approaches should not been seen as exclusive, they differ in the degree of abstraction of intervening variables[1].

Black box modeling

The purpose of black box models is to produce outcomes or decisions that are maximally similar to those resulting from the operation of naturally occurring systems, disregarding both the processes whereby these outcomes are attained as well as the structures involved (see also Phelps and Musgrove 1986, p. 161).

Although such models provide little information concerning the mechanisms involved, they are very useful for practical decision making and for providing a sound grounding for theoretical and empirical study. In particular, they can help to investigate necessary and sufficient variables. System performance (e.g. quality of classification and computational economy) as well as cost of data gathering are important criteria for assessing the quality of the chosen computational model. Since black box models focus on the input-output relationship, they make few claims whatsoever concerning the nature of the underlying processes.

The simplified example in figure 1 illustrates the idea:

[1] I also think that models that include underlying mechanisms are necessarily based on black box models at a certain level of abstraction.

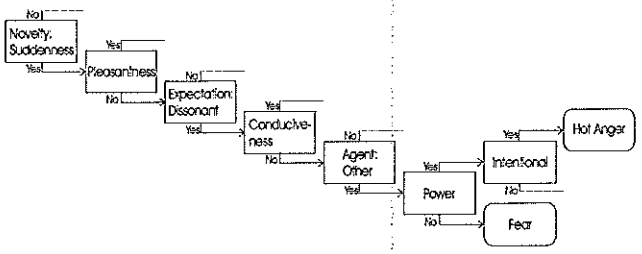

Fig. 1: Part of a decision tree representation of Scherer's Appraisal Theory concerning the cognitive component. Note that the theoretically postulated sequence of appraisal is violated (the intention check should precede the power check) but the black box output is correct.

To my knowledge most existing implementations (whether they simulate or reason about emotions) are based on black box models (and mostly symbol systems).

Process modeling

The purpose of process modeling is usually the attempt to simulate naturally occurring processes using hypothesized underlying mechanisms. Clearly, this approach is considerably more ambitious than the black box approach. In the case of psychobiological process models, one needs to specify the effects of causal input factors on intervening process components (often hypothetical constructs), as well as the structural interdependencies of the internal organismic mechanisms involved.

To my knowledge few systems have attempted to synthesize emotional behavior on this level of biologically plausible mechanisms (e.g. Armony et al. 1995). Given the complexity of involved components, models of this kind seem to be as difficult to realize as they are useful to increase our knowledge. In the Geneva Emotion Research Group we have tried to implement a very simple emotional problem solver inspired by Toda's Social Fungus Eater (Toda 1962, 1982). The system is described elsewhere (Wehrle 1994a 1994b, and on my web page [1]). Toda describes the behavior of the hypothesized Fungus Eater in terms of urges and cost functions (internal vs. external). In our proposed model we used a simple hedonistic principle based on an energy concept which also allows adaptation. Whereas for Toda the concept of urges seems similar to emotions, we have so far only implemented some very basic urges as value scheme, and regard emotional behavior as an emergent property of the latter.

Possible contributions of appraisal theory

In the following, several theoretical postulates from two prominent theories concerning appraisal and emotion are presented to illustrate the function and conceptualization of emotion in human beings from a psychological point of view. I feel that appraisal theory could potentially offer some heuristics for the design of emotional systems. Since it represents the mainstream of current emotion psychology

it may also be used as a source of ideas about how psychologists envisage the complexity of affective behavior. Later on in my argumentation I will also use the assumption that the mechanisms underlying emotions are functionally and structurally heterogeneous to question the notion of grounding emotions in an artifact. Even though the theory is quite general with respect to the emotional situations, it has nevertheless been based upon the specific physical properties of the human organism.

> Scherer has argued that the elicitation and differentiation of emotion can be most economically explained by a process of cognitive event appraisal that reflects the significance of an event for an individual's goal and value structure, his or her coping potential, and the socio-normative significance of the event. The component process theory (see Scherer, 1988, 1993, for details) posits relatively few basic evaluation criteria and assumes a sequential processing of these criteria in the appraisal process. Figure 2 shows the major appraisal dimensions or "stimulus evaluation checks" (SECs) which are considered to be sufficient to account for the differentiation of all major emotions. (Wehrle, Kaiser, Schmidt, and Scherer, submitted)

> Scherer has suggested that emotion can be defined as an episode of temporary synchronization of all major subsystems of organismic functioning represented by five components (cognition, physiological regulation, motivation, motor expression, and monitoring/feeling) in response to the evaluation of an external or internal stimulus event as relevant to central concerns of the organism. It is claimed that while the different subsystems or components operate relatively independently of each other during non- emotional states, dealing with their respective function in overall behavioral regulation, they are recruited to work in unison during emergency situations, the emotion episodes. These require the mobilization of substantial organismic resources in order to allow adaptation or active responses to an important event or change of internal state. The emotion episode begins with the onset of synchronization following a particular stimulus evaluation pattern and ends with the return to independent functioning of the subsystems (although systems may differ in responsivity and processing speed). Since stimulus evaluation is expected to affect each subsystem directly and since all systems are seen to be highly interrelated during the emotion episode, regulation is complex and involves multiple feedback and feedforward processes. For this reason, it is assumed that there is a large number of highly differentiated emotional states, of which the current emotion labels capture only clusters or central tendencies of regularly recurring modal states. ... Scherer has suggested that subjective experience or feeling can be conceptualized as the reflection of the changes in all other emotion components, including the different neurophysiological and motor subsystems as well as changes in motivation and particularly the cognitive appraisal system. Leventhal and Scherer (1987) have made a first attempt to illustrate the way in which Scherer's stimulus evaluation checks could be performed on the sensory-motor, the schematic and the conceptual level. Obviously, the nature of the resulting emotion is likely to be quite different depending on the level of its antecedent appraisal, particularly with respect to the conscious experience of the episode. (Kaiser and Scherer 1998)

> Arnold (1960) defined emotions as "felt action tendencies" because, as she argues, a felt tendency is what characterizes such experience and differentiates it from mere feelings of pleasantness or unpleasantness; different action tendencies are what characterize different emotions. Action tendencies or, more generally, changes in action readiness are not only important in emotional experience but are also central in the analysis of emotion as such. Action readiness is what links experience and behavior; felt action readiness can be considered a reflection of the actual state of behavioral readiness.... Emotions involve states of action readiness elicited by events appraised as emotionally relevant... Events are appraised as emotionally relevant when they appear to favor or harm the individual's concerns. (Frijda, Kuipers, and ter Schure)

An attempt to summarize the basic concepts can be found in figure 2. I took some freedom to include some further assumptions about the postulated sequential evaluation and to apply the concept of the different levels of processing also to the expressive component.

Note however, that the theory is based on constructs (for which there is empirical evidence). More directly accessible to us are observable behaviors, some of which I list below:

- Social behavior, problem solving, action selection, etc.
- Facial expression
- Voice and utterances (prosody and syntax)
- Physiological responses (skin conductance, EMG, ECG, EEG, etc.)
- Verbal report of subjective feeling and mood
- Expression in arts

I prefer to leave it open whether this behavior can be seen as an antecedent, as a result, or as a constituent of emotions. My point here is that the great variety of behavior in which emotions play a role seems to reflect the functionally and phylogenetically different mechanisms underlying emotion that we assume.

To further complicate the issue there is evidence that the roles that the different subsystems play, and their connections to the expressive components, vary significantly among different emotions. I refer to the respective literature (concerning psychophysiological measures e.g., Stemmler 1989; and concerning mechanisms, e.g., LeDoux 1996; Panksepp 1993; Davidson 1994).

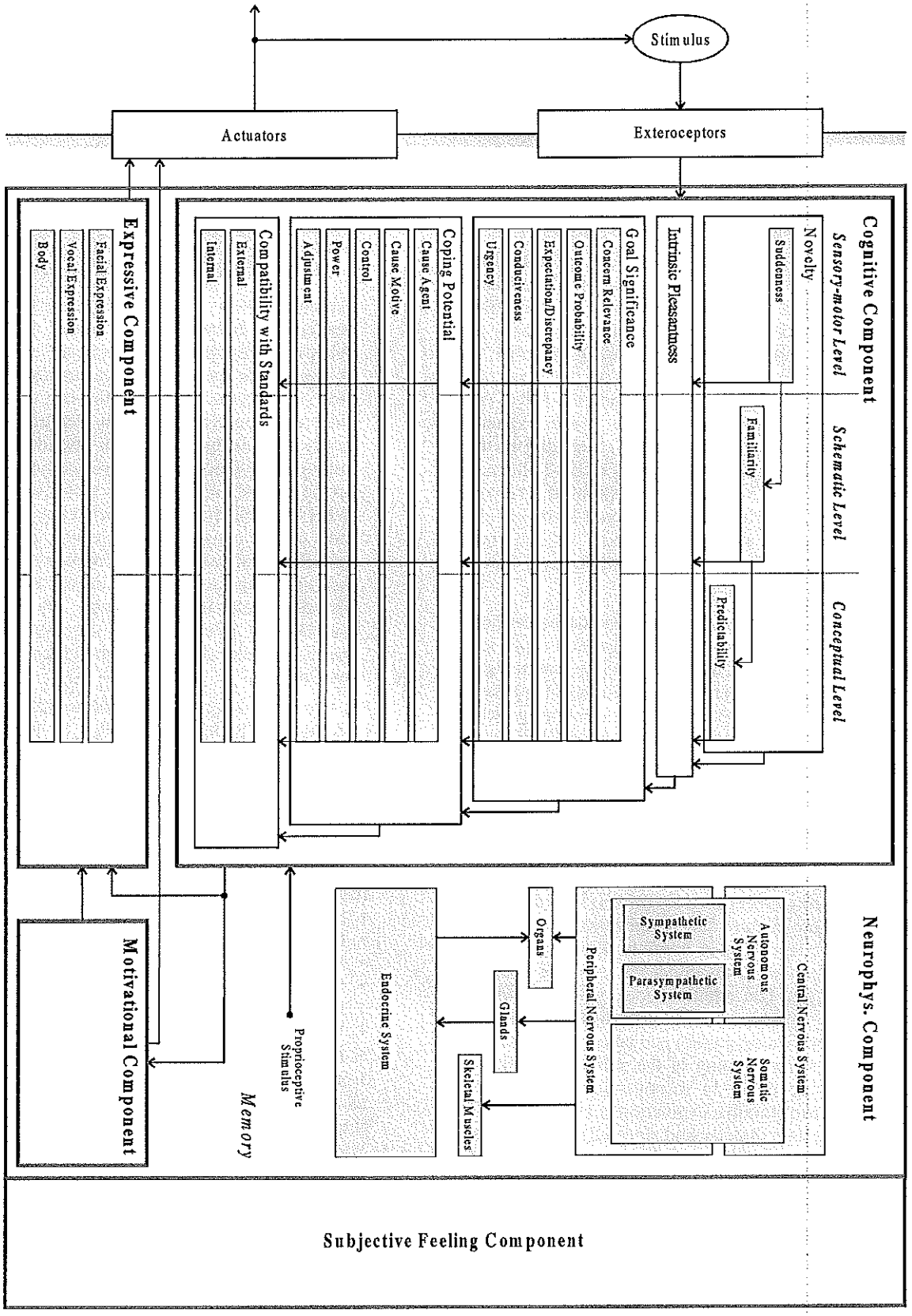

Fig. 2: An attempt of a formal representation of Scherer's Appraisal Theory

Adaptation

If we regard emotions as the result of the interaction and synchronization of rather complex subsystems in response to situational, environmental, and physiological properties, then emotions must reflect an adaptive system. There is evidence that affective behavior is also much more flexible than behavior generated by simpler fixed pattern response systems. There are large individual and cultural differences, and emotional reactions also vary over different life stages. One could speculate about three different levels of processing on which individual adaptation might be implemented:

- Level of single values: One likely candidate where adaptive mechanisms might play a role is in the parameters of the involved stimulus evaluation systems and their influence on the organization of behavior. Example: The attribution of self-competence and power can be derived from experience and changes over time and domain.
- Level of emotion specific value patterns: There seems to be evidence that the significance of different appraisal dimensions varies among emotions. I propose that this is a second candidate for the implementation of adaptive mechanisms. Example: A diplomat might have to learn to reduce the importance of his or her estimated coping potential to avoid aggressing other diplomats in an anger situation.
- Level of action readiness patterns: Since it has been argued that the affective systems allow an organism to generate behavior in an efficient and adapted fashion, the association of an action repertory to an emotional state might also be modified in terms of reflection and success evaluation. Example: In elevators where people are necessarily *physically* closer to each other than they actually would like to be (embarrassment), many people learn to establish a larger *social* distance by actively avoiding eye contact.

Empirical data

Eventually, the quality of a synthetic approach aiming to model human emotion will be compared with the observation of naturally occurring systems. Empirical studies serve also to improve our knowledge about the nature of emotion, the expressive components, and the function of emotion. Most empirical studies rely on verbal report, which is not necessarily the most promising way, since the necessary degree of awareness of an emotional state, and the influence of self-reflection on the emotional process are not yet clear. Isolated studies of certain aspects of emotion such as facial expression on the basis of strong static stimuli seem similarly unpromising. Masanao Toda's proposal of the Fungus Eater experiment, although already described more than 35 years ago, still seems to be a fresh and rarely pursued idea (see also Pfeifer 1994). The Geneva Emotion Research Group is conducting empirical studies within this paradigm with several NSF projects (see [2]; Kaiser and Wehrle, 1996).

Questions and conclusions

I have tried to show that emotions seem to be involved in many functionally different behaviors, and that this impression is also reflected in the many heterogeneous mechanisms that we suppose underlie emotions. I also tried to show that there are different motives behind affective computing, and accordingly different techniques for modeling and representing emotions in an artifact which abstract from the underlying mechanisms to different degrees. I addressed the issue of adaptation by making some speculations about how it could be realized within the framework of appraisal theory.

Unfortunately, it was not possible to fully elaborate the relations between the different aspects of modeling emotions that I described, but I would like to at least conclude with some questions concerning the attempt to *ground* emotion in an *adaptive* system.

As Braitenberg (1984) and others have shown, the observed complexity of behavior does not necessarily need to be reflected in the underlying mechanism. However, if we are to believe that there are different mechanisms underlying emotions, and that they get expressed in functionally different behaviors, the question remains to what extent the supposed structural complexity of involved mechanisms can be abstracted. For certain purposes it has successfully been done in several systems, but the question becomes more important if a system attempts to ground emotions.

If emotion categories are either seen as emergent labels for the evaluation of prototypical situations or events (modal emotions), or as evolutionarily achieved response programs (basic emotions), then we have to expect that emergent emotions in robots will be different than human emotions, i.e. all emotion systems of natural or artificial agents should be expected to be incommensurable if the niches, and probably even more importantly, the physical properties of the agent bodies differ significantly.

Even if we can introduce a type of value system to a robot, I personally feel that grounding somehow implies that we allow the robot to establish its own emotional categorization which refers to its own physical properties, the task, the properties of the environment, and the ongoing interaction with this environment. In this case talking about mechanisms seems unavoidable.

We can put *human* emotion categories into an artificial agent. It might be useful to use emotions as design heuristics for adaptive systems, or to describe their behavior, but can we hope to ground these categories that have evolved in a different system? Is that a reasonable goal? With the framework of appraisal theory I also tried to demonstrate that a theory might be able to abstract from

the niche to a certain extent but to a smaller extent from the properties of the agent.

One might argue that implementing a certain value system for an autonomous agent is equivalent to introducing a sort of emotion system in a broad sense. If one takes the appraisal dimensions proposed in emotion psychology as a basis for the value system, one might profit from the possibility to describe the resulting behavior in known emotion terms. Whether or not the chosen values are appropriate depends on the task an agent is designed for, its physical properties, and the properties of the environment. In summary, we can put human emotion categories into an artifact but we will probably not be able to ground them if the properties of the artifact, its tasks, and its environment are significantly different. Or we can let a system develop its own categories, but those might not be the categories that we are used to, and we do not necessarily have to dub these system states with emotion terms.

References

Armony, J. L., Servan-Schreiber, D., Cohen, J. D., and LeDoux, J. E. 1995. An anatomically constrained neural network model of fear conditioning. *Behavioral Neuroscience* 109:246-257.

Braitenberg, V. 1984. *Vehicles: Experiments in synthetic psychology*. Cambridge, Massachusetts: MIT Press.

Davidson, R. J. 1994. Asymmetric brain function, affective style, and psychopathology. The role of early expression and plasticity. *Development and Psychopathology* 6:741-758.

Frijda, N. H., Kuipers, P., and ter Schure, E. 1989. Relations among emotion, appraisal, and emotional action readiness. *Journal of Personality and Social Psychology* 57:212-228.

Kaiser, S., and Scherer, K. R. 1998. Models of 'normal' emotions applied to facial and vocal expressions in clinical disorders. In *Emotions in psychopathology*, eds. W. F. Flack, Jr., and J. D. Laird, 81-98. New York: Oxford University Press.

Kaiser, S. and Wehrle, T. 1996. Situated emotional problemsolving in interactive computergames. In *Proceedings of the VIIIth Conference of the International Society for Research on Emotions, ISRE'96*. 276-280. Storrs, CT: ISRE Publications.

LeDoux, J. E. 1996. *The emotional brain*. New York: Simon and Schuster.

Oatley, K. 1987. Editorial: Cognitive Science and the understanding of emotions. *Cognition and Emotion* 1:209-216.

Panksepp, J. 1993. Neurochemical control of moods and emotions: From amino acids to neuropeptides. In *Handbook of emotions*, eds. M. Levis and M. J. Haviland, 87-107. New York: The Guilford Press.

Pfeifer, R. 1988. Artificial intelligence models of emotion. In *Cognitive perspectives on emotion and motivation*, eds. V. Hamilton, G. H. Bower, and N. H. Frijda, 287-320. Dordrecht: Kluwer Academic Publishers.

Pfeifer, R. 1994. The 'fungus eater approach' to emotion: A view from artificial intelligence. *Cognitive Studies, The Japanese Society for Cognitive Scienc*, 1:42-57.

Phelps, R. I., and Musgrove, P. B. 1986. Artificial intelligence approaches in statistics. In *Artificial intelligence and statistics*, ed. W. A. Gale. Reading, Massachusetts: Addison-Wesley.

Picard, R. W. 1997. *Affective computing*. Cambridge: The MIT Press.

Scherer, K. R. 1988. Criteria for emotion-antecedent appraisal: A review. In *Cognitive perspectives on emotion and motivation*, eds. V. Hamilton, G. H. Bower, and N. H. Frijda, 89-126. Dordrecht: Kluwer Academic Publishers.

Scherer, K. R. 1993. Studying the emotion-antecede[illegible] appraisal process: An expert system approach. *Cogniti[illegible] and Emotion* 7:325-355.

Stemmler, G. 1989. The autonomic differentiation of emotions revisited: Convergent and discriminant validation. *Psychophysiology* 26:617-632.

Toda, M. 1962. Design of a Fungus-Eater. *Behavioral Science* 7:164-183. (Reprinted in Toda 1982, 100-129).

Toda, M. 1982. *Man, robot and society*. The Hague: Martinus Nijhoff Publishing.

Wehrle, T. 1994a. New fungus eater experiments. In *From perception to action*, eds. P. Gaussier and J.-D. Nicoud. Los Alamitos: IEEE Computer Society Press.

Wehrle, T. 1994b. *Eine Methode zur psychologischen Modellierung und Simulation von Autonomen Agenten*. Ph.D. diss, University of Zürich.

Wehrle, T., and Scherer, K. 1995. Potential pitfalls in computational modeling of appraisal processes: A reply to Chwelos and Oatley. *Cognition and Emotion* 9:599-616.

Wehrle, T., Kaiser, S., Schmidt, S., and Scherer, K. Submitted. Studying dynamic models of facial expression of emotion using synthetic animated faces. *Journal of Personality and Social Psychology*.

WWW links:

[1] Author's home page:
http://www.unige.ch/fapse/emotion/members/wehrle/wehrle.htm

[2] Geneva Emotion Research Group:
http://www.unige.ch/fapse/emotion/members

[3] The *Emotion Home Page* maintained by Jean-Marc Fellous and Eva Hudlicka
at the Salk Institute, Computational Neurobiology Lab:
http://emotion.salk.edu/emotion.html

[4] Affective Computing MIT Media Laboratory:
http:www-white.media.mit.edu/vismod/demos/affect/affect.html

[5] The Berkeley Psychophysiology Laboratory at the University of California:
http://socrates.berkeley.edu/~ucbp1/bpl.html

achines may seem to be the stuff of fantasy. But mechanisms for understanding and
notions may be essential if we are to improve the way that computers and humans
comprehensive collection of essays presents the state of the art on this fascinating
ging research topic. I recommend it to anyone who wants to understand how
puters will eventually understand what it feels like to have a bad day at the office."
ooldridge, Department of Computer Science, University of Liverpool

of explaining the functional basis of feeling is undoubtedly the hardest (some think
oblem of cognitive science. Whether they succeed or fail, it is undeniable that the
o this volume are facing this problem head-on."
rnad, Research Chair of Canada, and Center for Cognitive Neurosciences, University

he MIT Press
lassachusetts Institute of Technology
ambridge, Massachusetts 02142
ttp://mitpress.mit.edu

0-262-20142-9

9 780262 201421 90000

Emotions in Humans and Artifacts

Trappl, Petta, and Payr, editors

Emotions in Humans and Artifacts

edited by

Robert Trappl,

Paolo Petta,

and Sabine Payr

Mandler, G. (1984): *Mind and Body: Psychology of Emotion and Stress*. W. W. Norton, New York.
Maturana, H., and Varela, F. (1980): *Autopoiesis and Cognition*. D. Reidel, Dordrecht, Holland.
Maturana, H., and Varela, F. (1987): *The Tree of Knowledge*. Shambhala, Boston.
McGinn, C. (1999): *The Mysterious Flame: Conscious Minds in a Material World*. Basic Books, New York.
Miller, R. (1981): *Meaning and Purpose in the Intact Brain*. Oxford University Press, Oxford, London, New York.
Minsky, M. (1985): *The Society of Mind*. Simon and Schuster, New York.
Moffat, D. (1999): Personal communication.
Norman, D. (1993): *Things That Make Us Smart: Defending Human Attributes in the Age of the Machine*. Addison Wesley, New York.
O'Brien, M. (1992): Playing in the MUD. Ask Mr. Protocol Column. *SUN Expert* 3 (5): 19–20; 23: 25–27.
Ogden, C. K., and Richards, I. A. (1923): *The Meaning of Meaning*. Harvest/HBJ Book, New York.
Pert, C., Candace, B., Ruff, M. R., Weber, R. J., and Herkenham, M. (1985): Neuro-peptides and Their Receptors: A Psychosomatic Network. *J. Immunol.* 135 (2): 820–826.
Picard, R. W. (1997): *Affective Computing*. MIT Press, Cambridge.
Polichar, V. E. (1996): An Office MUD for Fun and Profit? Or Maybe Just Better Communication; *login Magazine*.
Polichar, V. E. (1997): On the value of MUDs as instructional and research tools. Open letter provided to Northern Arizona University.
Riner, R., and Clodius, J. (1995): Simulating Future Histories. *Anthropol. Educ. Q.* 26 (1): 95–104. On-line. Available: ⟨http://www.dragonmud.org/people/jen/solsys.html⟩. (Availability last checked 5 Nov 2002)
Rolls, E. T. (1999): *The Brain and Emotion*. Oxford University Press, Oxford, London, New York.
Sacks, O. (1995): *An Anthropologist on Mars*. Alfred A. Knopf, New York.
Sagan, C. (1977): *The Dragons of Eden: Speculations on the Evolution of Human Intelligence*. Random House, New York.
Schachter, S., and Singer, J. E. (1962): Cognitive, Social, and Physiological Determinants of Emotional State. *Psychol. Rev.* 69: 379–399.
Schwartz, J. (1994): A Terminal Obsession. *Washington Post*, Style Section, 27 March.
Searle, J. (1999): No Limits Hinder UC Thinker. Article by Tersy McDermott in *Los Angeles Times*, 28 December, p. A23.
Sloman, A. (1997): Synthetic Minds. In W. Lewis Johnson ed., *Proceedings of the First International Conference on Autonomous Agents*, ACM SIGART, Marina Del Rey, Calif., 534–535. Association for Computing Machinery, New York.
Smith, B. C. (1986): Varieties of Self-Reference. In J. Y. Halpern, ed., *Theoretical Aspects of Reasoning about Knowledge, Proceedings of TARK 1986*, 19–43. AAAI Publication, Morgan Kaufman Publishers, Los Altos, Calif.
Smith, B. C. (1996): *On the Origins of Objects*. MIT Press, Cambridge.
Thomas, L. (1974): *The Lives of a Cell: Notes of a Biology Watcher*. Bantam Books, New York.
Turkle, S. (1995): *Life on the Screen*. Simon and Schuster, New York.
Varela, F., Thompson, E., and Rosch, E. (1991): *The Embodied Mind*. MIT Press, Cambridge.
von Uexküll, J. [1934] (1957): A Stroll through the World of Animals and Men. In C. Schiller, trans., *Instinctive Behavior: The Develoment of a Modern Concept*. International Universities Press, New York.
Zukav, G. (1999): *The Seat of the Soul*. Simon and Schuster, New York.

6 On Making Believable Emotional Agents Believable

Andrew Ortony

Abstract

How do we make an emotional agent a believable *emotional agent? Part of the answer is that we have to be able to design agents whose behaviors and motivational states have some consistency. This necessitates ensuring situationally and individually appropriate internal responses (emotions), ensuring situationally and individually appropriate external responses (behaviors and behavioral inclinations), and arranging for sensible coordination between internal and external responses. Situationally appropriate responses depend on implementing a robust model of emotion elicitation and emotion-to-response relations. Individual appropriateness requires a theory of personality viewed as a generative engine that provides coherence, consistency, and thus some measure of predictability.*

6.1 Making Believable Emotional Agents Believable

What does it take to make an emotional agent a *believable* emotional agent? If we take a broad view of believability—one that takes us beyond trying to induce an illusion of life through what Stern (chapter 12 of this volume) refers to as the "Eliza effect," to the idea of generating behavior that is genuinely plausible—then we have to do more than just arrange for the coordination of, for example, language and action. Rather, and certainly in the context of *emotional* agents, the behaviors to be generated—and the motivational states that subserve them—have to have some *consistency*, for consistency across similar situations is one of the most salient aspects of human behavior. If my mother responds with terror on seeing a mouse in her bedroom today, I generally expect her to respond with terror tomorrow. Unless there is some consistency in an agent's emotional reactions and motivational states, as well as in the observable behaviors associated with such reactions and states, much of what the agent does will not make sense. To be sure, people do not always react in the same way in the same

kind of situation—there must be variability within consistency, but equally surely there is *some* consistency—enough in fact for it to be meaningful to speak of people behaving in character. An agent whose behaviors were so arbitrary that they made no sense would probably strike us as psychotic, and Parry (e.g., Colby 1981) notwithstanding, building psychotics is not generally what we have in mind when we think about building believable emotional agents or modeling human ones.

But consistency is not sufficient for an agent to be believable. An agent's behavior also has to be *coherent*. In other words, believability entails not only that emotions, motivations, and actions fit together in a meaningful and intelligible way at the local (moment-to-moment) level, but also that they cohere at a more global level—across different kinds of situations, and over quite long time periods. For example, I know that my daughter intensely dislikes meat—it disgusts her to even think about eating something that once had a face. Knowing this, I know that she would experience disgust if she were to suddenly learn that she was eating something that contained meat (e.g., beef bouillon, not vegetable bouillon), and I would expect her disgust to influence her behavior—she would grimace, and push the plate away, and make some hideous noise. In other words, I expect her emotion-related behaviors to be consonant with (i.e., appropriate for) her emotions. But I also expect coherence with other realms of her life. Accordingly, I would be amazed if she told me that just for the fun of it, she had taken a summer job in a butcher's shop (unless perhaps I learned that she had taken the job with a view to desensitizing herself). Clearly, the issue of coherence is an important part of the solution to the problem of how to construct believable emotional agents.

6.2 Consistency and Variability in Emotions

It is an interesting fact about humans that they are often able to predict with reasonable accuracy how other individuals will respond to and behave in certain kinds of situations. These predictions are rarely perfect, partly because when we make them, we generally have imperfect information, and partly because the people whose behavior and responses we are predicting do not always respond in the same way in similar situations. Nevertheless, it is certainly possible to predict to some degree what other people

(especially those whom we know well) will do and how they will feel and respond (or be inclined to respond) under varying circumstances. We also know that certain kinds of people tend to respond in similar ways. In other words, to some extent, there is both within-individual consistency and cross-individual consistency.

So what makes it possible to predict and understand with any accuracy at all other people's feelings, inclinations, and behavior? At least part of the answer lies in the fact that their emotions and corresponding behavioral inclinations are not randomly related to the situations in which they find themselves, for if they were, we'd be unable to predict anything. But if the emotions, motivations, and behaviors of people are not randomly associated with the situations whence they arise, there must be some psychological constraints that limit the responses that are produced. And indeed, there are. Sometimes the constraints are very limiting (as with reflexes, such as the startle response) and sometimes they are less so—merely circumscribing a set of possibilities, with other factors, both personal and contextual, contributing to the response selection. But either way, there are constraints on the internal responses to situations—that is, on the internal affective states and conditions that arise in people—and on the external actions that are associated with those states and conditions.

Discussion

Sloman: You used the word "behavior" several times, and I suspect you are talking about intentions rather than behavior.

Ortony: Yes, that's why I called it motivational-behavioral component.

Sloman: But it's absolutely a crucial thing, for example, with regard to your daughter. She might well be going to work at a butcher's for the same kind of reason as somebody who belongs to a police group might join a terrorists' organization. It's the intention that is important, and the behavior might just be an appropriate means.

Ortony: Right. And actually I meant to mention this in the imaginary context of my daughter going to work at the butcher's, because one thing we would try to do to maintain our belief that people's behavior is coherent is that we would come up with an explanation, such as: "she is trying to desensitize herself." We

would not feel comfortable letting these two parts of behavior coexist—we would think that she was crazy or something.

There are two classes of theories in psychology that are relevant to these issues. Theories of emotion, and theories of personality. Consider first, emotion theories—especially cognitive ones, which are often incorporated into affective artifacts. The principal agenda of cognitive theories of emotion is the characterization of the relation between people's construals of the situations in which they find themselves and the kinds of emotions that result. The specification of such relationships is a specification of the constraints that construals of the world impose on emotional states. And these constraints are a major source of consistency, both within and across individuals. At the same time, they are only constraints—they do not come close to fully determining what a particular individual will feel or do on a particular occasion because they work in concert with several sources of variation. These are (1) individual differences in the mappings from world situations to construals (e.g., members of the winning and losing teams in a football game have different mappings from the same objective event), (2) individual differences in something that we might call emotionality (e.g., some of the team members might be more prone to respond emotionally to good or bad outcomes than others), and (3) the current state of the individual at the time (e.g., current concerns, goals, mood).

Mappings from particular types of emotions to classes of behavioral inclinations and behaviors are similarly constrained, and thus constitute another source of consistency. This is an area that only a few psychologists (e.g., Averill 1982, on anger) have studied in any very deep way, except with respect to facial expressions (e.g., Ekman 1982), although it was of considerable interest to Darwin who first wrote about it at length in his 1872 (first edition) book, *The Expression of Emotions in Man and Animals*. However, probably because the linkage between emotions and behaviors is often very flexible, there has been little effort to develop systematic accounts of it. But again, we know that the relation cannot be random, and this means that it ought to be possible to identify some principles governing constraints on the relation between what we feel and what we do, or are inclined to do. And again, whereas there are some constraining principles governing the emotion-behavior connection—principles that are the source of some consistency—there are also various factors (e.g., emotionality, again) that give rise to variation.

People only get into emotional states when they *care* about something (Ortony, Clore, and Foss 1987)—when they view something as somehow good or bad. If there's no caring, there's no emoting. This suggests that the way to characterize emotions is in terms of the different ways there might be for feeling good or bad about things. Furthermore, many traits can be regarded as chronic propensities to get into corresponding emotional states. For example, an anxious person is one who experiences fear emotions more easily (and therefore more frequently) than most people, and an affectionate person is one who is likely to experience (and demonstrate) affection more readily than less affectionate people. This means that if we have a way of representing and creating internal states that correspond to emotions, we can capture many traits too. This is important because, at the level of individuals—and this is one of my main points—traits are a major source of emotional and behavioral consistency.

Many psychologists (e.g., Ortony, Clore, and Collins 1988; Roseman, Antoniou, and Jose 1996; Scherer 1997) have proposed schemes for representing the conditions under which emotions are elicited. In our own work (which in affective computing circles is often referred to as the OCC model), we proposed a scheme that we thought accommodated a wide range of emotions within the framework of twenty-two distinct emotion types. Over the years, Gerald Clore and I, together with some of our students, collected considerable empirical support for many of the basic ideas. However, for the purposes of building believable artifacts, I think we might want to consolidate some of our categories of emotions. So, instead of the rather cumbersome (and to some degree arbitrary) analysis we proposed in 1988, I think it is worth considering collapsing some of the original categories down to five distinct positive and five negative specializations of two basic types of affective reactions—positive and negative ones—as shown in table 6.1.

I think that these categories have enough generative capacity to endow any affective agent with the potential for a rich and varied emotional life. As the information processing capabilities of the agent become richer, more elaborate ways of characterizing the good and the bad become possible, so that one can imagine a system starting with only the competence to differentiate positive from negative and then developing progressively more elaborate

Table 6.1 Five specializations of generalized good and bad feelings (collapsed from Ortony, Clore, and Collins 1988)

Positive reactions
because something good happened (joy, happiness etc.)
about the possibility of something good happening (hope)
because a feared bad thing didn't happen (relief)
about a self-initiated praiseworthy act (pride, gratification)
about an other-initiated praiseworthy act (gratitude, admiration)
because one finds someone/thing appealing or attractive (love, like, etc.)
Negative reactions
because something bad happened (distress, sadness, etc.)
about the possibility of something bad happening (fear, etc.)
because a hoped-for good thing didn't happen (disappointment)
about a self-initiated blameworthy act (remorse, self-anger, shame, etc.)
about an other-initiated blameworthy act (anger, reproach, etc.)
because one finds someone/thing unappealing or unattractive (hate, dislike, etc.)

The first entry in each group of six is the undifferentiated (positive or negative) reaction. The remaining five entries are specializations (the first pair goal-based, the second standards-based, and the last taste-based).

categories. A simple example of this idea is that fear can be viewed as a special case of a negative feeling about something bad happening—with the bad thing being the *prospect* of something bad happening. If one adopts this position, then one is left with the idea that the main driving force underlying all emotions is the registration of good and bad and that discrete emotions can arise to the extent that the nature of what is good and bad for the agent can be and is elaborated. Indeed, this may well be how humans develop increasingly sophisticated emotion systems as they move from infancy through childhood to adulthood.

So, specifying a mechanism that generates distinct emotions and other affective conditions seems not so hard—what is hard is to make it all believable. As I just indicated, a key issue is the need for affective artifacts to be able to parse the environment so as to understand its beneficial and harmful affordances—a crucial requirement for consistency, and thus also for believability. And a prerequisite for doing this is a coherent and relatively stable value system in terms of which the environment is appraised. As we indicated in OCC (and as illustrated in figure 6.1), such a system, at least in humans, is an amalgam of a goal hierarchy in which at least some of the higher-level goals are sufficiently enduring that

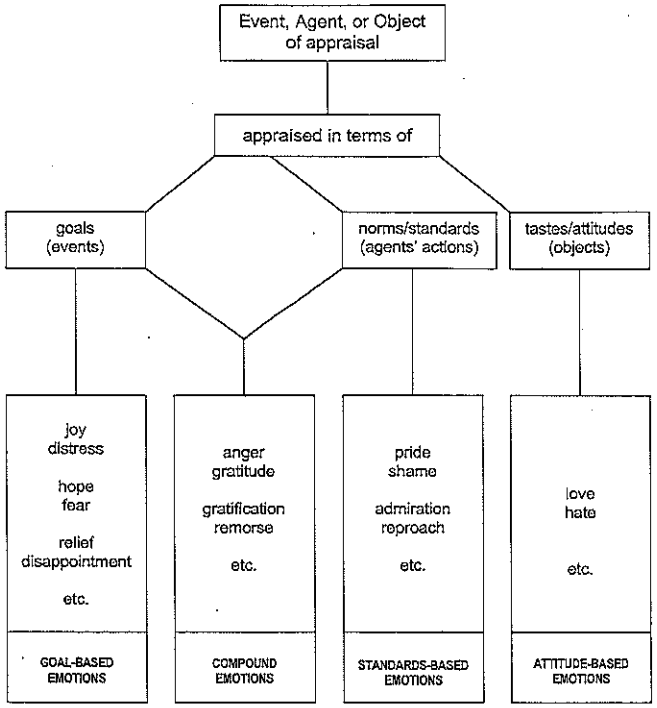

Figure 6.1 The relation between different things being appraised, the representations in terms of which they are appraised, and the different classes of resulting emotions.

they influence behavior and emotions over an extended period (rather than transiently), a set of norms, standards, and values that underlie judgments of appropriateness, fairness, morality, and so on, and tastes and preferences whence especially value-laden sensory stimuli acquire their value.

Another respect in which emotional reactions and their concomitant behaviors need some degree of consistency has to do with emotion intensity. It is not sufficient that similar situations tend to elicit similar emotions within an individual. Similar situations also elicit emotions of comparable intensity. In general, other things (external circumstances, and internal conditions such as moods, current concerns, etc.) being equal, the emotions that individuals experience in response to similar situations, and the intensity with

which they experience them, are reasonably consistent. Emotionally volatile people explode with the slightest provocation while their placid counterparts remain unmoved. In this connection, I'm reminded of a colleague (call him G) whom my (other) colleagues and I know to be unusually "laid back" and unemotional. One day several of us were having lunch together in an Italian restaurant when G managed to splash a large amount of tomato sauce all over his brilliant white, freshly laundered shirt. Many people would have become very angry at such an incident—I for example, would no doubt have sworn profusely, and for a long time! G, on the other hand, said nothing; he revealed no emotion at all—not even as much as a mild kind of "oh dear, what a bother" reaction; he just quietly dipped his napkin into his water and started trying to wipe the brilliant red mess off his shirt (in fact making it worse with every wipe), while carrying on the conversation as though nothing had happened. Yet, unusual as his nonreaction might have been for people in general, those of us who witnessed this were not at all surprised by G's reaction (although we were thoroughly amused) because we all know G to be a person who, when he emotes at all, consistently does so with very low intensity—that's just the kind of person he is, that's his personality.

6.3 Consistency and Variability in Emotion-Related Response Tendencies

The tomato sauce episode not only highlights questions about emotion intensity, it also, for the same reason, brings to the fore the question of the relation between (internal) emotional states and their related behaviors. To design a computational artifact that exhibits a broad range of believable emotional behavior, we have to be able to identify the general principles governing the relation between emotions and behavior, or, more accurately, behavioral inclinations, because, as Ekman (e.g., 1982) has argued so persuasively, at least in humans, social and cultural norms (display rules) often interfere with the "natural" expression (both in the face, and in behavior) of emotions.

Associated with each emotion type is a wide variety of reactions, behaviors, and behavioral inclinations, which, for simplicity of exposition, I shall refer to collectively as "response tendencies" (as distinct from responses). Response tendencies range from involuntary expressive manifestations, many (e.g., flushing) having immediate physiological causes, through changes in the way in which

information is attended to and processed, to coping responses such as goal-oriented, planned actions (e.g., taking revenge). From this characterization alone, it is evident that one of the most salient aspects of emotional behavior is that some of it sometimes is voluntary and purposeful (goal-oriented, planned, and intentional) and some of it is sometimes involuntary and spontaneous—as when a person flies into an uncontrollable rage, trembles with fear, blushes with embarrassment, or cries with joy.

Figure 6.2 sketches a general way of thinking about the constraints on the response tendencies for emotions. It shows three major types of emotion response tendencies (labeled "expressive," "information-processing," and "coping"), each of which is elaborated below its corresponding box. The claim is that *all* emotion responses have these three kinds of tendencies associated with them. Note, however, that this is *not* the same as saying that in every case of every emotion, these tendencies have observable concomitants—they are *tendencies* to behave in certain ways, not actual behaviors. The first group—the *expressive* tendencies—are the usually spontaneous, involuntary manifestations of emotions that are often referred to by emotion theorists (following Darwin) as emotional expressions. These expressive tendencies are of three kinds: *somatic* (i.e., bodily), *behavioral*, and *communicative* (both verbal and nonverbal). Consider first the somatic tendencies. These are almost completely beyond the control of the person experiencing the emotion. For instance, the box marked "somatic" in figure 6.2 has a parenthetical "flushing" in it. This (and the other parenthetical entries) is presented (only) as an example of the kind of response tendencies that one might expect to find in the case of anger; it should be interpreted as indicating that when someone is angry, one possible somatic manifestation is that the person grows red in the face. Notice that this is not something that he or she chooses to do. We do not choose to flush—our physiology does it for us, without us asking.

The next class of expressive tendencies are the behavioral ones. Again, these tendencies are fairly automatic, often hardwired, and relatively difficult (although not always impossible) to control; they are spontaneous actions that are rarely truly instrumental (although they might have vestigial instrumentality), such as kicking something in anger. So, to continue with the example of anger, I have in mind not the reasoned planful behaviors that might be entertained as part of a revenge strategy (they belong to the

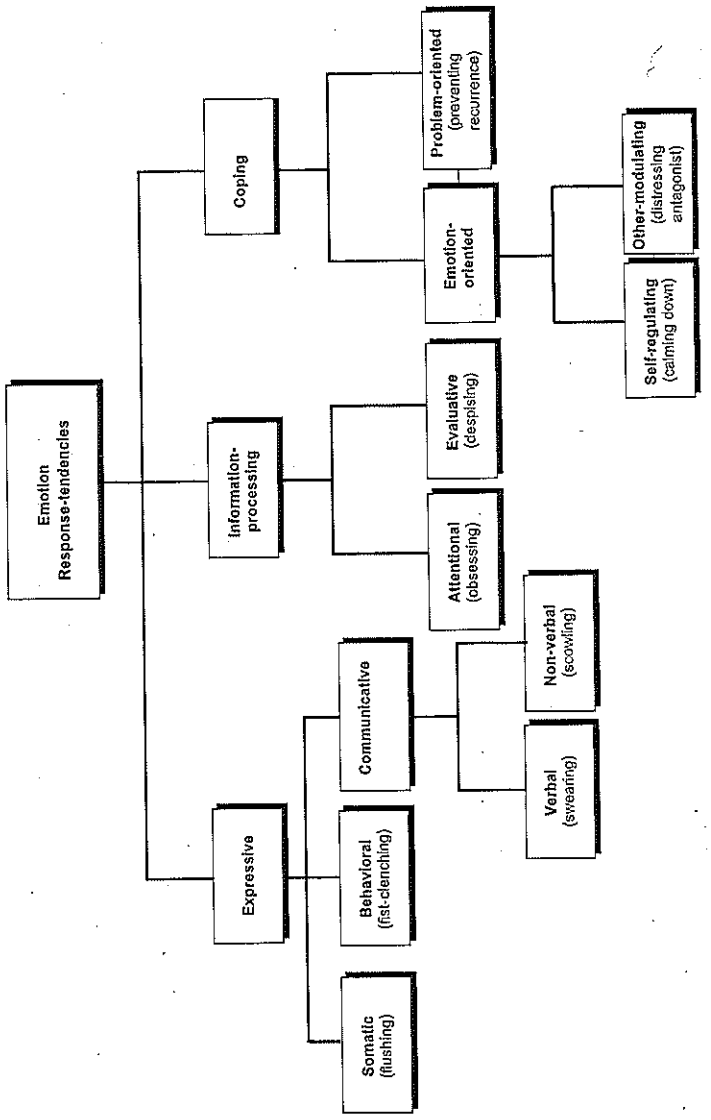

Figure 6.2 Proposed analysis of the behavioral structure of emotions. The parenthetical entries in the leaf nodes are intended as examples of the different kinds of response tendencies, in this case instances of response tendencies that might be associated with anger are indicated.

"coping" category), but the more spontaneous tendencies to exaggerate actions (as when one slams a door that one might have otherwise closed quietly), or the tendency to perform almost symbolic gestural actions (albeit, often culturally learned ones) such as clenching one's fist.

Finally, I have separated out communicative tendencies (while realizing that symbolic acts such as fist clenching also have communicative value) as a third kind of expressive response tendency. Still, I wish here to focus more on communication through the face, because historically this has been so central to emotion research. Communicative response tendencies are those that have the capacity to communicate information to others, even though they are often not intended to do so. They have communicative value because they are (sometimes pan-culturally) recognized as *symptoms* of emotions. They include *nonverbal* manifestations in the face, including those usually referred to by emotion theorists as "facial expressions" (e.g., scowling, furrowing of the brow), as well as *verbal* manifestations (e.g., swearing, unleashing torrents of invectives), and other kinds of oral (but nonverbal) responses such as growling, screaming, and laughing.

The second, *information processing*, component has to do with changes in the way in which information is processed. A major aspect of this is the diversion of *attention* (again often quite involuntary) from those tasks that were commanding resources prior to the emotion-inducing event to issues related to the emotion-inducing event. One of the most striking cases of the diversion of attentional resources is the all-consuming obsessive focus that people often devote to situations that are powerfully emotional. In humans, this obsessive rumination can be truly extraordinary and often quite debilitating, as so convincingly depicted in much of the world's great literature—consider, for example, Shakespeare's *Othello*. The second part of the information processing response has to do with updating beliefs, attitudes, and more generally *evaluations* about other agents and objects pertinent to the emotion-inducing event—you increasingly dislike your car when it repeatedly infuriates you by breaking down on the highway, whereas your liking for an individual increases as he or she repeatedly generates positive affect in you (Ortony 1991).

Finally, there are coping strategies, of which I have identified two kinds. One of these, *problem-oriented coping*, is what emotion theorists usually have in mind when they talk about coping;

namely, efforts to bring the situation under control—to change or perpetuate it—with the goal of improving a bad situation, or prolonging or taking advantage of a good one. In the case of anger, people often seek to do something that they think might prevent a recurrence of the problem, or that might somehow fix the problem.

The more interesting kind of coping is *emotion-oriented coping.* This kind of coping has to do with managing emotions themselves—either one's own, or those of some other agent or agents involved in the emotion-inducing situation. *Self-regulating* emotion-oriented coping responses focus on one's own emotions. For example, if I am angry I might try to calm down, perhaps as a precursor to developing a sensible plan to solve the problem, or perhaps simply because I don't like the feeling of being out of control. The *other-modulating* emotion management strategies can serve various purposes. For instance, if I induce distress in you because of what you did to me, not only might it make me feel better (i.e., it might help me to manage my own emotion of anger, hence the association in the figure between self-regulating and other-modulating responses), but it might also make you more likely to fix the problem you caused me (hence the link between emotion-oriented and problem-oriented responses). So, for example, suppose you are angry at somebody for smashing into your car. Developing or executing a plan to have the car fixed is a problem-oriented response, as would be a desire to prevent, block, or otherwise interfere with the antagonist's prospects for doing the same kind of thing again. But one might also try to modulate the antagonist's emotions by retaliating and getting one's own back so as to "make him pay for what he did to me," or one might try to induce fear or shame in him to make him feel bad, all with a view to making one's self feel better. There is no requirement that any of these responses be "rational." Indeed, if we designed only rational emotion response-tendencies into our emotional agents, we would almost certainly fail to make our emotional agents believable.

So the general claim is that a major source of consistency derives from the fact that all emotions constrain their associated response tendencies and all emotions have all or most of these tendencies. It should be clear from this discussion, and from table 6.2 (which indicates how the various constraints might be manifested in the emotions of anger and fear) that there is plenty of room for individual variation. Just as in the case of the emotions themselves, much of this variation is captured by traits—so many, although

Table 6.2 Sample manifestation of the different components for fear emotions (upper panel) and for anger emotions (lower panel)

Expressive	Somatic	Trembling, shivering, turning-pale, piloerection
	Behavioral	Freezing, cowering
	Communicative nonverbal	Screaming
Information Processing	Attentional	Obsessing about event, etc.
	Evaluative	Disliking source, viewing self as powerless/victim
Coping	Emotion Self-regulating	Calming down, getting a grip
	Emotion Other-modulating	Scaring away
	Problem-oriented coping	Getting help/protection, escaping, eliminating threat

Expressive	Somatic	Shaking, flushing
	Behavioral	Fist-clenching
	Communicative verbal	Swearing
	Communicative nonverbal	Scowling, frowning, stomping, fist-pounding, etc.
Information Processing	Attentional	Obsessing about event, etc.
	Evaluative	Disliking and despising source
Coping	Emotion Self-regulating	Calming down, getting a grip
	Emotion Other-modulating	Causing distress to antagonist
	Problem-oriented coping	Preventing continuation or recurrence of problem

not all, of the ways in which a timid person responds to anger-inducing situations are predictably different from the ways in which an aggressive person responds.

6.4 Why Personality?

Traits are the stuff of personality theory. Personality psychologists disagree as to whether personality should be viewed merely as an empirical description of observed regularities, or whether it should be viewed as a driver of behavior. But for people interested in building affective artifacts, personality can only be interesting and relevant if one adopts the second position. If one really wants to build believable emotional agents, one is going to need to ensure situationally and individually appropriate internal responses (emotions), ensure situationally and individually appropriate external responses (behaviors and behavioral inclinations), and arrange for sensible coordination between internal and external responses. Situationally appropriate responses are controlled by incorporating models of emotion elicitation and of emotion to emotion-responses of the kind I have just outlined. But to arrange

for individual appropriateness, we will have to incorporate personality, not to be cute, but as a generative engine that contributes to coherence, consistency, and predictability in emotional reactions and responses. The question is, how can we incorporate personality into an artifact without doing it trait by trait for the thousands of traits that make up a personality? In their famous 1938 monograph, Trait Names: A Psycho-lexical Study, Allport and Odbert identified some 18,000 English words as trait descriptors, and even though many of the terms they identified do not in fact refer to traits, the number still remains very large.

Trying to construct personalities in a more or less piecemeal fashion, trait by trait, is probably quite effective if the number of traits implemented is relatively small and if the system complexity is relatively limited. To some extent, this appears to be the way in which emotional behaviors and expressions are constrained in Cybercafé—part of Hayes-Roth's Virtual Theater Project at Stanford (e.g., Rousseau 1996), and to an even greater extent, in Virtual Petz and Babyz (see Stern, chapter 12 of this volume), and anyone who has interacted with these characters knows how compelling they are. However, if one has more stringent criteria for believability—as one might have, for example, in a soft-skills business training simulation, where the diversity and complexity of trait and trait constellations might have to be much greater—I suspect that a more principled mechanism is going to be necessary to produce consistent and coherent (i.e., believable) characters. Note, incidentally, that this implies that "believability" is a context-, or rather application-dependent notion. A character that is believable in an entertainment application might not be believable in an education or training application.

One solution to the problem of how to achieve this higher level of believability is to exploit the fact that traits don't live in isolation. If we know that someone is friendly we know that he has a general tendency or disposition to be friendly relative to people in general; we know that in a situation that might lead him to be somewhere on the unfriendly-friendly continuum, he is more likely to be toward the friendly end. However, we also know some other very important things—specifically, we know that he is likely to be kind, generous, outgoing, warm, sincere, helpful, and so on. In other words, we expect such a person to exhibit a number of correlated traits. This brings us back to the question of behavioral coherence. There is much empirical evidence that traits cluster together and that trait space can be characterized in terms of a small number of factors—varying in number from two to five, depending on how one decides to group them. For our purposes here, the question of which version of the factor structure of personality one adopts may not be crucial (although I do have a personal preference). What matters is that the factor structure of trait space provides a meaningful way to organize traits. What matters is that it provides a meaningful and powerful reduction of data to note that people whom we would normally describe as being outgoing or *extroverted* (as opposed to introverted) tend to be sociable, warm, and talkative, and that people who are forgiving, good-natured, and softhearted we generally think of as *agreeable* or likeable (as opposed to antagonistic). Similarly, people who are careful, well organized, hard working, and reliable we tend to think of as being *conscientious* (as opposed to negligent). These (extroversion, agreeableness, and conscientiousness) are three of the "big five" (e.g., McCrae and Costa 1987) dimensions of personality—the other two being *openness* (as opposed to closed to new experiences), and *neuroticism* (as opposed to emotional stability).

The key point here is that such clusters, such groups of tendencies to behave and feel in certain kinds of ways, constitute one source of behavioral and emotional consistency and hence predictability of individuals. Viewed in this way, personality is the engine of behavior. You tend to react this way in situations of this kind *because* you are that kind of person. Personality is a (partial) *determiner* of, not merely a summary statement of, behavior. Consistent with this view (which is certainly not shared by all personality theorists) is the fact that some components of personality appear to be genetically based. All this suggests that to build truly believable emotional agents, we need to endow them with personalities that serve as engines of consistency and coherence rather than simply pulling small groups of traits out of the thin air of intuition.

A general approach to doing this would be to identify generative mechanisms that might have the power to spawn a variety of particular states and behaviors simply by varying a few parameters. Many of the proposals in the personality literature provide a basis for this kind of approach. For example, one might start with the distinction between two kinds of regulatory focus (e.g., Higgins 1998), namely, *promotion* focus in which agents are more

concerned with attempting to achieve things that they need and want (e.g., they strive for nurturance, or the maintenance of ideals). Promotion focus is characterized as a preference for gain-nongain situations. In contrast, with *prevention* focus, agents seek to guard against harm (e.g., they strive for security) and exhibit a preference for nonloss-loss situations. Thus regulatory focus is a fundamental variable that characterizes preferred styles of interacting with the world. Different people at different times prefer to focus on the achievement of pleasurable outcomes (promotion focus), or on the avoidance of painful outcomes (prevention focus). These are essentially the same constructs as approach motivation and avoidance motivation (e.g., Revelle, Anderson, and Humphreys 1987), and are closely related to the idea that individuals differ in their sensitivity to cues for reward and punishment (Gray 1994). This can be clearly seen when we consider people's gambling or sexual behavior (sometimes there's not much difference): Those who are predominantly promotion focused (sensitive to cues for reward) focus on the possible gains rather the possible losses—they tend to be high (as opposed to low) on the personality dimension of impulsivity; those with a prevention focus (sensitive to cues for punishment) prefer not to gamble so as to avoid the possible losses—these people tend to be high as opposed to low on the anxiety dimension.

If an individual prefers one regulatory strategy over another, this will be evident in his behaviors, in his styles of interaction with the world, and with other agents in the world, and as such, it constitutes one aspect of personality. Probably the most productive way to think about regulatory focus is that in many of our encounters with the world, a little of each is present—the question then becomes, which one dominates, and to what degree. Different people will have different degrees of each, leading to different styles of interacting with the world. Still, some of each is what we would ordinarily strive for in designing an affective artifact. Without some counterbalancing force, each is dysfunctional. For example, unbridled promotion focus is associated with a high tolerance for negativity (including a high threshold for fear, pain, and the like), and that comes pretty close to being pathologically reckless behavior.

I think it is possible to exploit these kinds of ideas in a principled way in designing our artifacts. We might start with the ideas of psychologists such as Eysenck, Gray, Revelle, and others (e.g., Rolls 1999; chapter 2 of this volume) who take the position that

there are biological substrates of personality (such as cue sensitivity). The virtue of this kind of approach is that it provides a biological basis for patterns of behaviors and, correspondingly, emotions, which can serve as the basis for generating some sort of systematicity and hence plausibility or believability of an artificial agent. Which particular activities a human agent actually pursues in the real world is of course also dependent on the particular situation and local concerns of that agent, as well, no doubt, as on other biologically based determinants of other components of personality. But at least we have a scientifically plausible and computationally tractable place to start, even though the specification of exactly how this can be done remains a topic for future research.

Discussion: On Modeling Emotion and Personality

Bellman: So you are telling me that personality would be this core biological basis that somehow constrains behavioral inclinations. I have been bothered for a long time about a lot of research in emotions, because I am confused by the tendency to want to have oversimplified models—why people keep trying to reduce the space to two or three bases. There are many disciplines that I have been in which try to take complex multidimensional problems and reduce them to two or three bases. And, yes, you can do that at some level, but you usually lose most of the interesting stuff when you do that.

Sloman: I have a worse problem: Why try to find any number of dimensions as opposed to finding what the underlying architecture is and generating these things?

Bellman: But the underlying architecture doesn't have to be something with only two or three reinforcement/nonreinforcement bases. Why should that be the underlying architecture?

Rolls: The only theory is that one tries to get a principled approach here instead of doing something like factor analysis. The idea is to say something like this: So, what is it that causes emotion? If one would recreate and operationalize things where reinforcers are involved—most people find it difficult to think of exceptions to that—then one ought to pursue that idea and ask: What sorts of reinforcers are there? You know, you can give positive and negative reinforcement, you can withdraw positive and negative reinforcement. The second question is: What comes out of that? The

nice thing that Andrew is pointing us towards here is that personality might drop out of that research. If some individuals, by their genes, were a bit less sensitive to nonreward, or a bit less sensitive to punishment, it turns out that you would categorize them as extraverts. And so, one gets the dimension of personality without having to buy into any sort of special engineering specifications.

Ortony: But personality has consequences, because it constrains behaviors down the road, individual behaviors and motivations, and it makes one more likely to gamble in casinos and more likely to engage in unsafe sex and more likely to do a whole host of things which actually make people look as though they are individuals with some sort of stable underlying behavioral patterns.

Rolls: The idea is that here, then, is a sort of scientifically principled way to get personality. And I agree with your notion about consistency, but it is one of the quite nice things that come with personality. Notice that consistency is slightly different to persistence of emotions: If you have a nonreward, it's helpful on the short-time scale to keep your behavior directed towards a reward. So, the emotional state ought to be continuing slightly. Consistency, on the other hand, is more of a long-term requirement for believable agents: Next time you come round to that similar situation for that individual, it makes sense if they behave in a somewhat similar way. At least we as biological organisms do, perhaps for the reasons that we have discussed.

Bellman: The comment was not that there aren't some important principles. It is exactly how we model those important principles. I will give you a simple example: If we take language generation, we can model it, as we have learned to do over time. It has been very difficult, with all sorts and kinds of generative grammars. That's a very different kind of modeling than from a simple combinatorics. I don't know any cases, but one could imagine language as having been modeled as if it were a simple combinatoric problem. And eventually, people shifted to more interesting underlying modeling with grammars. That's part of my point: I don't see any reason why we should suppose, just because of a positive modeling, a simple combinatoric space. That's what my comment was directed towards, not the lack of principles.

Sloman: But are you talking about building synthetic artifacts for some purpose, which may be useful, or are you talking about how human beings work? Because, if it's the latter, there are going to be

constraints, and you can go and find out the biological bases, you can go and find out how the brain is involved.

Bellman: But there are lots of constraints that we know about in human language generation.

Sloman: So, the answer to your question is: Technically, you can take as many constraints as you get a handle on. And you do the best you can. And then someone else may come and have a better solution.

Bellman: Ok. I am suggesting a different style of modeling for it. I think that there is a long history of emotional modeling.

Petta: If you talk about human beings, the society and the environment as such are so complex that perhaps you have to leave out that part and just concentrate on the individual. You make an analysis of how the individual describes itself and end up with this collection of traits, which always just refers to the first person point of view. What I think is important however, especially when you are talking about artifacts, is that perhaps way more efforts should be put into performing a thorough lifeworld analysis—to use the terminology of Agre.[1] You would have to look at the whole system, what the environment provides and what kinds of couplings there are between a single architecture and between different instances of the architecture and/or what could happen in the environment, what kinds of dimensions, effects, and kinds of dynamics are relevant for your artifact. Personality, after all, is also something that is perceived about the other.

Sloman: Could you give an example of the kind of coupling you talked about?

Petta: Especially when you design an artifact, you want it to behave within given constraints. These constraints typically are characterized by, for instance, what kinds of interactions you want to occur, what kind of dynamics, how you want to stabilize its behavior. And once you know that, you can take a look at what happens in that environment, what can take place over a certain amount of time and how that relates to the possibilities of the agent, what its perceptional capabilities are, what its choice of actions is. Then you can start to consider how to constrain or direct those elements. Where do you put the incentives, where do you rather expand, where do you rather suppress? And these, in some, turn out to be biases, constraints, which again can have their own dynamics coupled to the environment, and which, I would

assume, should lead to something recognizable as a certain personality.

Ortony: Part of this, I think, has to do with a fact that I did not talk about, the appraisal side. Also, you want a level of description, if you are thinking about this in general, that goes beyond any particular design intention we may have. So, in thinking in general about what you want to do when you construct believable agents, you are not going to have one set of criteria for a person or a system, and a different set of criteria for what makes it believable. For a pet robot, for example, you have to have some level of description that characterizes the interaction which organisms are likely to have in a physical world. So, this gets you to things like this, but stuff happens that gets appraised. Again, there are individual differences, but there are also similarities with respect to our goals. The norms or standards that we use to make judgments of the kind that lead us to approve or disapprove of things, which is different from goals, although they could in fact be collapsed if you said that you had a goal to maintain order or something. But the point is, it does not matter what the goals of the organism are. It only matters that it indeed *has* some goals. It is very difficult to imagine building an emotionally believable artifact which didn't have goals.

So, once you admit that it has got goals, the architecture should be such that it's impervious to the particular kind of goals. It only cares what happens when goals are satisfied or blocked or failed. This really comes down to how you characterize the underlying value system in terms of which appraisals of the environment you are constructing. Is that an answer to your question or not?

Petta: Partly, because, actually, I refer to the box with norms and standards (cf. figure 6.1), because this is where you introduce aspects of the society which are beyond the individual. So, this is just one very evident spot where you gather stuff that is external to the single individual and put it into your picture.

Sloman: But if that does not get internalized by the individual, it has no effect on the individual's behavior.

Petta: Yes, sure. Obviously, there must be a connection, there must be a coupling.

Sloman: Can I, in this context, say something which, at first, will contradict with what you say? You have been saying that you need consistency because of the predictability of personality. Now, if you actually look at human beings, but not necessarily at other

animals, you will find that there may be consistency in a particular individual's behavior in a certain context only. If you put him in another context, you will get a different collection of behaviors, which is itself consistent. So, at home with your family, you may be kind and generous and thoughtful, whereas being in the aggressive MIT AI Lab or in the office, where there is a lot of competition and insecurity, you may suddenly behave in a very different way, but consistently in itself. I think that would not be possible for other animals.

Ortony: It might actually: animal as parent and animal as hunter, for example.

Sloman: Yes. In that case, it may be a very general characterization that, in some sense, there is not a single personality, there are subpersonalities which get turned on and off by the context.

Petta: By the context, by the environment. That's precisely my point.

Ortony: Yes and no, is my reaction. At some level, of course, it's true that we are going to behave less aggressively in an environment in which the interpersonal interactions are characterized by love, affection, and familial relations, than when we are in a hostile or in a competitive environment—at the workplace, for example.

Elliott: You can have one personality that appraises everything always the same way, but it's appraising different things at different times.

Sloman: But even cues for punishment or reward might be a variable factor. They may be consistently low in this situation and consistently high when you are in that situation.

Elliott: Also, you can't leave out that individuals are highly affected by moods, too, in that you can characterize being in the workplace as placing yourself in an aggressive mood.

Ortony: Let's take a dimension like friendliness—a person you characterize as a friendly person would be represented on one side of a curve for "friendliness" in a group. But, for this particular individual, it is still true that there is a distribution of behaviors relative to his own behaviors, such that some of his behaviors are friendly and some are unfriendly. But this distribution lies inside the "friendly" sector with respect to the reference group.

Sloman: And that distribution might shift with context for the same individual.

Ortony: Yes, it could shift, but it's still probably the case that a person whom you would regard as aggressive is going to have more aggressive behaviors in the aggressive roles.

Sloman: I think what you are saying is that there may be an even deeper consistency. I suspect that for some people, there is a degree of integration that is higher than for others. And in an extreme case you get personality disorders where this mechanism is going badly wrong.

Rolls: So, is the bottom line of this that your sensitivity to reward and punishment could be relatively set, but the actual coping strategies that you are bringing into play in different environments are appropriate for that particular environment? For example, if there is something you can do about it, you might be angry; but if there's nothing you can do about it, you might be sad? Is that one way to try to rescue a sort of more biological approach?

The basic biology might—successive to the reward and punishment—be unchanged, but then you have different, as it were, coping strategies.

Sloman: This is an empirical question, and I have no reason to think that, at least for humans, it's more like what I said than like what you said. But I could be wrong. We have to investigate.

Rolls: Yes, that's right. But I think it's worth underlining the fact that there are at least two possibilities to explain context dependency of personality.

Note

1. P. E. Agre, *Computation and Human Experience* (Cambridge University Press, Cambridge, 1997).

References

Allport, G. W., and Odbert, H. S. (1938): Trait Names: A Psycho-Lexical Study. *Psychol. Monogr.* 47 (1): whole no. 211.

Averill, J. R. (1982): *Anger and Aggression: An Essay on Emotion.* Springer, Berlin, Heidelberg, New York.

Colby, K. M. (1981): Modeling a Paranoid Mind. *Behav. Brain Sci.* 4: 515–560.

Darwin. C. (1872): *The Expression of Emotions in Man and Animals.* Murray, London.

Ekman, P., ed. (1982): *Emotion in the Human Face.* Cambridge University Press, Cambridge.

Gray, J. A. (1994): Neuropsychology of Anxiety. Oxford University Press, Oxford, London, New York.

Higgins, E. T. (1998): Promotion and Prevention: Regulatory Focus as a Motivational Principle. In M. P. Zanna, ed., *Advances in Experimental Social Psychology.* Academic Press, New York.

McCrae, R. R., and Costa, P. T. (1987): Validation of a Five-Factor Model of Personality across Instruments and Observers. *J. Pers. Soc. Psychol.* 52: 81–90.

Ortony, A. (1991): Value and Emotion. In W. Kessen, A. Ortony, and F. Craik, eds., *Memories, Thoughts, and Emotions: Essays in Honor of George Mandler.* Laurence Erlbaum Associates, Hillsdale, N.J.

Ortony, A., Clore, G. L., and Collins, A. (1988): *The Cognitive Structure of Emotions.* Cambridge University Press, Cambridge.

Ortony, A., Clore, G. L., and Foss, M. A. (1987): The Referential Structure of the Affective Lexicon. *Cogn. Sci.* 11: 341–364.

Revelle, W., Anderson, K. J., and Humphreys, M. S. (1987): Empirical Tests and Theoretical Extensions of Arousal Based Theories of Personality. In J. Strelau and H. J. Eysenck, eds., *Personality Dimensions and Arousal,* 17–36. Plenum Press, London.

Rolls, E. (1999): *The Brain and Emotion.* Oxford University Press, Oxford, London, New York.

Roseman, I. J., Antoniou, A. A., and Jose, P. E. (1996): Appraisal Determinants of Emotions: Constructing a More Accurate and Comprehensive Theory. *Cogn. Emotion* 10: 241–277.

Rousseau, D. (1996): Personality in Computer Characters. Paper Presented at the Annual Meeting of the American Association of Artificial Intelligence. In: H. Kitano, ed., "Entertainment and AI/A-Life", AAAI Workshop Technical Report WS-96-03, AAAI Press, Menlo Park, CA, pp. 38–43.

Scherer, K. R. (1997): Profiles of Emotion-Antecedent Appraisal: Testing Theoretical Predictions Across Cultures. *Cogn. Emotion* 11: 113–150.

1

AFFECT ASSESSMENT THROUGH SELF-REPORT METHODS

JOHN HUMRICHOUSE, MICHAEL CHMIELEWSKI, ELIZABETH A. McDADE-MONTEZ, AND DAVID WATSON

Affect emerged as a central concept in psychology during the 1980s, and affective science has flourished ever since. To document this explosion of interest, we conducted a search of the PsycINFO database using the keywords *affect*, *emotion*, *emotions*, *emotional states*, and *mood*. This search generated 7,083 hits during the 5-year period from 1980 to 1984. Since 1984, the number has increased dramatically, rising to 11,374 (1985–1989), 16,478 (1990–1994), 24,602 (1995–1999), and finally to 33,828 during the most recent 5-year period (2000–2004).

The explosion of scientific interest has created an acute need for reliable and valid measures of affect. In this chapter, we provide a brief introduction to the self-report assessment of affect. The chapter is organized into three main sections. First, we define and distinguish among affective terms and address the basic underlying structure of affect. Second, we review and evaluate existing self-report measures of affect. Third, we discuss basic considerations in affect measurement (e.g., construct validity, reliability, and sources of measurement error) and make general recommendations about assessing affect through self-report methods.

Before one can assess affect intelligently, it is important to define several constructs that are relevant to this area of investigation (i.e., affect, emotions, and moods). *Affect* is a broad overarching construct, encompassing both emotions and moods. It is what one is experiencing or feeling, either pleasant or unpleasant, with varying levels of intensity, duration, and triggers or patterns of activation (Gray & Watson, in press). Watson (2000) argued that one's waking experience invariably is spent in some affective state and referred to this continuous affective experience as a "stream of affect [that] typically is experienced as mildly to moderately pleasant" (p. 15).

Within the domain of affect, *emotions* have been defined as biobehavioral systems comprising at least four core components: (a) a subjective experience, (b) a physiological reaction, (c) an expressive component (e.g., a facial expression), and (d) a behavioral response (Watson & Vaidya, 2003). For example, the emotion of fear comprises the subjective experience of apprehension, the physiological reaction of increased heart rate and general sympathetic activation, the facial expression of raised eyebrows and wide-open eyes, and the behavioral response of either freezing or fleeing. These components occur as part of an intense, coordinated response that lasts for a very brief period of time, usually for only a matter of seconds or minutes. Affective experiences comprising these four components have been conceptualized as basic emotions. Although there is no consensual taxonomy of basic emotions, most models include anger, disgust, fear, sadness, interest, joy, and surprise (Watson, 2000).

Moods are similar to the subjective components of emotions—they represent the subjective aspects of one's experience. Moods have an evaluative quality of being either positive or negative and can vary in relative intensity and duration. However, many mood states do not reflect basic emotional responses because they do not clearly exhibit all four components mentioned earlier (Watson, 2000; Watson & Vaidya, 2003). For example, the classic emotion of anger involves dramatic manifestations of all four components; in contrast, the experience of an irritable mood need not implicate the other three components (e.g., the face may not exhibit a prototypic anger expression). Additionally, the concept of mood includes an array of low-intensity states (e.g., calm, quiet, sleepy) and mixed states (e.g., nostalgia) that are not traditionally considered to represent emotions. Therefore, moods encompass a broader range of subjective states than classically defined emotions.

Whereas emotions are brief and intense, moods can be less dramatic and can last much longer. Considering that emotions are generally of greater relative intensity than moods, there are adaptive advantages (e.g., conservation of energy and bodily resources) to experiencing these states relatively infrequently (see Watson, 2000). Indeed, experiencing prolonged, intense emotional states may be maladaptive and indicative of psychopathology (Clark & Watson, 1994). For example, panic disorder is characterized by an intense

negative emotional episode that essentially represents a prolonged (and situationally inappropriate) fear response.

On the basis of an analysis of thousands of momentary observations, Watson (2000) estimated that as little as 17% of our waking time may be spent experiencing intense affective states reflecting classical emotions. Therefore, moods comprise a much larger part of the continuous "stream of affect." Additional analyses have indicated that people show consistent and stable individual differences in their tendency to have both positive and negative mood states. These findings have led researchers to distinguish between *state affect* (i.e., current short-term fluctuations in mood) and *trait affect* (i.e., stable individual differences in the tendency to experience different types of mood; see Watson, 2000; Watson, Clark, & Tellegen, 1988).

Another distinction between emotions and moods lies in the activation or triggers of these affective states. Generally, emotions have an identifiable trigger or event that activates the coordinated response. On the other hand, moods often seem to arise without a clear trigger or reference point and later dissipate without a clear intervention or change in the environment. Thus, for instance, a person may experience a low, dysphoric mood without knowing why. One reason for this is that moods are strongly influenced by a variety of endogenous processes, such as circadian rhythms (see Watson, 2000, chap. 4; Watson, Wiese, Vaidya, & Tellegen, 1999). Likewise, psychopathology may involve strong affective responses without clear precipitants, as is seen in *generalized anxiety disorder*, which is defined by "prolonged, moderately intense anxiety in the absence of an overt stressor" (Clark & Watson, 1994, p. 135).

Researchers should consider how these key distinctions between mood and emotions (i.e., intensity, duration, and breadth of activation) may be relevant to particular investigations of affect and psychopathology. On the basis of the considerations we have discussed, most self-report measures of affect—regardless of their names or intended target constructs—are better viewed as assessing moods rather than emotions, because they invariably assess responses ranging widely in intensity and duration, even when administered in clinical samples. We emphasize, however, that emotions and moods are not mutually exclusive and that both may be influenced by similar processes and share some common components (Clark & Watson, 1999), an insight captured nicely by Davidson (1994): "It also appears to be the case that moods and emotions dynamically interact in important ways. Emotions can lead to particular moods and moods can alter the probability that particular emotions will be triggered" (p. 53).

THE STRUCTURE OF AFFECT

To optimally define and assess affective experiences, it is necessary to understand the underlying structure of affect. Historically, there have been

two basic approaches to assessing the structure of affect: (a) discrete or specific affect models and (b) dimensional schemes. First, we briefly discuss discrete affect models and review issues associated with them. Next, we discuss the basic characteristics of dimensional models, focusing on two particularly popular schemes. Third, we describe an integrated three-level hierarchical model that neatly integrates all of these other schemes (Tellegen, Watson, & Clark, 1999).

Specific or Discrete Affect Models

Specific or discrete affect models posit the existence of specific types of affect and generally focus on emotions such as happiness, fear, anxiety, sadness, and anger (Gray & Watson, in press; Watson & Vaidya, 2003). This approach has been supported by a wide array of empirical evidence. Most notably, structural analyses of mood terms repeatedly have identified well-defined content factors corresponding to these specific affect states. Moreover, a common core of discrete affects—including fear, sadness, and anger—have emerged consistently in analyses based on widely varying item pools and diverse samples of respondents (Watson & Clark, 1997, 1999; Watson & Vaidya, 2003).

Having said that, however, we also must discuss two key problems associated with this approach. First, although a few core states are universally recognized, we still lack a compelling taxonomy of affect at the discrete, lower order level. That is, even after more than 50 years of study, affect researchers have not reached a consensus regarding the basic discrete states that must be included in any complete and comprehensive assessment of affect (see Watson & Clark, 1997; Watson & Vaidya, 2003). Second, structural analyses have established that measures created to assess specific affects typically show only limited discriminant validity (Watson & Clark, 1997). That is, measures of similarly valenced affects tend to be strongly intercorrelated, establishing a substantial level of nonspecificity in the data. For example, people who report feeling anxious also report feeling sad, angry, and guilty. In fact, multitrait–multimethod analyses consistently demonstrate much stronger evidence for nonspecificity (i.e., significant positive correlations among measures of different, similarly valenced affects) than for specificity (i.e., unique relations between different measures of the same target affect; Diener, Smith, & Fujita, 1995; Watson, 2000; Watson & Clark, 1992).

Dimensional Models of Affect

This evidence of strong nonspecificity indicates that mood can be characterized by a smaller number of general dimensions, thereby stimulating the development of dimensional models. Although three-dimensional schemes also have been proposed (see Watson & Tellegen, 1985), most attention has

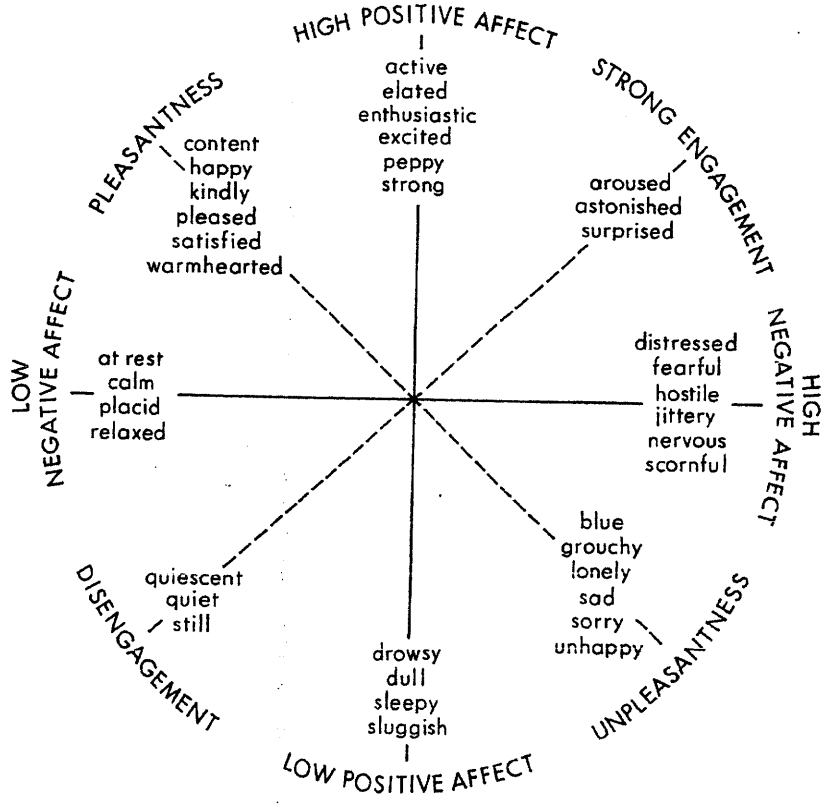

Figure 1.1. The two-factor structure of affect. From "Toward a Consensual Structure of Mood," by D. Watson and A. Tellegen, 1985, *Psychological Bulletin, 98,* p. 221. Copyright 1985 by the American Psychological Association.

been given to two-dimensional structures. On the basis of analyses of facial expressions and similarity ratings of mood terms, Russell (1980) proposed a two-factor dimensional model of affect that can be visually represented as a circumplex (i.e., mood terms mapped onto the perimeter of a circle within a two-dimensional space). In Russell's model, the circumplex is defined by two strongly bipolar dimensions: Pleasure–Displeasure and Arousal–Sleep. This model particularly emphasizes the bipolarity of the Pleasure–Displeasure dimension; that is, pleasant (e.g., happy, cheerful) and unpleasant (e.g., sad, lonely) affect are viewed as opposite ends of a single continuum. This model therefore suggests that one cannot simultaneously experience negative and positive affects and that the ends of the continuum should be highly negatively correlated.

On the basis of self-report data from several studies, Watson and Tellegen (1985) created a similar circumplex structure that is a rotational variant of Russell's model. Specifically, both schemes can be mapped onto a common circumplex structure with approximately a 45-degree rotation separating the two sets of reference axes (see Figure 1.1). Watson and Tellegen's model

emphasized two factors—Positive Affect (PA; also labeled *Positive Activation*) and Negative Affect (NA; also called *Negative Activation*)—that are largely unipolar and independent of one another. PA comprises mood terms such as *active, excited, energetic,* and *strong,* whereas NA is defined by terms such as *distressed, guilty, hostile,* and *nervous.* This model, with the largely independent factors of PA and NA, allows for the simultaneous experience of positive affect states and negative affect states; thus, according to this scheme, a person can feel both nervous and enthusiastic at the same time.

Watson and Tellegen's (1985) model addresses the issues of discriminant validity and specificity by positing that (a) similarly valenced mood states (e.g., sad and fearful) will be substantially correlated, whereas (b) oppositely valenced mood states (e.g., fearful and energetic) will be only weakly related; it is the latter property that is largely responsible for the quasi-independence of the NA and PA dimensions. It should be noted, however, that the independence of PA and NA mainly applies to low-intensity states (Watson, 1988; Watson et al., 1999). When experiencing an extremely intense, negative emotional state (e.g., terror), it is highly unlikely that one would simultaneously be experiencing an intense, positive emotional state (e.g., elation). However, mood states are generally less intense and tend to last for greater lengths of time, thereby allowing for the simultaneous experience of positive and negative feelings.

The key distinction between the Russell and the Watson and Tellegen schemes appears to be in their conceptualization of positive and negative affect as either largely independent (Watson & Tellegen) or as opposite ends of a single bipolar continuum (Russell). This has led to some confusion in the recent literature. It must be emphasized, however, that bipolarity and independence actually are key features of both models. In fact, a close inspection of Figure 1 indicates that some positive and negative descriptors are placed 180 degrees apart and should be strongly negatively correlated (e.g., happy vs. sad), whereas others are only 90 degrees apart and should be very weakly related to one another (e.g., enthusiastic vs. fearful). Thus, the circumplex model actually hypothesizes both independence and bipolarity, depending on the descriptors involved; it should be noted, moreover, that these hypothesized relations have been consistently confirmed empirically (see Tellegen et al., 1999; Watson, 1988; Watson & Tellegen, 1999).

Furthermore, a diverse array of research has subsequently demonstrated the empirical and heuristic value of this proposed independence between positive and negative affect. Research on hemispheric asymmetry in the prefrontal cortex (Tomarken & Keener, 1998) has linked two key biobehavioral systems—the behavioral activation system (BAS) and the behavioral inhibition system (BIS)—to PA and NA, respectively. PA essentially can be viewed as the affective component of the BAS, which exists in synchrony with its other components (e.g., cognitive, biological, and behavioral) to motivate the organism to seek out pleasure or rewards. Conversely, NA, the

affective component of the BIS, exists in synchrony with its other components to motivate the organism to avoid aversive stimuli or punishment (see Watson, 2000; Watson et al., 1999). In clinical research, the existence of these independent PA and NA dimensions has helped to elucidate the comorbidity of anxiety and depression (Watson, Clark, & Carey, 1988): Although both types of disorders are characterized by high negative affect, they show differential relations with positive affect (i.e., low positive affect is more prominent in depression than in anxiety).

Integrated Hierarchical Model

Thus far, we have treated discrete affect and dimensional models separately. Although these two approaches often are seen as antagonistic and incompatible, they actually can be integrated into a three-level hierarchical structure that incorporates key features from all of the models we have discussed (Tellegen et al., 1999). At its lowest or first-order level, this model includes nine specific affect dimensions (i.e., calm–ease, joy, interest, surprise, fear, anger–disgust–contempt, shame–guilt, sadness–distress, and low energy). The second-order level of this model consists of the relatively independent dimensions of PA and NA. Finally, at the highest or third-order level of this model is a bipolar happiness–unhappiness dimension that accounts for the bipolarity between pleasant and unpleasant affect.

One important implication of this hierarchical structure is that any complete examination of affect must model and assess it at both the discrete affect and dimensional levels; otherwise significant information likely will be lost. The relative importance of these levels will differ, however, according to the nature and goals of the research. For instance, the overarching Happiness dimension may be most relevant to the study of life satisfaction (Tellegen et al., 1999), whereas the second-level NA and PA dimensions may be more useful for many areas of personality and psychopathology research. Finally, a focus on at the lowest, discrete-affect level may be particularly informative in certain circumstances (e.g., in the study of self-esteem; see Watson, Suls, & Haig, 2002).

Affective Structure Across Samples

The bulk of the structural evidence we have reviewed is based on college student responses. This raises questions about the robustness of this structure across samples. In particular, psychopathology researchers may question whether these results will generalize to clinical samples. Accordingly, we present data to establish that the basic structure of affect is highly robust across samples. Table 1.1 shows correlations among four content-valid affect scales in four large samples: college students, psychiatric patients, community-dwelling adults, and high school students. The scales assess sad, depressed

TABLE 1.1
Correlations Between Mood Scales Created From the Iowa Depression and Anxiety Scales Item Pool

Scale	1	2	3	4
College students ($N = 673$)				
1. Depression	(.88)			
2. Anxiety	.76	(.91)		
3. Anger	.72	.71	(.93)	
4. Well-Being	−.47	−.38	−.32	(.84)
Psychiatric patients ($N = 353$)				
1. Depression	(.91)			
2. Anxiety	.77	(.93)		
3. Anger	.55	.57	(.94)	
4. Well-Being	−.49	−.38	−.20	(.88)
Community adults ($N = 362$)				
1. Depression	(.92)			
2. Anxiety	.78	(.93)		
3. Anger	.69	.74	(.94)	
4. Well-Being	−.54	−.43	−.39	(.90)
High school students ($N = 247$)				
1. Depression	(.90)			
2. Anxiety	.78	(.90)		
3. Anger	.73	.72	(.94)	
4. Well-Being	−.52	−.40	−.37	(.86)

Note. Internal consistency reliabilities (coefficient alphas) are shown in parentheses. All correlations are significant at $p < .01$, two-tailed.

mood (Depression, 5 items; e.g., "I felt depressed"); anxious, fearful mood (Anxiety, 9 items; e.g., "I felt anxious"); angry, irritable mood (Anger, 11 items; e.g., "I felt irritable"); and pleasant, positive mood (Well-Being, 8 items; e.g., "I felt hopeful about the future"). All respondents rated the extent to which they experienced each item "during the past two weeks, including today" on a 5-point scale ranging from *not at all* to *extremely*.

Several aspects of these data are noteworthy. First, the overall pattern is highly consistent across the four samples, indicating that the structural evidence we have reviewed should generalize well to clinical populations. Second, the correlations among the Depression, Anxiety, and Anger scales consistently are high, demonstrating the strong influence of the general NA dimension. Third, the correlations among the Depression and Anxiety scales are particularly high, which has potentially important implications for our understanding of psychopathology. Most notably, they establish a strong link between sad–depressed mood (a core, defining element of the mood disorders) and fearful–anxious mood (a key feature of the anxiety disorders). On

the basis of this and other evidence, Watson (2005) has suggested that the current distinction between the mood and anxiety disorders is not particularly useful and that it should be replaced by an empirically based taxonomy that reflects the actual similarities among disorders. Finally, the Well-Being scale consistently has significantly stronger negative correlations with Depression than with either Anxiety or Anger; this again establishes that (low) PA is more strongly linked to depression than to other types of negative affect.

SELF-REPORT MEASURES OF AFFECT

Because of the large number of self-report affect measures that currently are available, our review necessarily must be selective. We focus on those measures that are the most important, the most recent, or the most widely used; that have demonstrated expected patterns of correlations with psychopathology symptoms; and/or that show the best psychometric properties (see Table 1.2 for a summary of the reviewed measures). For a more detailed review of self-report measures, see Watson and Vaidya (2003) and Gray and Watson (in press).

Specific or Discrete Affect Measures

The Mood Adjective Check List (MACL; Nowlis, 1965) assesses 12 factor-analytically derived affects (e.g., Fatigue, Aggression, Surgency, Anxiety, Elation, Concentration, Social Affection, Sadness, Skepticism, Egotism, Vigor, and Nonchalance) using adjectives that are rated on a 4-point scale. The MACL was designed to assess current mood as well as to detect changes in mood. Although it had a highly influential role in the early literature on the structure and assessment of affect, the MACL failed to become a standard measure in the field, in part because its basic psychometric properties were never clearly established (see Watson & Vaidya, 2003).

The Profile of Mood States (POMS; McNair, Lorr, & Droppleman, 1971) consists of 65 adjectives that are rated on a 4- or 5-point scale. It assesses six specific affects: anger–hostility, vigor–activity, fatigue–inertia, confusion–bewilderment, tension–anxiety, and depression–dejection. The scale was originally created to assess mood fluctuations in psychiatric patients and has since been validated in other settings (Lane & Lane, 2002). Multiple short forms assessing fewer factors were subsequently created as well (e.g., Curran, Andrykowski, & Studts, 1995; Shacham, 1983). Evidence indicates that the POMS scales (a) demonstrate acceptable internal consistency reliabilities and moderate short-term stability, (b) are sensitive to changes due to therapy, and (c) show good concurrent and predictive valid-

TABLE 1.2
Self-Report Measures of Affect

Measure	State or trait	Reliability	Subscales
			Discrete measures
MACL	State	N/A	Fatigue, Aggression, Surgency, Anxiety, Elation, Concentration, Social Affection, Sadness, Skepticism, Egotism, Vigor, Nonchalance
POMS	Both	Moderate	Anger–Hostility, Vigor–Activity, Fatigue–Inertia, Confusion–Bewilderment, Tension–Anxiety, Depression–Dejection
DES	Both	Fair	Interest, Joy, Surprise, Sadness, Anger, Disgust, Contempt, Fear, Shame/Shyness,
MAACL–R	Both	Good	Guilt/Anxiety, Depression, Hostility, Positive Affect, Sensation Seeking
PANAS–X	Both	Good	Fear, Sadness, Guilt, Hostility, Joviality, Self-Assurance, Attentiveness, Shyness, Fatigue, Serenity, Surprise
			Dimensional measures
AD-ACL	State	Good	General Activation, Deactivation-Sleep, High Activation, General Deactivation
Affect Grid	State	N/A	Pleasure–Displeasure, Arousal–Sleep
CMQ	State	Fair–moderate	Pleasure–Displeasure, Arousal–Sleep
PANAS	Both	Good	Positive Affect, Negative Affect
UMACL	State	Good	Tense Arousal, Energetic Arousal, Hedonic Tone

Note. Reliability is internal consistency reliability. The rating scale is Poor, Fair, Moderate, or Good. MACL = Mood Adjective Check List; POMS = Profile of Mood States; DES = Differential Emotions Scale; MAACL–R = Multiple Affect Adjective Check List—Revised; PANAS–X = Positive and Negative Affect Schedule—Expanded Form; AD-ACL = Activation–Deactivation Adjective Check List; CMQ = Current Mood Questionnaire; PANAS = Positive and Negative Affect Schedule; UMACL = UWIST Mood Adjective Checklist.

ity (Lane & Lane, 2002; McNair & Lorr, 1964; Payne, 2001). However, the negative mood scales of the POMS are strongly intercorrelated and fail to demonstrate strong discriminant validity with one another (Watson & Clark, 1999).

The Differential Emotions Scale (DES; Izard, Libero, Putnam, & Haynes, 1993) originally was designed to measure 10 discrete emotions: interest, joy, surprise, sadness, anger, disgust, contempt, fear, shame/shyness and guilt. There are multiple versions of the DES using different response formats and instructions; depending on the instructions used, the DES can be modified to assess current, past week, or long-term affect. In its most recent version, the Differential Emotions Scale—IV, shyness and shame are measured separately and a new subscale, Inner-Directed Hostility, has been added, creating a total of 12 subscales (Izard et al., 1993). Although the subscales are stable over time, many of them show moderate to high intercorrelations and only low to moderate internal consistencies (see Watson & Vaidya, 2003), which is largely due to the low number of items per scale (typically only three apiece across the various versions of the instrument).

The Multiple Affect Adjective Check List—Revised (MAACL–R; Zuckerman & Lubin, 1985) assesses both state and trait affect. It contains five scales: Anxiety, Depression, Hostility, Positive Affect, and Sensation Seeking. The original MAACL (Zuckerman & Lubin, 1965) was revised in 1985 because of widespread evidence of poor discriminant validity among its Anxiety, Depression, and Hostility scales (despite these concerns, however, the original MAACL continues to be used and actually has been cited hundreds of times since the publication of the MAACL–R). The MAACL–R has many strong psychometric properties, including generally good coefficient alphas and test–retest reliability. Unfortunately, the negative affective scales continue to demonstrate high intercorrelations and questionable discriminant validity (Watson & Vaidya, 2003). Finally, it should be noted that one of the original authors has developed a short-form version of the MAACL–R (Lubin, Whitlock, Reddy, & Petren, 2001).

The Positive and Negative Affect Schedule—Expanded Form (PANAS–X; Watson & Clark, 1999) is a factor-analytically derived measure that assesses 11 specific affects; this includes four negative mood states (Fear, Sadness, Guilt, and Hostility), three positive mood states (Joviality, Self-Assurance, and Attentiveness), and four scales that are less consistently related to the higher order NA and PA dimensions (Shyness, Fatigue, Serenity, and Surprise). The PANAS–X contains mood terms rated on a 5-point scale ranging from *not at all* to *extremely*; instructions can be varied to assess either state or trait affect. Extensive reliability and validity data from multiple samples have been reported on the PANAS–X (Watson & Clark, 1997, 1999). The longer scales are highly reliable (e.g., coefficient alphas .83 or higher), and its shorter scales still consistently demonstrate adequate reliabilities (e.g., above .76).

Dimensional Affect Measures

The Activation–Deactivation Adjective Check List (AD-ACL; Thayer, 1967) is one of the earliest dimensional measures of affect. The AD-ACL was developed out of research using the MACL (Nowlis, 1965). The AD-ACL contains a series of affect adjectives that are rated on a 4-point scale. It consists of four factor-analytically derived factors: General Activation (Energy), Deactivation–Sleep, High Activation, and General Deactivation (Calmness). Further analyses have shown two higher order activation dimensions—Energy versus Tiredness and Tension versus Inactivation (Thayer, 1967, 1978)—that are broadly similar to PA and NA, respectively (Yik, Russell, & Feldman Barrett, 1999). The AD-ACL demonstrates excellent test–retest reliability and validity and can be administered in approximately 2 minutes (Thayer, 1978).

The Affect Grid (Russell, Weiss, & Mendelsohn, 1989) represents affect in the two-dimensional space representing Russell's version of the circumplex. The 9 × 9 grid contains affect descriptors in each corner and at each midpoint along the sides (i.e., in each cell). Respondents are asked to place a check in the cell that best captures their current affect. The descriptors, moving clockwise from the top left corner of the grid, include *stress*, *high arousal*, *excitement*, *pleasant feelings*, *relaxation*, *sleepiness*, *depression*, and *unpleasant feelings*. Thus, the two major dimensions of Pleasure–Displeasure and Arousal–Sleep are represented as bipolar opposites (see Russell, Weiss, & Mendelsohn, 1989, Figure 1, p. 494). The Affect Grid has demonstrated acceptable interrater reliability and good convergent validity with other dimensional measures of affect, and it is particularly well suited for quick and frequent assessment (Russell et al., 1989). However, one encounters obvious difficulties in assessing internal consistency reliability with single item measures; moreover, such instruments are likely more affected by systematic or random error than other types of affect scales.

The Current Mood Questionnaire (CMQ; Feldman Barrett & Russell, 1998) assesses all eight octants of Russell's circumplex using multiple response formats for each dimension; the inclusion of multiple measures for each construct allows researchers to use structural equation modeling to correct for both random and systematic error (Feldman Barrett & Russell, 1998). The CMQ was developed to test key aspects of Russell's model, particularly the assertion that the dimensions of pleasure versus displeasure and arousal versus sleep are fully bipolar. The CMQ scales demonstrate good convergent and discriminant validity, and the Pleasure–Displeasure scale consistently demonstrates acceptable internal consistency reliability (Feldman Barrett & Russell, 1998; Watson & Vaidya, 2003). However, the Arousal–Sleep dimension shows less acceptable internal consistency, and the scales defining this factor do not appear to be fully bipolar (Watson & Vaidya, 2003). In addition, the use of multiple response formats makes the CMQ longer to

complete and more cumbersome than the other dimensional measures included in this review; this, in turn, makes it less attractive for use in many contexts. Recent CMQ citations are low in number relative to other dimensional measures of affect.

The Positive and Negative Affect Schedule (PANAS; Watson et al., 1988)—which later was subsumed into the PANAS–X (see Watson & Clark, 1997, 1999)—is a brief measure of the two major dimensions in the Watson and Tellegen model and can be used to assess either state or trait affects with a slight modification in instructions. The PANAS was factor-analytically derived from Zevon and Tellegen's (1982) set of 60 adjectives. PA and NA are each assessed with 10 adjectives, which participants rate on a 5-point scale. The scales show excellent internal consistency and convergent and discriminant validity.

The UWIST Mood Adjective Checklist (UMACL; Matthews, Jones, & Chamberlain, 1990) was designed to synthesize the competing structural models proposed by Watson and Tellegen (1985), Russell (1980), and Thayer (1986). The UMACL includes adjectives rated on a 4-point response scale that tap the affective dimensions of tense arousal, energetic arousal (reflecting constructs from both the Thayer and Watson and Tellegen models), and hedonic tone (i.e., Russell's pleasure–displeasure dimension). In addition, the adjectives can also be scored to yield a General Arousal Index, corresponding to Russell's arousal–sleep dimension. Most of the UMACL scales show excellent internal consistency and convergent and discriminant validity, although the arousal scales correlate moderately with hedonic tone.

BASIC CONSIDERATIONS IN AFFECT MEASUREMENT

In many ways, the basic considerations involved in assessing affect are the same as those involved in measuring any psychological construct. Any construct must be assessed in a way that is both valid and reliable. However, because of the internal and subjective nature of affect, measurement in this area also faces some unique assessment issues. For example, to provide valid assessments, participants must be able to both (a) synthesize information regarding their affective experiences and then (b) accurately report that information. In the following section, we outline the evidence supporting the validity and reliability of affect measurement and report on some factors that contribute to measurement error.

CONSTRUCT VALIDITY OF TRAIT AFFECT MEASURES

Self–Other Agreement of Trait Affect

Personality researchers have long used self–other agreement, that is, the convergence between a self-rating and a peer rating of the same target

(e.g., Person A's self-rating vs. Person B's ratings of Person A), as evidence for the validity of personality measures. Until recently, few studies have examined self–other agreement in trait affect. This neglect can be attributed to the *trait visibility effect* (i.e., highly observable traits will yield better self–other agreement than more internalized traits; see Watson et al., 2000). Affective experiences are highly internal and subjective, and thus researchers anticipated finding relatively poor convergence because of this effect. However, recent studies consistently have found significant correlations between self- and other-ratings of trait affect (Diener et al., 1995; Watson, Hubbard, & Wiese, 2000). Watson and Vaidya (2003) reported PANAS–X data from four samples: 279 friendship dyads, 68 dating couples, 136 dating couples, and 74 married couples. The PANAS–X scales (with the single exception of Surprise) consistently showed significant, moderate levels of convergent validity (weighted mean correlations across the samples ranged from .25 to .42.) Furthermore, there was clear evidence of the *acquaintanceship effect*, that is, the tendency for individuals who know each other well to generate higher self–other agreement correlations; thus, higher convergent correlations were found in the married couples than in the dating samples, whereas the dating couples tended to produce higher correlations than the friendship dyads. These results help to establish the construct validity of trait measures of affect.

Although these correlations are encouraging, it is interesting to note that they tend to be lower than those of standard personality measures. For instance, scales assessing the Big Five personality traits had agreement correlations ranging from .42 to .53 in three of these same samples (see Watson et al., 2000). Even though affective experiences are less "visible" than most personality traits, the trait visibility effect cannot fully explain the differences in self–other agreement found between neuroticism and trait negative affect. These two constructs are highly correlated; the mean correlation between neuroticism and the PANAS–X Negative Affect scale was .60 (self-ratings) and .69 (other-ratings) in the three samples reported in Watson et al. (2000). Nevertheless, the neuroticism scales produced higher self–other agreement correlations in all three groups (.37–.59) than did PANAS–X Negative Affect (.20–.44). This cannot be attributed to differences in the internal consistency of the scales (Watson et al., 2000; Watson & Vaidya, 2003) and becomes even more intriguing when one considers their content. For instance, the Neuroticism scale of the Big Five Inventory (BFI; John & Srivastava, 1999) is strongly affective in character (Pytlik Zillig, Hemenover, & Dienstbier, 2002); nevertheless, it consistently shows better self–other agreement than the PANAS–X Negative Affect scale.

Temporal Stability of Trait Affect

Temporal stability is a necessary property of any construct that is defined as a trait. Watson (2004) examined both short-term stability (2 months; 465 students) and stability over a much longer time span (approximately 2.5

years; 392 participants). Strong retest correlations were found for the PANAS–X in the short-term stability study; coefficients ranged from .67 to .76, with a mean of .70 for the negative affect scales and .71 for the positive affect scales (Watson, 2004). The data from the long-term study demonstrated that trait affect as measured by the PANAS–X is moderately stable over a 2.5-year period; correlations ranged from .46 to .55, with mean coefficients of .49 (negative affect scales) and .51 (positive affect scales; Watson, 2004). These data enhance the construct validity of trait affect measures by demonstrating that they have a stable dispositional component.

Paralleling the self–other agreement data, however, the temporal stabilities of trait affect measures are lower than those of standard personality scales. For example, in these same two studies, Watson (2004) reported that the short-term BFI retest correlations ranged from .79 to .89 (mean r = .82), whereas the long-term BFI stabilities ranged from .59 to .72 (mean r = .64). Many of these correlations are significantly higher than those of the PANAS–X. This pattern is not entirely unexpected (see Watson, 2004) and, in general, could be explained by systematic differences in content between these scales. However, we again see a puzzling discrepancy between the BFI Neuroticism and PANAS–X Negative Affect scales: Watson (2004) found that BFI Neuroticism was significantly more stable in both the long- and short-term retests than PANAS–X Negative Affect. Given that this discrepancy cannot be attributed to differences in internal consistency reliability or item content, other, more subtle factors must be responsible.

Retrospective Recall of Affective States

We now turn our attention to possible sources of measurement error that potentially could lessen the validity and reliability of self-report affect scales. One potential problem with the measurement of longer term affect (e.g., mood rated over the past week, past month, or in general) is its retrospective nature. Participants have to remember their past experiences and recall their previous affective states, which could lead to several problems. For example, raters could experience *duration neglect*, that is, they could be relatively insensitive to how long a particular affect lasted and instead give more attention to the overall intensity of the experience. They may also be influenced by the *recency effect*, that is, the tendency for individuals to weight recent experiences more heavily than earlier experiences. Furthermore, there is evidence indicating that a person's current, momentary mood at the time of assessment may influence his or her retrospective ratings of past affective experiences (see Stone, Shiffman, & deVries, 1999). Given these problems, several researchers have advocated an alternative method for assessing trait affect. Specifically, they have proposed that a large number of state affect ratings be aggregated into one composite score (i.e., averaging several ratings of current mood assessed across time; see Stone et al., 1999).

In general, this process results in aggregated ratings that display moderate to strong levels of convergence with traditional, global measures of trait affect. Watson and Vaidya (2003) reported convergent correlations ranging from .37 to .60 (*Mdn* = .51) with aggregated daily ratings and from .45 to .64 (*Mdn* = .53) with aggregated weekly ratings. Although these correlations demonstrate that the two approaches converge well, the correlations are far from +1.00. This, then, raises an interesting question: Which type of rating is more valid? We already have discussed the potential problems associated with retrospective ratings of longer term affect; it must be emphasized, however, that the aggregation of state ratings produces problems of its own. The most serious problem is the reduced discriminant validity of aggregated measures of specific, lower order affects. Diener et al. (1995) were the first to demonstrate this effect. They collected global ratings of four negative affects (fear, anger, sadness, shame) and corresponding aggregated daily ratings from 212 participants. Although the convergence between the two methods was strong (mean r = .62), there was a striking difference in discriminant validity across the two methods. In the global ratings, correlations among the four negative affect scales ranged from .54 to .61, with a mean value of .58; in marked contrast, the corresponding correlations in the aggregated ratings ranged from .70 to .79, with a mean value of .75. Watson and Tellegen (2002) replicated this finding with the PANAS–X scales and also extended it by demonstrating that the same pattern emerged using positive affect scales.

The most likely explanation for this phenomenon is that systematic measurement errors (e.g., acquiescence) are inflated during the aggregation process; this, in turn, artificially increases the correlations between scales assessing similarly valenced constructs (e.g., fear and sadness; see Watson & Tellegen, 2002; Watson & Vaidya, 2003). This may seem counterintuitive, given that one of the frequently touted advantages of aggregation is that it reduces measurement error. However, this is only true when the errors are random and, therefore, increasingly cancel each other out with repeated assessment. When the errors are instead systematic, they can be correlated across the assessments; therefore, instead of neutralizing systematic error, aggregation will actually increase it in some cases. Furthermore, it appears that this same basic process also is responsible for the almost complete lack of bipolarity in aggregated affect ratings (Watson & Tellegen, 2002). In light of these data, we conclude that although aggregated ratings may be advantageous in some contexts, global trait ratings generally are superior and should be the preferred method (see also Watson & Tellegen, 2002; Watson & Vaidya, 2003).

Unidentified Sources of Error

With a few exceptions, affect researchers have given little attention to the response formats of their measures or to the specific instructions that are

given to participants. Recent data strongly suggest that these issues may play an important role in the reliability and validity of affect measures. There is increasing evidence that different information processing systems—such as schematic memory, semantic memory, and autobiographical–episodic memory—may be activated depending on the specific assessment approach taken (Robinson & Clore, 2002). In this regard, Watson (2004) showed that highly correlated measures of obsessive–compulsive disorder symptoms, dissociative tendencies, personality traits, and trait affectivity yielded significantly different levels of stability over a 2-month period. Furthermore, broad differences in item content could not account for the discrepancy in stability between matched pairs of scales. This evidence of differential stability becomes even more important when one considers the short retest interval (i.e., 2 months) between assessments in this study; this makes it unlikely that these effects are due to true changes on these dimensions. It therefore seems likely that this differential stability results from different levels of measurement error across instruments. This, in turn, leads to the intriguing possibility that subtle differences between scales (e.g., instructions, response formats, wording effects) can lead to varying amounts of error variance.

To investigate this idea, Watson (2004) created the Temperament and Emotion Questionnaire (TEQ) by taking PANAS–X items and embedding them in sentences. For example, the PANAS–X item "sad" became "I often feel a bit sad." In addition, the response format was changed to a 5-point agree–disagree scale. This process ensured that the TEQs content was extremely similar to the PANAS–X. Furthermore, strong convergent correlations between corresponding PANAS–X and TEQ scales (mean r = .70) demonstrated that these instruments are assessing the same basic affective constructs. Nevertheless, the 2-month stability correlations for the TEQ negative affect scales were higher than those of their PANAS–X counterparts; these differences were statistically significant for the Fear, Sadness, and Hostility scales. These results suggest that, in some cases, measurement error can be reduced simply by embedding standard affect descriptors in sentences; however, the BFI Neuroticism scale still was significantly more stable than the TEQ Negative Affect scale. Therefore, other subtle differences also contribute to the increased error variance in standard measures of trait affect. We are currently investigating these differences in new studies of temporal stability and self–other agreement.

The Influence of Social Desirability

Because self-report affect measures almost invariably contain face-valid items whose content is not hidden, it is possible that participants could respond in a manner that is not entirely accurate. For example, participants may respond defensively and distort their answers (either consciously or unconsciously) in a self-enhancing manner. One way to investigate the extent

to which social desirability is a problem in affect measurement is by obtaining other-ratings from judges who know the target individual well. Because others should not be as inclined to rate targets in a socially desirable manner, comparisons of other-ratings with the targets' own self-ratings help to determine the extent to which social desirability introduces error into mood measurement. Thus, if social desirability were a significant problem, then self-ratings would be expected to yield lower levels of negative affect and higher levels of positive affect (e.g., responses in the socially desirable direction) than other-ratings. To investigate this issue, Watson and Vaidya (2003) analyzed self- and other-ratings of the trait version of the PANAS–X from friendship, dating, and married dyads. Comparisons of the self- and other-ratings indicated that social desirability does not play a substantial role in self-ratings of affect.

Reducing Error and Improving Measurement

Given the findings we have examined, we believe it is essential that affect researchers pay careful attention to measurement-related issues. We especially encourage studies that allow for side-by-side comparisons of scales that tap the same (or very similar) constructs. Comparing the reliabilities and validities of similar scales in the same sample enables researchers to identify the measures that best suit their needs. Furthermore, such comparisons enhance our understanding of the issues that influence validity and reliability. Too often, researchers view stability and validity as dichotomous, "either–or" properties; they are eager to conclude that their instruments are adequate or satisfactory without giving these issues much real thought (see Watson, 2004). A more nuanced dimensional approach in which researchers investigate how specific factors influence validity, reliability, and error would lead to the creation of more valid and reliable assessment instruments.

CONCLUSION

We conclude this brief introduction to the assessment of self-rated affect by emphasizing several basic points. First, affect assessment should be guided by a thorough understanding of the underlying structure of this domain. As we have discussed, any complete and comprehensive assessment should acknowledge the hierarchical structure of this domain and, accordingly, should include measures of both general dimensions and specific, discrete affects. Second, researchers can choose between instruments that show very different psychometric properties. We therefore urge investigators to examine these properties carefully—including internal consistency, test–retest reliability, and convergent and discriminant validity—before selecting instruments for their research. Third, we now have extensive evidence establishing the

reliability and construct validity of commonly used self-report measures. As we have discussed, however, measures of longer term affect, including both global and aggregated ratings, are subject to a variety of forces that may lessen their reliability and validity. Further research is needed to identify specific sources of measurement error that, if minimized, will enable researchers to create a new generation of even better self-report affect measures.

REFERENCES

Clark, L. A., & Watson, D. (1994). Distinguishing functional from dysfunctional affective responses. In P. Ekman & R. J. Davidson (Eds.), *The nature of emotion: Fundamental questions* (pp. 131–136). New York: Oxford University Press.

Clark, L. A., & Watson, D. (1999). Temperament: A new paradigm for trait psychology. In L. Pervin & O. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 399–423). New York: Guilford Press.

Curran, S. L., Andrykowski, M. A., & Studts, J. L. (1995). Short form of the Profile of Mood States (POMS–SF): Psychometric information. *Psychological Assessment, 7*, 80–83.

Davidson, R. J. (1994). On emotion, mood and related affective constructs. In P. Ekman & R. J. Davidson (Eds.), *The nature of emotion: Fundamental questions* (pp. 51–55). New York: Oxford University Press.

Diener, E., Smith, H., & Fujita, F. (1995). The personality structure of affect. *Journal of Personality and Social Psychology, 69*, 130–141.

Feldman Barrett, L., & Russell, J. A. (1998). Independence and bipolarity in the structure of current affect. *Journal of Personality and Social Psychology, 74*, 967–984.

Gray, E. K., & Watson, D. (in press). Assessing positive and negative affect via self report. In J. J. B. Allen & J. A. Coan (Eds.), *The handbook of emotion elicitation and assessment*. New York: Oxford University Press.

Izard, C. E., Libero, D. Z., Putnam, P., & Haynes, O. M. (1993). Stability of emotion expression experiences and their relations to traits of personality. *Journal of Social and Personality Psychology, 64*, 847–860.

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A Pervin & O. P. John (Eds.), *Handbook of personality* (2nd ed., pp. 102–138). New York: Guilford Press.

Lane, A. M., & Lane, H. J. (2002). Predictive effectiveness of mood measures. *Perceptual and Motor Skills, 94*, 785–791.

Lubin, B., Whitlock, R. V., Reddy, B., & Petren, S. (2001). A comparison of the short and long forms of the Multiple Affect Adjective Check List—Revised (MAACL–R). *Journal of Clinical Psychology, 57*, 411–416.

Matthews, G., Jones, D. M., & Chamberlain, A. G. (1990). Refining the measurement of mood: The UWIST Mood Adjective Checklist. *British Journal of Psychology, 81*, 17–42.

McNair, D. M., & Lorr, M. (1964). An analysis of mood in neurotics. *Journal of Abnormal and Social Psychology*, 69, 620–627.

McNair, D. M., Lorr, M., & Droppleman, L. F. (1971). *Manual: Profile of Mood States*. San Diego, CA: Educational and Industrial Testing Service.

Nowlis, V. (1965). Research with the Mood Adjective Check List. In S. S. Tompkins & C. E. Izard (Eds.), *Affect, cognition, and personality: Empirical studies* (pp. 352–389). New York: Springer Publishing Company.

Payne, R. (2001). Measuring emotions at work. In R. L. Payne & C. L. Cooper (Eds.), *Emotions at work: Theory, research and applications in management* (pp. 107–129). West Sussex, England: Wiley.

Pytlik Zillig, L. M., Hemenover, S. H., & Dienstbier, R. A. (2002). What do we assess when we assess a Big 5 trait? A content analysis of the affective, behavior, and cognitive processes represented in Big 5 personality inventories. *Personality and Social Psychology Bulletin*, 28, 847–858.

Robinson, M. D., & Clore, G. L. (2002). Belief and feeling: Evidence for an accessibility model of emotional self-report. *Psychological Bulletin*, 128, 934–960.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39, 1161–1178.

Russell, J. A., Weiss, A., & Mendelsohn, G. A. (1989). Affect Grid: A single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology*, 57, 493–502.

Shacham, S. (1983). A shortened version of the Profile of Mood States. *Journal of Personality Assessment*, 47, 305–306.

Stone, A. A., Shiffman, S. S., & deVries, M. W. (1999). Ecological momentary assessment. In D. Kahneman, E. Diener, & N. Schwarz (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 26–29). New York: Russell Sage Foundation.

Tellegen, A., Watson, D., & Clark, L. A. (1999). On the dimensional and hierarchical structure of affect. *Psychological Science*, 10, 297–303.

Thayer, R. E. (1967). Measurement of activation through self-report. *Psychological Reports*, 20, 663–678.

Thayer, R. E. (1978). Factor analytic and reliability studies on the Activation–Deactivation Adjective Check List. *Psychological Reports*, 42, 747–756.

Thayer, R. E. (1986). Activation–Deactivation Adjective Check List: Current overview and structural analysis. *Psychological Reports*, 58, 607–614.

Tomarken, A. J., & Keener, A. D. (1998). Frontal brain asymmetry and depression: A self-regulatory perspective. *Cognition and Emotion*, 12, 387–420.

Watson, D. (1988). The vicissitudes of mood measurement: Effects of varying descriptors, time frames, and response format on measures of positive and negative affect. *Journal of Personality and Social Psychology*, 55, 128–141.

Watson, D. (2000). *Mood and temperament*. New York: Guilford Press.

Watson, D. (2004). Stability versus change, dependability versus error: Issues in the assessment of personality over time. *Journal of Research in Personality, 38,* 319–350.

Watson, D. (2005). Rethinking the mood and anxiety disorders: A quantitative hierarchical model for *DSM–V. Journal of Abnormal Psychology, 114,* 122–136.

Watson, D., & Clark, L. A. (1992). Affects separable and inseparable: On the hierarchical arrangements of the negative affects. *Journal of Personality and Social Psychology, 62,* 489–505.

Watson, D., & Clark, L. A. (1997). Measurement and mismeasurement of mood: Recurrent and emergent issues. *Journal of Personality Assessment, 68,* 267–296.

Watson, D., & Clark, L. A. (1999). *The PANAS–X: Manual for the Positive and Negative Affect Schedule—Expanded Form.* Retrieved September 8, 2006, from http://www.psychology.uiowa.edu/Faculty/Watson/Watson.html

Watson, D., Clark, L. A., & Carey, G. (1988). Positive and negative affectivity and their relation to anxiety and depressive disorders. *Journal of Abnormal Psychology, 97,* 346–353.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54,* 1063–1070.

Watson, D., Hubbard, B., & Wiese, D. (2000). Self–other agreement in personality and affectivity: The role of acquaintanceship, trait visibility, and assumed similarity. *Journal of Personality and Social Psychology, 78,* 546–558.

Watson, D., Suls, J., & Haig, J. (2002). Global self-esteem in relation to structural models of personality and affectivity. *Journal of Personality and Social Psychology, 83,* 185–197.

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin, 98,* 219–235.

Watson, D., & Tellegen, A. (1999). Issues in the dimensional structure of affect—Effects of descriptors, measurement error, and response formats: Comment on Russell and Carroll (1999). *Psychological Bulletin, 125,* 601–610.

Watson, D., & Tellegen, A. (2002). Aggregation, acquiescence, and the assessment of trait affectivity. *Journal of Research in Personality, 38,* 589–597.

Watson, D., & Vaidya, J. (2003). Mood measurement: Current status and future directions. In J. A. Schinka & W. Velicer (Eds.), *Comprehensive handbook of psychology: Vol. 2. Research methods* (pp. 351–375). New York: Wiley.

Watson, D., Wiese, D., Vaidya, J., & Tellegen, A. (1999). The two general activation systems of affect: Structural findings, evolutionary considerations, and psychobiological evidence. *Journal of Personality and Social Psychology, 76,* 820–838.

Yik, M. S. M., Russell, J. A., & Feldman Barrett, L. (1999). Structure of self-reported current affect: Integration and beyond. *Journal of Personality and Social Psychology, 77,* 600–619.

Zevon, M. A., & Tellegen, A. (1982). The structure of mood change: An idiographic/nomothetic analysis. *Journal of Personality and Social Psychology, 43*, 111–122.

Zuckerman, M., & Lubin, B. (1965). *Manual for the Multiple Affect Adjective Check List*. San Diego, CA: Educational and Industrial Testing Service.

Zuckerman, M., & Lubin, B. (1985). *Manual for the MAACL–R: The Multiple Affect Adjective Check List—Revised*. San Diego, CA: Educational and Industrial Testing Service

A Survey of Affect Recognition Methods: Audio, Visual and Spontaneous Expressions

Zhihong Zeng[1], Maja Pantic[2], Glenn I. Roisman[1] and Thomas S. Huang[1]
[1]University of Illinois at Urbana-Champaign, USA
[2]Imperial College London, UK / University of Twente, Netherlands
{zhzeng,huang}@ifp.uiuc.edu, m.pantic@imperial.ac.uk, roisman@uiuc.edu

ABSTRACT
Automated analysis of human affective behavior has attracted increasing attention from researchers in psychology, computer science, linguistics, neuroscience, and related disciplines. Promising approaches have been reported, including automatic methods for facial and vocal affect recognition. However, the existing methods typically handle only deliberately displayed and exaggerated expressions of prototypical emotions--despite the fact that deliberate behavior differs in visual and audio expressions from spontaneously occurring behavior. Recently efforts to develop algorithms that can process naturally occurring human affective behavior have emerged. This paper surveys these efforts. We first discuss human emotion perception from a psychological perspective. Next, we examine the available approaches to solving the problem of machine understanding of human affective behavior occurring in real-world settings. We finally outline some scientific and engineering challenges for advancing human affect sensing technology.

Categories and Subject Descriptors
A.1 [**Introduction and Survey**]
H.1.2 [**User/Machine Systems**]: Human information processing
H.5.1 [**Multimedia Information Systems**]: Evaluation/ methodology
I.5.4 [**Pattern Recognition Applications**]

General Terms
Algorithms, Performance.

Keywords
Multimodal human computer interaction, multimodal user interfaces, affective computing, human computing, affect recognition, emotion recognition.

1. INTRODUCTION
A widely accepted prediction is that computing will move to the background, weaving itself into the fabric of our everyday living spaces and projecting the human user into the foreground. Consequently, the future "ubiquitous computing" environments will need to have human-centered designs instead of computer-centered designs [15], [20], [57], [63], [64]. A change in the user's affective state is a fundamental component of human-human communication. Some affective states motivate human actions and others enrich meaning of human communication. Consequently, the traditional HCI that ignores the user's affective states filters out a large portion of the information available in the interaction process. Human Computing paradigm suggests that user interfaces of the future need to be proactive and human-centered, based on naturally occurring multimodal human communication [57]. More specifically, human-centered interfaces must have the ability to detect subtleties of and changes in the user's behavior, especially his or her affective behavior, and to initiate interactions based on this information, rather than simply responding to the user's commands.

Fig 1 illustrates a prototype of such an affect-sensitive, multimodal computer-aided learning system. The system was built during the NSF ITR project titled "Multimodal Human Computer Interaction: Toward a Proactive Computer"[1]. In this learning environment, the user explores Lego gear games by interacting with a computer avatar. Multiple sensors are used to detect and track the user's behavioral cues and his or her task. More specifically, the useful information recognized from these sensors includes the user's emotional state, engagement state, the utilized speech keywords, and the gear state. Based on this information, the avatar offers an appropriate tutoring strategy in this interactive learning environment. Other examples of affect-sensitive, multimodal HCI systems include the system of Duric et al. [22], which applies a model of embodied cognition that can be seen as a detailed mapping between the user's affective states and the types of interface adaptations, and the proactive HCI tool of Maat and Pantic [51] capable of learning the user's context-dependent behavioral patterns from multi-sensory data and of adapting the interaction accordingly, and the automated Learning Companion of Kapoor et al. [43] that combines information from cameras, a sensing chair and mouse, and wireless skin sensor to detect frustration in order to predict when the user need help. These systems demonstrate a rough picture of future multimodal human-computer interaction.

Except in standard HCI scenarios, potential commercial applications of automatic human affect recognition include affect-sensitive systems for customer services, call centers [46], intelligent automobile system [40], and game and entertainment industry. These systems will change the nature of human-computer interaction in our daily lives. Another important

ICMI'07, November 12–15, 2007, Nagoya, Aichi, Japan.

[1] http://itr.beckman.uiuc.edu

application of automated systems for human affect recognition is in affect-related research (e.g. in psychology, psychiatry, behavioral and neuroscience), where such systems can improve the quality of the research by improving the reliability of measurements and speeding up the currently tedious, manual task of processing data on human affective behavior [27], [66].

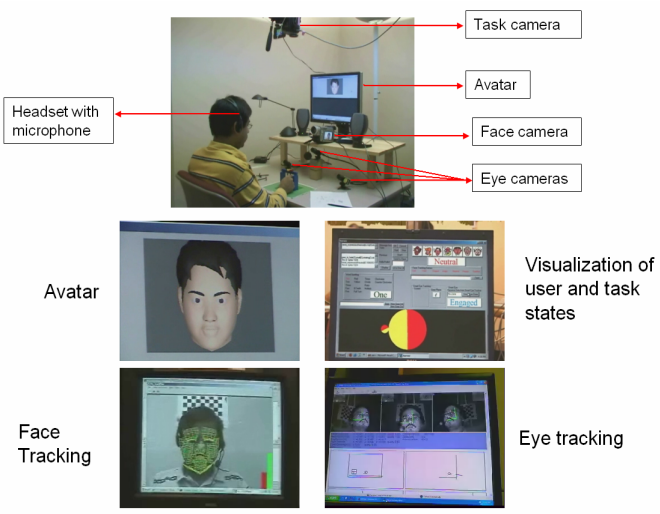

Fig. 1. A prototype of multimodal computer-aided learning system

Because of this practical importance and the theoretical interest of cognitive scientists, automatic human affect analysis has attracted the interest of many researchers. However, most of the existing approaches to automatic human affect analysis are uni-modal (e.g., visual-only or audio-only) approaches, based on deliberately displayed affective expressions, and aimed at prototypical (basic) emotions. Accordingly, the efforts toward uni-modal analysis of artificial affective expressions have been focused in the previously published survey papers [20], [55], [57], [58], [59], [61], [69], [75] among which the papers of Cowie et al. in 2001 [20] and of Pantic and Rothkrantz in 2003 [59] have been most comprehensive and widely cited in this field to date.

Due to the criticisms received from both cognitive and computer scientist that the existing methods for automatic human affect analysis are not applicable in real-life situations, where subtle changes in expressions typify the displayed affective behavior rather than the exaggerated changes that typify posed expressions, the focus of the research in the field has started to shift to automatic analysis of spontaneously displayed affective behavior, i.e., spontaneous facial expressions (e.g., [5], [15], [70], [78]) and audio expressions (e.g., [7], [46]). In addition, more and more researchers realize that integrating the information from audio and visual channels leads to an improved recognition of affective behavior occurring in real-world settings. As a result, an increased number of studies on audiovisual human affect recognition have emerged in recent years (e.g., [10], [30], [86]).

This paper introduces and surveys these recent advances in the research on human affect recognition. In contrast to those previous survey papers in the field, it focuses on the approaches that can handle audio and/or visual recordings of spontaneous (as opposed to posed) displays of affective states.

It is organized as follows. Section 2 describes human perception of affect from a psychological perspective. Section 3 provides a detailed review of related studies, specifically available audio/visual computing methods. Section 4 discusses the challenges in enhancing and extending these reviewed studies. A summary and closing remarks conclude the paper.

2. HUMAN AFFECT (EMOTION) PERCEPTION

Constructing an affect analyzer is dependent on our understanding of the nature of affect. This knowledge of affect includes the description of affect, and the association between observed signals (audio and visual signals in this paper) and affective states. There is no doubt that the progress in automatic affect recognition is in part contingent on the progress of psychologists' and linguists' understanding of human affect perception [26], [67].

2.1 The Description of Affect

Perhaps the most longstanding way that affect has been described by psychologists is in terms of discrete categories, an approach that is rooted in the language of daily life [20], [26], [67]. The most popular example of this description is the prototypical (basic) emotion categories, which include happiness, sadness, fear, anger, disgust, and surprise. The description of basic emotions was supported especially by the cross-cultural studies conducted by Ekman [23]. This influence of basic emotion theory resulted in the fact that most of existing studies of automatic affect recognition focus on recognizing these basic emotions. However, discrete lists of emotions fail to describe the range of emotions occurring in natural communication settings. In particular, basic emotions cover a rather small part of our daily emotional displays. Selection of affect categories that people show in daily interpersonal interactions needs to be done in a pragmatic and context-dependent manner.

An alternative to category description is the dimensional description [20], [32] where an affective state is represented as a point of a set of dimensions defined by psychological concepts. One of the popular methods to describe affective is in terms of dimensions of evaluation and activation [20]. The evaluation dimension measures how human feels, from positive to negative. The activation dimension measures whether humans are more or less likely to take an action under the emotional state, from active to passive. In contrast to category representation, dimensional representation enables raters to label a range of emotions. However, this projection of the high-dimensional emotional states onto a rudimentary 2D space results to some degree in the loss of information. Some emotions become indistinguishable (e.g., fear and anger) and some emotions lie outside the space (e.g., surprise). Some studies [33] use the additional dimension (e.g., dominance) to add discriminability of emotions.

2.2 Association Between Affects, Audio and Visual Signals

The face plays a significant role in human emotion perception and expression. The association between face and affective arousal was confirmed by a series of impressive and systematic studies in the field of psychology [26], [67].

Different from the traditional message judgment in which the aim is to infer what underlies a displayed behavior, such as affect or personality, another major approach to human behavior measurement is the sign judgment [15]. The aim of sign judgment is to describe the appearance rather than meaning of the shown behavior. While message judgment is focused on interpretation, sign judgment attempts to be objective, leaving the inference about the conveyed message to higher order decision making. The most commonly used sign judgment method used for manual labeling of facial behavior is the Facial Action Coding System (FACS) proposed by Ekman et al. [25]. FACS is a comprehensive and anatomically based system that is used to measure all visually discernible facial movements in terms of atomic facial actions called Action Units (AUs). These AUs can be used for any higher order decision making process including recognition of basic emotions according to Emotional FACS (EMFACS) rules[2] and a variety of affective states according to FACS Affect Interpretation Database (FACSAID)[2], as well as for recognition of other complex psychological states such as depression [27] or pain [49]. AUs of the FACS are very suitable to be used in studies on human naturalistic facial behavior as the thousands of anatomically possible facial expressions (independently of their higher-level interpretation) can be described as combinations of 27 basic AUs and a number of AU descriptors. It is not surprising, therefore, that an increasing number of studies on human spontaneous facial behavior aimed at automatic AU recognition (e.g., [5], [16], [78]).

Speech is another important communication device in human communication. It delivers affective information through explicit (linguistic) message, and implicit (paralinguistic) message that reflects the way the words are spoken. Although cognitive scientists have not identified the optimal set of vocal cues that reliably discriminate among affective and attitudinal states, listeners seem to be rather accurate in decoding some basic emotions from prosody [41] and some non-basic affective states such as distress, anxiety, boredom, and sexual interest from nonlinguistic vocalizations like laughs, cries, sighs, and yawns [67]. The basic-emotion-related prosodic features extracted from audio signal include pitch, energy, and speech rate. Cowie et al. [20] provided a comprehensive summary of qualitative acoustic correlations for prototypical emotions.

Linguistic content of speech definitely carries emotional information. Some of this information can be inferred directly from the surface features of words which were summarized in some affective word dictionaries and lexical affinity [80], [65]. The rest of this information lies below the text surface and can only be detected when the semantic context (e.g., discourse information) is taken into account. The association between linguistic content and emotion is language-dependent and generalizing from one language to another is very difficult to achieve.

A large number of studies in psychology and linguistics confirm the correlation between some affective displays (especially prototypical emotions) and specific audio and visual signals (e.g., [26], [67]). Ekman [24] found that the relative contributions of facial expression, speech and body cues to affect judgment depend both on the affective state and the environment where the affective behavior occurs. Many studies indicate that the human judgment agreement is typically higher for facial expression modality than it is for vocal expression modality. The amount of the agreement drops considerably when the stimuli are spontaneously displayed expressions of affective behavior rather than posed exaggerated displays. In addition, facial expression and vocal expression of emotion are often studied separately. This precludes finding evidence of the temporal correlation between them. On the other hand, a growing body of research in cognitive sciences argues that the dynamics of human behavior are crucial for its interpretation (e.g., [15], [27], [67]). For example, it has been shown that temporal dynamics of facial behavior represents a critical factor for distinction between spontaneous and posed facial behavior (e.g., [15], [27], [78]) as well as for categorization of complex behaviors like pain, shame, and amusement (e.g., [27]). Based on these findings, we may expect that temporal dynamics of each modality separately (facial and vocal) and temporal correlations between the two modalities play an important role in interpretation of human affective behavior. However, these are largely unexplored areas of research. Another unexplored area of research is that of context dependency. The interpretation of human behavioral signals is context dependent. For example a smile can be a display of politeness, irony, joy, or greeting. To interpret a behavioral signal, it is important to know the context in which this signal has been displayed – where the expresser is (e.g., inside, on the street, in the car), what his or her current task is, who the receiver is, and who the expresser is [67].

3. THE STATE OF THE ART

Rather than providing exhaustive coverage of all past efforts in the field of automatic recognition of human affect, we focus here on the efforts recently proposed in the literature that address the problem of automatic analysis of spontaneous affective behavior recorded in real-world settings. Keeping in mind the complexity of affective computing, we also briefly examine studies that represent exemplary approaches to treating a specific problem relevant for advancing human affect sensing technology.

For exhaustive surveys of the past efforts in the field, readers are referred to [20], [55], [57], [58], [59], [61], [69], [75].

This section is focused on an overview of the existing computing methods for automatic human affect recognition based on audio and/or visual displays. For the surveys of existing databases of spontaneous human affective behavior, the readers are referred to [18], [34], [62].

3.1 Facial Expression Recognition

The current research of facial expression recognition can be divided into two directions [15]: recognition of affect and recognition of facial muscle action (facial action units).

As far as automatic facial affect recognition is concerned, most of the existing efforts studied the expressions of the six basic emotions due to their universal properties, their marked reference representation in our affective lives, and the availability of the relevant training and test materials (e.g., [42]). There are a few tentative efforts to detect non-basic affective states from deliberately displayed facial expressions including fatigue [40], pain [49], and mental states like agreeing, concentrating, disagreeing, interest, frustration, thinking and unsure [28], [43], [82].

[2] http://face-and-emotion.com/dataface/general/homepage.jsp

Growing efforts are recently reported toward automatic analysis of spontaneous facial expression data [5], [6], [15], [16], [17], [39], [49], [50], [70], [78], [84]. Some of them study automatic recognition of AUs rather than emotions from spontaneous facial displays [5], [6], [15], [16], [78]. Several of these studies [17], [78] investigated the difference between spontaneous and deliberate facial behavior. The study [17] showed that many types of spontaneous smiles (e.g., polite) are smaller in amplitude, longer in total duration, and slower in onset and offset time than posed smiles. In addition, it has been shown in [78] that spontaneous brow actions (AU1, AU2 and AU4 in the FACS system) have different morphological and temporal characteristics (intensity, duration, and occurrence order) than posed brow actions.

The usually extracted facial features are either geometric features such as the shapes of the facial components (eyes, mouth, etc.) and the location of facial salient points (corners of the eyes, mouth, etc.) or appearance features representing the facial texture including wrinkles, bulges, and furrows. Typical examples of geometric-feature-based methods are those of Chang et al. [13], who used a shape model defined by 58 facial landmarks, and of Pantic and her colleagues [56], [60], [78], who used a set of facial characteristic points around the mouth, eyes, eyebrows, nose, and chin. Typical example of hybrid, geometric- and appearance-feature-based method, is that of Zhang and Ji [90], who used 26 facial points around the eyes, eyebrows, and mouth and the transient features like crow-feet wrinkles and nasal-labial furrows. Typical examples of appearance-feature-based methods are those of Bartlett et al. [5], [6] and Guo and Dyer [36], who used Gabor wavelets or eigenfaces, of Anderson and McOwen [1], who used a holistic spatial ratio face template, of Valstar et al. [77], who used temporal templates, and of Chang et al. [11], who built a probabilistic recognition algorithm based on the manifold subspace of aligned face appearances. An exemplar method of using both geometric and appearance features is that proposed by Lucey et al. [50], that uses Active Appearance Model (AAM) to capture the characteristics of the facial appearance and the shape of facial expressions.

Most of the existing 2D-feature-based methods are suitable for analysis of facial expressions under a small range of head motions. Thus, most of these methods focus on recognition of facial expressions in near-frontal-view recordings. An exception is the study of Pantic and Patras [56], who have explored automatic analysis of facial expressions from the profile-view of the face.

Few approaches to automatic facial expression analysis are based on 3D face models. Huang and his colleagues (i.e., [14], [70], [84]) used the geometry or appearance features extracted by a 3D face tracker called Piecewise Bezier Volume Deformation Tracker [74]. Cohn et al. [16] focused on analysis of brow action units and head movement based on a cylindrical head model [81]. Chang et al. [12] and Yin et al. [83] used 3D expression data for facial expression recognition. The progress of the methodology based on 3D face models may yield view-independent facial expression recognition, which is important for spontaneous facial expression recognition because the subject can be recorded in less controlled, real-world settings.

Relatively few studies investigated the fusion of the information from facial expressions and head movements [16], [40], [90], and the fusion of facial expression and body gesture [4], [35], [43], with the aim to improve affect recognition performance. Except for few studies, e.g., the studies [60], [29] that investigated interpretation of facial expressions in terms of user-defined interpretation labels, and the study [40] that investigated the influence of context (work condition, sleeping quality, circadian rhythm, and environment, physical condition) on fatigue detection, the existing automatic facial expression analyzers are context insensitive.

3.2 Audio Expression Recognition

Research on audio expression recognition is also influenced by basic emotion theory so that most of the existing efforts toward this direction chose the basic emotions or a subset of them as recognized targets. There are a few tentative studies that have investigated the detection of certain application-dependent affective states. Examples of these studies are those of Hirschberg et al. [37], who attempted deception detection, of Liscombe et al. [47], who focused on detecting certainness, Kwon et al. [45], who focused on detecting stress, of Zhang et al. [89], who focused on detecting confidence, confusion, and frustration, of Batliner et al. [7], who focused on detecting trouble, of Ang et al. [2], who focused on detecting annoyance and frustration, and of Steidl et al. [71], who conducted detection of motherese and empathy. More recently, few efforts towards automatic recognition of nonlinguistic vocalizations like laughters [76] and cries [54] have also been reported.

Some researchers started to turn their focus to investigation of spontaneous emotion recognition by using the audio data collected in call centers [46], [52], meetings [52], wizard of OZ [7] or other dialogue systems [8], [48]. In this natural interaction data, affective expressions are often subtle, and basic emotion expressions seldom occurred. Accordingly, these studies always chose to detect coarse affective states, i.e., positive, negative and neutral in [46], [52], [48], or application-dependent states as described above.

When the research shifts from posed emotion expression to spontaneous emotion expression, only acoustic information is not enough to detect the change of audio affective expression, as indicated by Batliner et al. [7] that "the closer we get to a realistic scenario, the less reliable is prosody as an indicator of the speakers emotional state". Thus, a few studies investigated the combination of acoustic features and linguistic features (language and discourse) to improve recognition performance. Typical examples of linguistic-paralinguistic-fusion methods are those of Litman et al. [48] and Schuller et al. [68], who used spoken words and acoustic features, of Lee and Narayanan [46], who used prosodic features, spoken words and information of repetition, and of Bartliner et al. [7], who used Part-of-speech (POS), dialogue act (DA), repetitions, corrections, and syntactic-prosodic boundary to infer the emotion. Litman et al. [48] investigated the role of the context information (e.g. subject, gender and problem, turn-level features representing local and global aspects of the prior dialogue) on audio affective recognition.

Although the above studies indicated recognition improvement by using information of language, discourse and context, automatic extraction of these related features is a difficult problem. First, existing automatic speech recognition systems cannot reliably recognize the verbal content of emotional speech [3]. Second how to extract semantic discourse information is more challenging. As

a result, most of these features have been extracted manually or directly from transcripts.

3.3 Audio-visual Expression Recognition

In the survey written by Pantic and Rothkrantz in 2003, [59], only four studies were found that were focused on audiovisual affect recognition. Since then, an increasing number of efforts are reported toward this direction. Although most of existing audio-visual affect recognition studies investigated recognition of basic emotions, fewer efforts are underway to detect non-basic emotion, i.e., those of Zeng et al. [85], [87], [88], who added 4 cognitive states (interest, puzzlement, frustration and boredom) considering the importance of these cognitive states in human computer interaction.

Recently a few studies have been reported toward audio-visual spontaneous emotion recognition [10], [30], [86]. These studies are that of Zeng et al. [86], who used the data collected in psychological research interview (Adult Attachment Interview), and of Fragopanagos and Taylor [30] and Caridakis et al. [10], who used the data collected in Wizard of OZ scenarios. Because their data were not sufficient to build classifiers for fine-grained affective states (e.g., basic emotions), they chose to recognize coarse affective states, e.g., positive and negative states in [86], or quadrants in evaluation-activation space [10], [30]. The studies [10], [30] applied the FeelTrace system that enables raters to continuously label the change of affective expressions. The study [30] noticed the considerable labeling variation among four raters using FeelTrace [19] due to subjectivity of audio-visual affect judgment. Specifically, one rater mainly relied on audio information to make judgment while another rater mainly relied on visual information. In order to reduce this variation, the studies [86] made the assumption that facial expression and vocal expression has the same coarse emotional states (positive and negative), and then directly used FACS-based labels of facial expressions as audio-visual expression labels.

Three fusion strategies (feature-level, decision-level and model-level fusions) are found to be used in the audio-visual affect recognition. A typical example of feature-level fusion is the study [9], which concatenated the prosodic features and facial features to construct joint feature vectors that are then used to build an affect recognizer. However, the different time scale and metric level of features from different modalities and increasing feature dimension influence the performance of the feature-level fusion. Most of the bimodal affect recognition studies applied decision-level fusion (e.g., [9], [31], [38], [79], [88]), which independently model audio-only and visual-only expressions, then combine these uni-modal recognition results at the end. Since humans display audio and visual expressions in a complementary and redundant manner, the conditional independent assumption of decision-level fusion actually loses the correlation information between audio and visual signals. Some interesting model-level fusion methods are introduced that can make use of the correlation between audio and visual streams, and relax the requirement of synchronization of these streams. Zeng et al. [87] presented Multi-stream Fused HMM to build an optimal connection among multiple streams from audio and visual channels according to maximum entropy and the maximum mutual information criterion. Zeng et al. [85] extended this fusion framework by introducing a middle-level training strategy under which a variety of learning schemes can be used to combine multiple component HMMs. Song et al. [73] presented tripled HMM to model correlation properties of three component HMMs that are based individually on upper face, lower face and prosodic dynamic behaviors. Fragopanagos and Taylor [30] proposed an artificial neural network with a feedback loop called ANNA to integrate the information from face, prosody and lexical content. Caridakis et al. [10] investigated combining face and prosody expressions by using Relevant Neural Networks.

4. CHALLENGES

The studies reviewed in the previous section indicate two new trends in the research on automatic human affect recognition: analysis of spontaneous affective behavior and multimodal analysis of human affective behavior including audiovisual analysis, combined linguistic and nonlinguistic analysis, and multi-cue visual analysis based on facial expressions, head movements, and/or body gestures. Several previously-recognized problems have been finally addressed. At the same time, several new challenging issues have been recognized, including the necessity of studying the temporal correlations between the different modalities (audio and visual) as well as between various behavioral cues (e.g., facial, head, and body gestures).

Here we focus on discussing the challenges in computing methods for developing of automatic spontaneous affect recognizer. As for the challenges to spontaneous emotion database collection and annotation, the readers are referred to [18], [21], [34], [59], [62].

4.1 Visual Input

Development of vision processing techniques that are robust in fully unconstrained environments is still in the relatively distant future. The existing visual face detection and tracking techniques are just able to reliably handle the near-front/profile view of face images with good resolution and lighting conditions. In a realistic interaction environment, the arbitrary movement of subjects, low-resolution and hand occlusion can cause these techniques to fail. The view-independent facial expression recognition based on 3D face model is worthy of further investigation [12], [83]. Development of a robust face detector, head and facial feature tracker forms the first step in the realization of facial expression analyzers capable of handling unconstrained environment.

In a realistic interaction environment, a facial expression analyzer should be able to deal with noisy and partial data and to generate its conclusion with confidence that reflects uncertainty of output of face and face point localization and tracking. Further efforts are needed toward modeling the static and dynamic structure of facial expression in order to handle noise features, temporal information, and partial data.

Except for few studies (e.g., [4], [16], [35], [40], [90]), the existing efforts analyzed facial expression behavior isolated from other visual cues (eye and head movement, and body gesture). It is suggested in the study [44] that multimodal coordination of facial expression, head movement and gesture is important to judge certain affect expression such as embarrassment. Integration of these multiple cues for automatic visual-based affect recognition is a largely unexplored research.

4.2 Audio Input

When our aim is to detect spontaneous emotion expressions, we have to take into account both linguistic and paralinguistic cues

that mingle together in audio channel. Although a number of linguistic and paralinguistic features (e.g. prosodic, dysfluency, lexicon, and discourse features) have been introduced for affect recognition in literature, the optimal feature set has not yet been established from the existing experiments.

Another challenge is how to reliably automatically extract these linguistic and paralinguistic feature from the audio channel. When we analyze the prosody in realistic conversation, we have to consider the multiple functions of prosody that include expression of affect and a variety of linguistic function [53]. Prosody features can be used to indicate discourse and segmentation information not only to express emotion. The prosodic event model that can reflect these functions simultaneously is worthy of further investigation. In addition, automatic extraction of spoken words from spontaneous emotional speech is also a difficult problem because the recognition rate of the exiting automatic speech recognition (ASR) system is far from perfect. The emotional aspects in speech further reduce ASR performance [3]. The automatic extraction of high-level underlying semantic linguistic information (e.g. dialogue act, repetitions, corrections, and syntactic information) is more challenging.

4.3 Fusion

Although the benefit of fusion (i.e., audio-visual fusion, linguistic and paralinguistic fusion, multi-visual-cue fusion from face, head and body gestures) for affect recognition is expected from engineering and psychological perspectives, our knowledge of how humans achieve this fusion is extremely limited. The neurological studies on fusion of sensory neurons [72] seem to more support early fusion (i.e., feature-level fusion) than late fusion (i.e., decision-level fusion). However, it is an open issue how to construct suitable joint feature vectors composed of features from different modalities with different time scales, different metric levels and different dynamic structures, based on existing methods. Due to these difficulties, most researchers choose decision-level fusion that simplifies the fusion problem by introducing the conditional dependent assumption. Model-level fusion or hybrid fusion that combines the benefits of both feature-level and decision-level fusion methods may be the best choice for this fusion problem. Based on existing knowledge and methods, how to model multimodal fusion is largely unexplored. A number of issues relevant to fusion require further investigation, such as the optimal level of integrating these different streams, the optimal function for the integration, as well as inclusion of suitable estimations of reliability of each stream.

4.4 Context

Investigation is clearly warranted to address how to make use of contextual information to improve the performance of affect recognition. Emotions are intimately related to a situation being experienced or imagined by human. Without context, human may misunderstand speaker's emotion expressions. Since the problem of context sensing is very difficult to solve, pragmatic approaches (e.g. activity- and user-profiled approaches) should be taken when learning the grammar of human affective behavior [57]. Yet, with the exception for a few studies (e.g., [29], [40], [48], [60]), virtually all existing approaches to machine analysis of human affect are context insensitive. Building a context model that includes person ID, gender, age, conversation topic, and workload need the help from other research field like face recognition, gender recognition, age recognition, topic detection, and task tracking.

4.5 Evaluation

Unfortunately, the diverse methods reviewed in this paper are difficult to compare because they are rarely tested on a common experimental condition (e.g., data and annotation). United efforts of different research communities are needed to address the evaluation of system performance based on a comprehensive, readily accessible benchmark database with annotation.

5. CONCLUSION

In the comprehensive survey written by Pantic and Rothkrantz in 2003 [59], almost all automatic affect recognition efforts were based small artificial emotion data, and only four studies were focused on audio-visual affect recognition. Since then, the picture has changed considerably. Increasing efforts are reported toward recognition of spontaneous affective expression by using audio and visual information and fusion methods. Some pilot studies have identified some problems that have been missed or avoided in uni-modal posed emotion recognition.

The shifts of perspective in affect recognition research, from uni-modal to multimodal and from posed emotion expression to spontaneous emotion expression, in turn highlight many challenges to our knowledge and existing techniques. Collaboration among related disciplines is certainly the most powerful means to advance our knowledge on the nature of affect, and in turn enhance automatic affect recognition performance.

6. Acknowledgment

The authors would like to thank reviewers for valuable comments on this paper. This paper is collaborative work. Thomas Huang is the lead of this team work but likes to be the last in the author list as usual. Zhihong Zeng wrote the first draft, Maja Pantic significantly improved it by rewriting it and offering important advices, and Glenn Roisman provided important comments and polished it. The work is supported in part by a Beckman Postdoctoral Fellowship and in part by National Science Foundation Grant CCF 04-26627.

7. REFERENCES

[1] Anderson K and McOwan P W (2006). A real-time automated system for recognition of human facial expressions. IEEE Trans. Systems, Man, and Cybernetics- Part B, Vol. 36, No. 1, 96-105

[2] Ang J, Dhillon R, Krupski A, et al. (2002). Prosody-based automatic detection of annoyance and frustration in human-computer dialog, ICSLP.

[3] Athanaselis T, Bakamidis S, Dologlou I, Cowie R, Douglas-Cowie E, Cox C (2005). ASR for emotional speech: Clarifying the issues and enhancing performance. Neural Networks, 18:437-444

[4] Balomenos, T., Raouzaiou, A., Ioannou, S., Drosopoulos, A., Karpouzis, K., Kollias, S. (2005). Emotion Analysis in Man-Machine Interaction Systems. Lecture Notes in Computer Science, vol. 3361, 318-328

[5] Bartlett, M.S., Littlewort, G., Frank, M., Lainscsek, C., Fasel, I., and Movellan, J.(2005), Recognizing Facial Expression: Machine Learning and Application to Spontaneous Behavior, IEEE International Conference on Computer Vision and Pattern Recognition, 568-573

[6] Bartlett M S, Littlewort G, Frank MG, Lainscsek C, Fasel I and Movellan J (2006). Fully automatic facial action recognition in

spontaneous behavior. Int. Conf. on Automatic Face and Gesture Recognition, 223-230

[7] Batliner A, Fischer K, Hubera R, Spilkera J and Noth E. (2003). How to find trouble in communication. Speech Communication, Vol. 40, 117-143.

[8] Blouin, C., and Maffiolo, V. (2005), "A study on the automatic detection and characterization of emotion in a voice service context", Interspeech, Lisbon, 469-472.

[9] Busso, C., Deng, Z., Yildirim, S., Bulut, M., Lee, C.M. et al. (2004), Analysis of Emotion Recognition using Facial Expressions, Speech and Multimodal Information.. Int. Conf. Multimodal Interfaces. 205-211

[10] Caridakis, G., Malatesta, L., Kessous, L., Amir, N., Paouzaiou, A. and Karpouzis, K.. (2006). Modeling Naturalistic Affective States via Facial and Vocal Expression Recognition. Int. Conf. on Multimodal Interfaces. 146-154

[11] Chang Y, Hu C, Turk, M (2004). Probabilistic expression analysis on manifolds. Proc. Computer Vision and Pattern Recognition, 2:520-527

[12] Chang Y, Vieira M, Turk M, and Velho L (2005), "Automatic 3D facial expression analysis in videos". Analysis and Modelling of Faces and Gestures, Proceedings. 3723, pp. 293-307.

[13] Chang Y, Hu C, Feris R and Turk M (2006). Manifold based analysis of facial expression. J. Image and Vision Computing, Vol. 24. No.6, 605-614

[14] Cohen, L., Sebe, N., Garg, A., Chen, L., and Huang, T. (2003), Facial expression recognition from video sequences: Temporal and static modeling, Computer Vision and Image Understanding, 91(1-2):160-187

[15] Cohn, J.F. (2006), Foundations of Human Computing: Facial Expression and Emotion, Int. Conf. on Multimodal Interfaces, 233-238

[16] Cohn JF, Reed LI, Ambadar Z, Xiao J, and Moriyama T. (2004). Automatic Analysis and recognition of brow actions and head motion in spontaneous facial behavior. Int. Conf. on Systems, Man & Cybernetics, 1, 610-616

[17] Cohn, J.F. and Schmidt, K.L.(2004). The timing of Facial Motion in Posed and Spontaneous Smiles, International Journal of Wavelets, Multiresolution and Information Processing, 2, 1-12

[18] Cowie R, Douglas-Cowie E and Cox C (2005). Beyond emotion archetypes: databases for emotion modeling using neural networks. Neural Networks, 18: 371-388

[19] Cowie, R., Douglas-Cowie, E., Savvidou, S., McMahon, E., Sawey, M., & Schröder, M. (2000). 'Feeltrace': an instrument for recording perceived emotion in real time. Proceedings of the ISCA Workshop on Speech and Emotion, 19–24

[20] Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., and Taylor, J.G. (2001), Emotion Recognition in Human-Computer Interaction, IEEE Signal Processing Magazine, January, 32-80

[21] Devillers L, Vidrascu L, and Lamel L (2005). Challenges in real-life emotion annotation and machine learning based detection. Neural Networks, 18: 407-422

[22] Duric, Z., Gray, W.D., Heishman, R., Li, F., Rosenfeld, A., Schoelles, M.J., Schunn, C., Wechsler, H. (2002). Integrating perceptual and cognitive modeling for adaptive and intelligent human-computer interaction. Proceedings of the IEEE, Vol. 90, No. 7, 1272-1289

[23] Ekman P. (1972). Universals and cultural differences in facial expressions of emotion. Nebr. Symp. Motiv. 1971, 207-283

[24] Ekman, P., editor (1982). Emotion in the human face. Cambridge University Press, New York, 2nd edition

[25] Ekman, P., Friesen, W.V., Hager, J.C. (2002). Facial Action Coding System. A Human Face, Salt Lake City, USA

[26] Ekman P. and Oster H. (1979). Facial expressions of emotion. Ann. Rev. Psychol. 1979, 30:527-554

[27] Ekman P. and Rosenberg E.L. (2005). What the face reveals: basic and applied studies of spontaneous expression using the facial action coding system. 2nd edition, Oxford University Press.

[28] El Kaliouby R and Robinson P (2004). Real-time Inference of complex mental states from facial expression and head gestures. Computer Vision and Pattern Recognition Workshop, Vol. 3, 154

[29] Fasel B, Monay F and Gatica-Perez D (2004). Latent semantic analysis of facial action codes for automatic facial expression recognition. ACM Int. Workshop on Multimedia Information Retrieval, 181-188

[30] Fragopanagos, F. and Taylor, J.G. (2005), Emotion recognition in human-computer interaction, Neural Networks, 18: 389-405

[31] Go H.J, Kwak KC, Lee DJ, and Chun MG. (2003). Emotion recognition from facial image and speech signal. Int. Conf. of the Society of Instrument and Control Engineers. 2890-2895

[32] Greenwald M, Cook E and Lang P. (1989). Affective judgment and psychophysiological response: dimensional covariation in the evaluation of pictorial stimuli. J. Psychophysiol. 3:51-64

[33] Grimm, M. and Kroschel, K. (2005). Evaluation of Natural Emotions Using Self Assessment Manikins, IEEE Workshop on Automatic Speech Recognition and Understanding, 381-383

[34] Gross, R. (2005).Face databases. In: Handbook of Face Recognition, Li S Z., Jain A.K., (Eds.), Springer, New York, USA, 301-328

[35] Gunes, H., Piccardi, M. (2005). Affect Recognition from Face and Body: Early Fusion vs. Late Fusion, In *Proc. Int'l Conf. Systems, Man and Cybernetics*, 3437- 3443

[36] Guo G and Dyer C R (2005). Learning from examples in the small sample case – face expression recognition. IEEE Trans. Systems, Man and Cybernetics – Part B, Vol.35, No.3, 477-488

[37] Hirschberg, J., Benus, S., Brenier, J.M., Enos, F., Friedman, S. (2005). Distinguishing Deceptive from Non-Deceptive Speech. Interspeech, 1833-1836

[38] Hoch, S., Althoff, F., McGlaun, G., Rigoll, G. (2005), Bimodal fusion of emotional data in an automotive environment, ICASSP, Vol. II, 1085-1088, 2005

[39] Ioannou, S., Raouzaiou, A., Tzouvaras, V., Mailis, T., Karpouzis, K., & Kollias, S. (2005). Emotion recognition through facial expression analysis based on a neurofuzzy method. Neural Networks: 18, 423-435.

[40] Ji Q, Lan P and Looney C (2006). A probabilistic framework for modeling and real-time monitoring human fatigue. IEEE SMC-Part A, Vol. 36, No.5, 862-875

[41] Juslin, P.N., Scherer, K.R. (2005). Vocal expression of affect. In The New Handbook of Methods in Nonverbal Behavior Research. Harrigan, J., Rosenthal, R., Scherer, K., Eds. Oxford University Press, Oxford, UK

[42] Kanade, T., Cohn, J., and Tian, Y. (2000), Comprehensive Database for Facial Expression Analysis, In Proceeding of International Conference on Face and Gesture Recognition, 46-53

[43] Kapoor, A., Burleson, W., and Picard, R. W. (2007), Automatic prediction of frustration. Int. Journal of Human-Computer Studies. Vol. 65(8), 724-736.

[44] Keltner D (1995). Signs of appeasement: evidence for the distinct displays of embarrassment, amusement and shame. Journal of Personality and Social Psychology, 68(3). 441-454

[45] Kwon, O.W., Chan, K., Hao, J., Lee, T.W (2003), Emotion Recognition by Speech Signals, EUROSPEECH.

[46] Lee C M Narayanan, S.S. (2005). Toward detecting emotions in spoken dialogs. IEEE Tran. Speech and Audio Processing, Vol. 13(2): 293-303

[47] Liscombe, J., Hirschberg, J., Venditti, J.J. (2005). Detecting Certainness in Spoken Tutorial Dialogues. Interspeech.

[48] Litman, D.J. and Forbes-Riley, K. (2004), Predicting Student Emotions in Computer-Human Tutoring Dialogues. In Proc. of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL), July

[49] Littlewort G, Bartlett M S and Lee K (2006). Faces of Pain: Automated measurement of spontaneous facial expressions of

genuine and posed pain, 13 Joint Symposium on Neural Computation. 1
[50] Lucey, S., Ashraf, A.B., and Cohn, J.F. (2007), Investigating Spontaneous Facial Action Recognition through AAM Representations of the Face. In Face Recognition, Delac, K. and Grgic, M., Eds. Vienna, Austria: I-Tech Education and Publishing, 275-286
[51] Maat, L., Pantic, M. (2006). Gaze-X: Adaptive affective multimodal interface for single-user office scenarios, Proc. ACM Int'l Conf. Multimodal Interfaces, 171-178
[52] Neiberg D, Elenius K, and Laskowski K. (2006). Emotion Recognition in Spontaneous Speech Using GMM. Int. Conf. on Spoken Language Processing, 809-812
[53] Mozziconacci, S. (2002). Prosody and Emotions. Int. Conf. on Speech Prosody.
[54] Pal P, Iyer A N and Yantorno R E (2006). Emotion detection from infant facial expressions and cries. In Proc. Int'l Conf. Acoustics, Speech & Signal Processing, 2, pp. 721-724, 2006.
[55] Pantic, M., and Bartlett, M.S. (2007). Machine analysis of facial expres-sions. In Face Recognition, Delac, K. and Grgic, M., Eds. Vienna, Austria: I-Tech Education and Publishing, 377-416
[56] Pantic M and Patras T (2006). Dynamics of facial expression: recognition of facial actions and their temporal segments form face profile imgae sequences. IEEE Trans. Systems, Man and Cybernetics – Part B, Vol. 36, No.2, 433-449
[57] Pantic, M., Pentland, A., Nijholt, A., and Huang, T.S. (2006), Human Computing and Machine Understanding of Human Behavior: A Survey, Int. Conf. on Multimodal Interfaces, 239-248
[58] Pantic M and Rothkrantz L J M (2000). Automatic analysis of facial expressions—the state of the art. IEEE PAMI, Vol.22, No.12, 1424-1445
[59] Pantic M., Rothkrantz, L.J.M. (2003), Toward an affect-sensitive multimodal human-computer interaction, Proceedings of the IEEE, Vol. 91, No. 9, Sept., 1370-1390
[60] Pantic M and Rothkrantz L J M (2004). Case-based reasoning for user-profiled recognition of emotions from face images. Int. Conf. Multimedia & Expo, 391-394
[61] Pantic, M., Sebe, N., Cohn, J.F., and Huang, T.S. (2005), Affective Multimodal Human-Computer Interaction, ACM Multimedia, 669-676
[62] Pantic, M., Valstar, M.F, Rademaker, R. and Maat, L. (2005), Web-based database for facial expression analysis, Int. Conf. on Multimedia and Expo, 317–321
[63] Pentland, A. (2005). Socially aware, computation and communication, IEEE Computer, Vol.38, 33-40
[64] Picard, R.W. (1997). Affective Computing, MIT Press, Cambridge.
[65] Plutchik R. (1980). Emotion: A psychoevolutionary synthesis. New York: Harper and Row.
[66] Roisman, G.I., Tsai, J.L., Chiang, K.S.(2004). The Emotional Integration of Childhood Experience: Physiological, Facial Expressive, and Self-reported Emotional Response During the Adult Attachment Interview, Developmental Psychology, Vol. 40, No. 5, 776-789
[67] Russell J.A., Bachorowski J. and Fernandez-Dols J. (2003). Facial and vocal expressions of emotion. Ann. Rev. Psychol. 54:329-349
[68] Schuller, B., Villar, R. J., Rigoll, G., Lang, M. (2005). Meta-Classifiers in acoustic and linguistic feature fusion-based affect recognition. Int. Conf. on Acoustics, Speech, and Signal Processing, 325-328
[69] Sebe, N., Cohen, I., and Huang, T.S. (2005). Multimodal Emotion Recognition, Handbook of Pattern Recognition and Computer Vision, World Scientific, 2005.
[70] Sebe, N., Lew, M.S., Cohen, I., Sun, Y., Gevers, T., Huang, T.S.(2004), Authentic Facial Expression Analysis, Int. Conf. on Automatic Face and Gesture Recognition
[71] Steidl, S., Levit, M., Batliner, A, Noth, E., and Niemann, H. (2005), "Off all things the measure is man" Automatic classification of emotions and inter-labeler consistency, ICASSP, vol.1, 317-320
[72] Stein, B., Meredith, M.A. (1993). The Merging of Senses. MIT Press, Cambridge, USA
[73] Song, M., Bu, J., Chen, C., and Li, N. (2004), Audio-visual based emotion recognition—A new approach, Int. Conf. Computer Vision and Pattern Recognition. 2004, 1020-1025
[74] Tao, H. and Huang, T.S. (1999), Explanation-based facial motion tracking using a piecewise Bezier volume deformation mode, IEEE CVPR, vol.1, pp. 611-617,
[75] Tian Y L, Kanade T and Cohn J F (2005). Facial expression analysis. In: Handbook of Face Recognition, Li S Z and Jain A K (Eds.), Springer, New York, USA, 247-276
[76] Truong K P and van Leeuwen D A (2005). Automatic detection of laughter. In Proc. Interspeech, pp. 485-488, 2005.
[77] Valstar M, Pantic M and Patras I (2004). Motion history for facial action detection from face video. Int. Conf. Systems, Man and Cybernetics, Vol.1, 635-640
[78] Valstar MF, Pantic M, Ambadar Z and Cohn JF. (2006). Spontaneous vs. Posed Facial Behavior: Automatic Analysis of Brow Actions. Int. Conf. on Multimedia Interfaces. 162-170
[79] Wang, Y. and Guan, L.(2005), Recognizing human emotion from audiovisual information, ICASSP, Vol. II, 1125-1128
[80] Whissell C M (1989). The dictionary of affect in language. In Plutchik R. and Kellerman H (Eds.). Emotion: Theory, research and experience. The measurement of emotions, Vol.4. 113-131. New York: Academic Press
[81] Xiao J, Moriyama T, Kanade T and Cohn J F (2003). Robust full-motion recovery of head by dynamic templates and re-registration techniques. Int. J. Imaging Systems and Technology, Vol. 13, No.1, 85-94
[82] Yeasin M., Bullot B. and Sharma R. (2006), Recognition of facial expressions and measurement of levels of interest from video, IEEE Trans. On Multimedia, Vol.8, No. 3, June, 500-507
[83] Yin L, Wei X , Sun Y, Wang J, Rosato M J (2006). A 3D facial expression database for facial behavior research. Int. Conf. on Automatic Face and Gesture Recognition, 211- 216
[84] Zeng, Z., Fu, Y., Roisman, G.I., Wen, Z., Hu, Y., and Huang, T.S. (2006). Spontaneous Emotional Facial Expression Detection. Journal of Multimedia, 1(5): 1-8.
[85] Zeng, Z., Hu, Y., Liu, M., Fu, Y. and Huang, T.S.(2006), Training Combination Strategy of Multi-stream Fused Hidden Markov Model for Audio-visual Affect Recognition, in Proc. ACM Int'l Conf. on Multimedia, 2006, 65-68
[86] Zeng Z, Hu Y, Roisman G I, Wen Z, Fu Y and Huang T S (2006): Audio-visual emotion recognition in adult attachment interview. Int. Conf. Multimodal Interfaces: 139-145
[87] Zeng, Z., Tu, J., Pianfetti , P., Liu, M., Zhang, T., Zhang Z., Huang T S and Levinson S (2005), Audio-visual Affect Recognition through Multi-stream Fused HMM for HCI, Int. Conf. Computer Vision and Pattern Recognition. 967-972
[88] Zeng, Z., Tu, J., Liu, M., Huang, T.S., Pianfetti, B., Roth D. and Levinson, S. (2007), Audio-visual Affect Recognition, IEEE Transactions on Multimedia, Vol. 9, No. 2, February, 424-428
[89] Zhang T, Hasegawa-Johnson M and Levinson S E (2004). Children's Emotion Recognition in an Intelligent Tutoring Scenario, Interspeech 2004.
[90] Zhang Y and Ji Q (2005). Active and Dynamic Information Fusion for Facial Expression Understanding from Image Sequences. IEEE Trans. Pattern Anal. Mach. Intell. 27(5): 699-714

Journal of Personality and Social Psychology
1988, Vol. 54, No. 6, 1063-1070

0022-3514/88/$00.75

Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales

David Watson and Lee Anna Clark
Southern Methodist University

Auke Tellegen
University of Minnesota

In recent studies of the structure of affect, positive and negative affect have consistently emerged as two dominant and relatively independent dimensions. A number of mood scales have been created to measure these factors; however, many existing measures are inadequate, showing low reliability or poor convergent or discriminant validity. To fill the need for reliable and valid Positive Affect and Negative Affect scales that are also brief and easy to administer, we developed two 10-item mood scales that comprise the Positive and Negative Affect Schedule (PANAS). The scales are shown to be highly internally consistent, largely uncorrelated, and stable at appropriate levels over a 2-month time period. Normative data and factorial and external evidence of convergent and discriminant validity for the scales are also presented.

Two dominant dimensions consistently emerge in studies of affective structure, both in the United States and in a number of other cultures. They appear as the first two factors in factor analyses of self-rated mood and as the first two dimensions in multidimensional scalings of facial expressions or mood terms (Diener, Larsen, Levine, & Emmons, 1985; Russell, 1980, 1983; Stone, 1981; Watson, Clark, & Tellegen, 1984; Zevon & Tellegen, 1982).

Watson and Tellegen (1985) have summarized the relevant evidence and presented a basic, consensual two-factor model. Whereas some investigators work with the unrotated dimensions (typically labeled *pleasantness-unpleasantness* and *arousal*), the varimax-rotated factors—usually called Positive Affect and Negative Affect—have been used more extensively in the self-report mood literature; they are the focus of this article. Although the terms *Positive Affect* and *Negative Affect* might suggest that these two mood factors are opposites (that is, strongly negatively correlated), they have in fact emerged as highly distinctive dimensions that can be meaningfully represented as orthogonal dimensions in factor analytic studies of affect.

Briefly, Positive Affect (PA) reflects the extent to which a person feels enthusiastic, active, and alert. High PA is a state of high energy, full concentration, and pleasurable engagement, whereas low PA is characterized by sadness and lethargy. In contrast, Negative Affect (NA) is a general dimension of subjective distress and unpleasurable engagement that subsumes a variety of aversive mood states, including anger, contempt, disgust, guilt, fear, and nervousness, with low NA being a state of calmness and serenity. These two factors represent affective state dimensions, but Tellegen (1985; see also Watson & Clark, 1984) has demonstrated that they are related to corresponding affective trait dimensions of positive and negative emotionality (individual differences in positive and negative emotional reactivity). Trait PA and NA roughly correspond to the dominant personality factors of extraversion and anxiety/neuroticism, respectively (Tellegen, 1985; Watson & Clark, 1984). Drawing on these and other findings, Tellegen has linked trait NA and PA, respectively, to psychobiological and psychodynamic constructs of sensitivity to signals of reward and punishment. He has also suggested that low PA and high NA (both state and trait) are major distinguishing features of depression and anxiety, respectively (Tellegen, 1985; see also Hall, 1977).

Numerous PA and NA scales have been developed and studied in a variety of research areas. Generally speaking, the findings from these studies indicate that the two mood factors relate to different classes of variables. NA—but not PA—is related to self-reported stress and (poor) coping (Clark & Watson, 1986; Kanner, Coyne, Schaefer, & Lazarus, 1981; Wills, 1986), health complaints (Beiser, 1974; Bradburn, 1969; Tessler & Mechanic, 1978; Watson & Pennebaker, in press), and frequency of unpleasant events (Stone, 1981; Warr, Barter, & Brownbridge, 1983). In contrast, PA—but not NA—is related to social activity and satisfaction and to the frequency of pleasant events (Beiser, 1974; Bradburn, 1969; Clark & Watson, 1986, 1988; Watson, 1988).

Anomalous and inconsistent findings have also been reported, however. For example, whereas most studies have found these NA and PA scales to have low or nonsignificant correlations with one another (e.g., Clark & Watson, 1986, 1988; Harding, 1982; Moriwaki, 1974; Warr, 1978; Wills, 1986), others have found them to be substantially related (Brenner, 1975; Diener & Emmons, 1984; Kammann, Christie, Irwin, & Dixon, 1979). There are many possible explanations for such inconsistencies (e.g., see Diener & Emmons, 1984), but one that must be considered concerns the various scales themselves. It

We wish to thank Lisa Binz, Sondra Brumbelow, Richard Cole, Mary Dieffenwierth, Robert Folger, Jay Leeka, Curt McIntyre, James Pennebaker, and Karen Schneider for their help in collecting the data reported in this article.

Correspondence should be addressed to David Watson, Department of Psychology, Southern Methodist University, Dallas, Texas, 75275.

may be, for example, that some scales are simply better, purer measures of the underlying factors than are others. Watson (in press) reported evidence supporting this idea. He found that some scale pairs (such as those used by Diener and his associates in a number of studies; e.g., Diener & Emmons, 1984; Diener & Iran-Nejad, 1986; Diener et al., 1985) yield consistently higher NA–PA correlations than do others (such as our own scales, to be described shortly).

More generally, one must question the reliability and validity of many of these measures. Some mood scales have been developed through factor analysis (e.g., Stone, 1981), but others have been constructed on a purely ad hoc basis with no supporting reliability or validity data (e.g., McAdams & Constantian, 1983). Watson (in press) analyzed the psychometric properties of several popular measures and found many of them to be wanting, at least for use in student populations. For example, Bradburn's (1969) widely used NA and PA scales were unreliable (coefficient α = .52 for NA, .54 for PA) and only moderately related to other measures of the same factor (for NA, the convergent correlations ranged from .39 to .52; for PA, they ranged from .41 to .53). The short PA and NA scales used by Stone and his colleagues (Hedges, Jandorf, & Stone, 1985; Stone, 1987; Stone, Hedges, Neale, & Satin, 1985) were also unreliable (in two samples, the NA scale had coefficient αs of .48 and .52, whereas the PA scale had corresponding values of .64 and .70).

Clearly there is a need for reliable and valid PA and NA scales that are also brief and easy to administer. In this article we describe the development of such scales, the 10-item NA and PA scales that comprise the Positive and Negative Affect Schedule (PANAS), and present reliability and validity evidence to support their use.

Development of the PANAS Scales

Much of our previous mood research has been concerned with identifying these dominant dimensions of affect and clarifying their nature (Clark & Watson, 1986, 1988; Tellegen, 1985; Watson, in press; Watson & Clark, 1984; Watson et al., 1984; Watson & Tellegen, 1985; Zevon & Tellegen, 1982). To have a broad and representative sample of mood descriptors, we have used questionnaires that contained a large number (57–65) of mood terms. Once the basic NA and PA factors were clearly identified, however, we wanted to measure them more simply and economically. We therefore turned our attention to the development of brief PA and NA scales.

Our greatest concern was to select terms that were relatively pure markers of either PA or NA; that is, terms that had a substantial loading on one factor but a near-zero loading on the other. As a starting point, we used the 60 terms included in the factor analyses reported by Zevon and Tellegen (1982). This sample of descriptors was constructed by selecting three terms from each of 20 content categories; for example, the terms *guilty, ashamed,* and *blameworthy* comprise the guilty category (see Zevon & Tellegen, 1982, Table 1). The categories were identified through a principal-components analysis of content sortings of a large sample of descriptors and provide a comprehensive sample of the affective lexicon.

From this list we selected those terms that had an average loading of .40 or greater on the relevant factor across both the R- and P-analyses reported in Zevon & Tellegen (1982). Twenty PA markers and 30 NA markers met this initial criterion. However, as noted previously, we were also concerned that the terms not have strong secondary loadings on the other factor. We therefore specified that a term could not have a secondary loading of |.25| or greater in either analysis. This reduced the pool of candidate descriptors to 12 for PA and 25 for NA.

Preliminary reliability analyses convinced us that 10 terms were sufficient for the PANAS PA scale; we therefore dropped 2 terms (*delighted* and *healthy*) that had relatively high secondary loadings on NA. This yielded the final list of 10 descriptors for the PA scale: *attentive, interested, alert, excited, enthusiastic, inspired, proud, determined, strong* and *active.*

The 25 NA candidate terms included all 3 terms from seven of the content categories (distressed, angry, contempt, revulsion, fearful, guilty, and jittery) and 2 from each of two others (rejected and angry at self). Because we wanted to tap a broad range of content, we constructed a preliminary 14-item scale that included 2 terms from each of the seven complete triads. We found, however, that the contempt and revulsion terms did not significantly enhance the reliability and validity of the scale. Moreover, these terms were less salient to our subjects and were occasionally left unanswered. We therefore settled on a final 10-item version that consisted of 2 terms from each of the other five triads: *distressed, upset* (distressed); *hostile, irritable* (angry); *scared, afraid* (fearful); *ashamed, guilty* (guilty); and *nervous, jittery* (jittery). The final version of PANAS is given in the Appendix.

Reliability and Validity of the PANAS Scales

Subjects and Measures

The basic psychometric data were gathered primarily from undergraduates enrolled in various psychology courses at Southern Methodist University (SMU), a private southwestern university. The students participated in return for extra course credit. In addition, groups of SMU employees completed questionnaires asking how they felt "during the past few weeks" (n = 164) and "during the past few days" (n = 50). A sample of 53 adults not affiliated with SMU also filled out a mood form with "today" time instructions. Preliminary analyses revealed no systematic differences between student and nonstudent responses, and they have been combined in all analyses. Nevertheless, because most of our data were collected from college students, it is important to establish that the PANAS scales also work reasonably well in adult and clinical samples. We briefly address this issue in a later section.

The mood questionnaire consisted of a single page with the 60 Zevon and Tellegen (1982) descriptors arrayed in various orders. The subjects were asked to rate on a 5-point scale the extent to which they had experienced each mood state during a specified time frame. The points of the scale were labeled *very slightly or not at all, a little, moderately, quite a bit,* and *very much,* respectively. The PANAS terms were randomly distributed throughout the questionnaire. It is important to note that we have since used the 20 PANAS descriptors without these additional terms and obtained essentially identical results (Clark & Watson, 1986; Watson, 1988).

We obtained ratings with seven different temporal instruc-

Table 1
Positive and Negative Affect Schedule (PANAS) Scale Means and Standard Deviations for Each Rated Time Frame

Time instructions	*n*	PANAS PA Scale		PANAS NA Scale	
		M	*SD*	*M*	*SD*
Moment	660	29.7	7.9	14.8	5.4
Today	657	29.1	8.3	16.3	6.4
Past few days	1,002	33.3	7.2	17.4	6.2
Past few weeks	586	32.0	7.0	19.5	7.0
Year	649	36.2	6.3	22.1	6.4
General	663	35.0	6.4	18.1	5.9

Note. PA = Positive Affect. NA = Negative Affect.

tions. Subjects were asked to rate how they felt (a) "right now (that is, at the present moment)" (*moment* instructions); (b) "today" (*today*); (c) "during the past few days" (*past few days*); (d) "during the past week" (*week*); (e) "during the past few weeks" (*past few weeks*); (f) "during the past year" (*year*); and (g) "in general, that is, on the average" (*general*). For six of these time frames, we collected data on large samples to be used for normative, internal consistency, and factor analyses. The *n*s are 660 (moment), 657 (today), 1,002 (past few days), 586 (past few weeks), 649 (year), and 663 (general). These samples are largely but not completely independent: Some subjects completed mood forms involving two or more different temporal instructions; such multiple ratings were always spaced at least 1 week apart. In addition, a subset of these subjects (n = 101) completed ratings on all seven time frames on two different occasions, providing retest data.

Normative and Reliability Data

Basic scale data. Table 1 presents basic descriptive data on the PANAS PA and NA scales for the various time instructions. Given the large sample sizes, these provide reasonably good college student norms. In our data, we have not found any large or consistent sex differences, so the data are collapsed across sex. Nevertheless, it seems advisable to test for sex differences in any new (especially nonstudent) sample.

Inspecting Table 1, one sees that subjects report more PA than NA, regardless of the time frame. Moreover, mean scores on both scales tend to increase as the rated time frame lengthens. This pattern is expectable: As the rated time period increases, the probability that a subject will have experienced a significant amount of a given affect also increases.

The PANAS scale intercorrelations and internal consistency reliabilities (Cronbach's coefficient α) are reported in Table 2. The alpha reliabilities are all acceptably high, ranging from .86 to .90 for PA and from .84 to .87 for NA. The reliability of the scales is clearly unaffected by the time instructions used.

The correlation between the NA and PA scales is invariably low, ranging from −.12 to −.23; thus, the two scales share approximately 1% to 5% of their variance. These discriminant values indicate quasi-independence, an attractive feature for many purposes, and are substantially lower than those of many other short PA and NA scales (see Watson, in press). Interestingly, our PA–NA correlation was unaffected by the rated time frame, whereas Diener and Emmons (1984) found that the correlation between their PA and NA scales decreased as the rated time frame lengthened. However, this discrepancy is beyond the scope of our article; see Watson (in press) for a detailed discussion of the effects of different temporal instructions on various mood scales.

Test–retest reliability. As noted previously, 101 SMU undergraduates filled out PANAS ratings for each of the seven time frames on two different occasions. The mood ratings were collected at weekly intervals. The first set of ratings was collected during Weeks 1–7 of the fall 1986 semester in the following order: year, past few days, today, past few weeks, general, moment, and week. Then, following a 1-week break, the PANAS scales were readministered during Weeks 9–15 in the same sequence. Thus, each scale was retested after an 8-week interval.

These reliability data are shown in Table 3. The NA and PA stability values were first compared at each rated time frame and no significant differences were found ($p > .05$, 2-tailed t test). Multiple comparisons were then made across the time frames for each affect separately ($p < .002$, Bonferroni corrected for 21 comparisons). Not surprisingly, the retest stability tends to increase as the rated time frame lengthens. Ratings of longer time periods, such as how one has felt during the past few weeks or the past year, are implicit aggregations. In a sense, subjects average their responses over a longer time frame and hence over more occasions. Thus, these data replicate the frequent finding that stability rises with increasing temporal aggregation (e.g., Diener & Larsen, 1984; Epstein, 1979). The stability coefficients of the general ratings are high enough to suggest that they may in fact be used as trait measures of affect.

It is also noteworthy that the PANAS scales exhibit a significant level of stability in every time frame, even in the moment ratings. These results are also consistent with earlier findings (e.g., Watson & Clark, 1984, Table 8) and reflect the strong dispositional component of affect. That is, even momentary moods are, to a certain extent, reflections of one's general affective level (Costa & McCrae, 1980; Watson & Clark, 1984).

Generalizability to nonstudent samples. Our largest nonstudent sample consisted of 164 SMU employees who rated how they had felt during the past few weeks. A separate analysis of this sample yielded results comparable with the values listed in

Table 2
Internal Consistency Reliabilities (Coefficient Alpha) and Scale Intercorrelations

Time instructions	*n*	Alpha reliabilities		PA–NA intercorrelation
		PANAS PA scale	PANAS NA scale	
Moment	660	.89	.85	−.15
Today	657	.90	.87	−.12
Past few days	1,002	.88	.85	−.22
Past few weeks	586	.87	.87	−.22
Year	649	.86	.84	−.23
General	663	.88	.87	−.17

Note. PANAS = Positive and Negative Affect Schedule. PA = Positive Affect. NA = Negative Affect.

Table 3
Test–Retest Reliabilities of the Positive and Negative Affect Schedule (PANAS) Scales (8-Week Retest Interval)

Time instructions	PANAS PA scale	PANAS NA scale
Moment	.54[ab]	.45[b]
Today	.47[b]	.39[b]
Past few days	.48[b]	.42[b]
Past week	.47[b]	.47[b]
Past few weeks	.58[ab]	.48[b]
Year	.63[ab]	.60[ab]
General	.68[a]	.71[a]

Note. $n = 101$. Coefficients not sharing the same superscript are different at $p < .05$ (two-tailed, Bonferroni corrected for multiple comparisons). PA = Positive Affect. NA = Negative Affect. Significance tests are computed separately for each scale. See text for further details.

Table 2. Specifically, the alpha reliabilities of the PANAS PA and NA scales were .86 and .87, respectively, and the correlation between the scales was −.09. Given these data, we believe that the PANAS scales will provide useful information in adult samples as well, although further data are desirable to establish this fully.

We have also collected data on a small ($n = 61$) psychiatric inpatient sample using the general instructions. Again, the PANAS scales were reliable (for PA, $\alpha = .85$; for NA, $\alpha = .91$) and only moderately intercorrelated with one another ($r = -.27$). Given the small sample size, these data cannot be considered definitive, but they are encouraging and suggest that the PANAS scales retain their reliability and quasi-independence in clinical samples. In addition, all but four of the patients retook the measure after a 1-week interval, and the resulting stability analyses yielded high test–retest reliabilities: .81 for NA and .79 for PA. Finally, consistent with previous studies (Watson & Clark, 1984), we found significant group differences for NA, with the patients considerably higher ($M = 26.6$) and more variable ($SD = 9.2$) than the normative group ($M = 18.1$, $SD = 5.9$; see Table 1). The corresponding differences for PA (patient group $M = 32.5$, $SD = 7.5$; normative group $M = 35.0$, $SD = 6.4$) were also statistically significant because of the very large n of the normative sample, but it would be premature to accept a mean scale difference of 2.5 points as clinically meaningful without further study.

Factorial Validity

Scale validity. An important step in evaluating the PANAS scales is to demonstrate that they adequately capture the underlying mood factors. To do this, we subjected ratings on the 60 Zevon and Tellegen (1982) mood descriptors in each of the six large data sets to a principal factor analysis with squared multiple correlations as the communality estimates. Two dominant factors emerged in each solution. Together, they accounted for roughly two thirds of the common variance, ranging from 62.8% in the moment solution to 68.7% in the general ratings. The first two factors in each solution were then rotated to orthogonal simple structure according to the varimax criterion.

Each of the six solutions generated two sets of factor scoring weights that can be used to compute regression estimates of the underlying PA and NA factors in those data. Within each data set, we then correlated these estimated factor scores with the PANAS PA and NA scales. The results, shown in Table 4, demonstrate the expected convergent/discriminant pattern: Both PANAS scales are very highly correlated with their corresponding regression-based factor scores in each solution, with convergent correlations ranging from .89 to .95, whereas the discriminant correlations are quite low, ranging from −.02 to −.18.

Item validity. It is also important to demonstrate the factorial validity of the individual PANAS items. To do this, we factored subjects' ratings on the 20 PANAS descriptors in each of the six data sets; as before, we used a principal factor analysis with squared multiple correlations as the initial communality estimates. Because the PANAS terms were selected to be relatively pure factor markers, it is not surprising that two dimensions accounted for virtually all of the common variance in these solutions (ranging from 87.4% in the moment data to 96.1% in the general ratings).

Median varimax loadings for the PANAS terms on these two factors are presented in Table 5. All of the descriptors have strong primary loadings (.50 and above) on the appropriate factor, and the secondary loadings are all acceptably low. Thus, all of the PANAS items are good markers of their corresponding factors.

Rating scale effects. The data shown in Tables 1 through 5 are all based on the same 5-point rating scale. Because the subjects were instructed to rate the extent to which they experienced each mood state, this may be termed an extent format. It seems reasonable to ask, however, whether different response formats might yield different results. Warr et al. (1983) have presented data indicating that the correlation between PA and NA scales varies according to the response scale used. Specifically, their PA and NA scales were highly correlated when they used a frequency-type format in which subjects rated the proportion of time they had experienced each mood state during a specified time period.

To test the effect of rating format, we collected ratings on 54 mood terms in two different student samples, both using past few weeks time instructions. In the first sample, 413 subjects rated their mood using the usual extent rating format. In the second, 338 students rated themselves on a 4-point frequency

Table 4
Correlations Between the Positive and Negative Affect Schedule (PANAS) Scales and Scores of the First Two Varimax Factors in Each Sample

Time instructions	n	PANAS PA scale correlations		PANAS NA scale correlations	
		Factor 1	Factor 2	Factor 1	Factor 2
Moment	660	−.02	.95	.91	−.15
Today	657	−.02	.95	.93	−.11
Past few days	1,002	−.15	.92	.93	−.10
Past few weeks	586	−.10	.92	.92	−.18
Year	649	−.17	.89	.93	−.09
General	663	−.08	.94	.93	−.12

Note. Factor analyses are based on the set of 60 mood terms reported in Zevon & Tellegen (1982). PA = Positive Affect. NA = Negative Affect.

Table 5
Median Varimax-Rotated Factor Loadings of the Positive and Negative Affect Schedule (PANAS) Descriptors Across the Six Solutions

PANAS descriptor	Loading on Positive Affect	Loading on Negative Affect
Enthusiastic	.75	−.12
Interested	.73	−.07
Determined	.70	−.01
Excited	.68	.00
Inspired	.67	−.02
Alert	.63	−.10
Active	.61	−.07
Strong	.60	−.15
Proud	.57	−.10
Attentive	.52	−.05
Scared	.01	.74
Afraid	.01	.70
Upset	−.12	.67
Distressed	−.16	.67
Jittery	.00	.60
Nervous	−.04	.60
Ashamed	−.12	.59
Guilty	−.06	.55
Irritable	−.14	.55
Hostile	−.07	.52

format (the options were *little or none of the time, some of the time, a good part of the time,* and *most of the time*).

In addition to the PANAS terms, the mood descriptors used in these samples allowed us to compare the factorial validity of our scales with those of other investigators. In both samples, we were able to measure the brief NA and PA scales developed by Diener and Emmons (1984, Studies 3 through 5; see also Diener & Iran-Nejad, 1986; Diener & Larsen, 1984; Diener et al., 1985), Stone and his associates (Hedges et al., 1985; Stone, 1987; Stone et al., 1985), and McAdams and Constantian (1983). Further, in the extent sample, 301 subjects rated themselves on Bradburn's (1969) widely used NA and PA scales; these were replaced by Warr et al.'s (1983) revised measures in the frequency sample.

The ratings in each sample were subjected to separate principal factor analyses with squared multiple correlations in the diagonal (these analyses are reported in detail in Watson, in press). Two large factors emerged in each solution, accounting for 75.4% and 73.3% of the common variance in the extent and frequency data, respectively. The first two factors in each solution were therefore rotated using varimax.

Table 6 presents correlations between the various mood scales and regression estimates of these factors. Considering first the PANAS scales, Table 6 demonstrates that they have excellent factorial validity even when a frequency response format is used: In both samples the convergent correlations are above .90 and the discriminant coefficients are all low. Thus, while we prefer an extent-type rating scale, other response formats can be used without diminishing the factorial validity of the scales.

Table 6 also demonstrates that the PANAS scales compare favorably with other brief affect measures. With the exception of the Bradburn scales, all of the mood scales have good convergent correlations (i.e., .76 to .92) with the appropriate factor, but none are higher than the corresponding values for the PANAS scales. Thus, in terms of convergent validity, most of these scales are reasonable approximations of the underlying factors, although some are clearly more precise representations than others. The discriminant correlations vary widely, however, especially in the frequency-format data, where many of the coefficients exceed −.30; across both samples, only the PANAS scales have discriminant correlations consistently under −.20. Overall, the PANAS scales offer the clearest convergent/discriminant pattern of any pair.

In summary, the PANAS scales provide reliable, precise, and largely independent measures of Positive Affect and Negative Affect, regardless of the subject population studied or the time frame and response format used.

External Validity

Correlations with measures of distress and psychopathology. It is also interesting to examine correlations between the PANAS scales and measures of related constructs, such as state anxiety, depression, and general psychological distress (for an extended discussion of how Positive and Negative Affect relate to anxiety, depression, and general psychological dysfunction, see Tellegen, 1985; Watson & Clark, 1984). We have used the PANAS scales in conjunction with a number of other commonly used measures and report here on three of them: the Hopkins Symptom Checklist (HSCL; Derogatis, Lipman, Rick-

Table 6
Correlations Between Various Positive Affect (PA) and Negative Affect (NA) Mood Scales and the Factor Scores From the Extent- and Frequency-Format Data

Mood scale	Extent format Factor 1	Extent format Factor 2	Frequency format Factor 1	Frequency format Factor 2
Positive Affect scales				
PANAS	.92	−.08	.92	−.12
Diener & Emmons (1984)	.89	−.22	.87	−.36
McAdams & Constantian (1983)	.90	−.19	.86	−.31
Stone, Hedges, Neale, & Satin (1985)	.88	−.04	.81	−.20
Warr, Barter, & Brownbridge (1983)	—	—	.81	−.30
Bradburn (1969)	.50	−.18	—	—
Negative Affect scales				
PANAS	−.08	.94	−.16	.91
Diener & Emmons (1984)	−.21	.92	−.35	.89
McAdams & Constantian (1983)	−.20	.81	−.43	.76
Stone, Hedges, Neale, & Satin (1985)	.06	.84	−.11	.81
Warr, Barter, & Brownbridge (1983)	—	—	−.32	.79
Bradburn (1969)	−.21	.51	—	—

Note. *ns* with the extent-format factors ranged from 301 to 413. *ns* with the frequency-format factors ranged from 336 to 338. PANAS = Positive and Negative Affect Schedule.

Table 7
Correlations Between the Positive and Negative Affect Schedule (PANAS) Scales and the Hopkins Symptom Checklist (HSCL), Beck Depression Inventory (BDI), and STAI State Anxiety Scale (A-State)

Measure and PANAS time instructions	*n*	Correlations with PANAS NA	Correlations with PANAS PA
HSCL			
Past few weeks	398	.74	−.19
Today[a]	53	.65	−.29
BDI			
Past few days	880	.56	−.35
Past few weeks	208	.58	−.36
A-State			
Past few weeks	203	.51	−.35

Note. Unless otherwise noted, subjects are college students. PA = Positive Affect. NA = Negative Affect.
[a] Normal adult sample.

els, Uhlenhuth, & Covi, 1974), the Beck Depression Inventory (BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961), and the State–Trait Anxiety Inventory State Anxiety Scale (A-State; Spielberger, Gorsuch, & Lushene, 1970).

The HSCL (Derogatis et al., 1974) is a measure of general distress and dysfunction. Subjects rate the extent to which they have experienced each of 58 symptoms or problems during the past week. The HSCL and a subsequent 90-item version, the SCL-90 (Derogatis, Rickels, & Rock, 1976), have been used frequently as measures of clinical symptomatology in both normal and clinical populations (e.g., Gotlib, 1984; Kanner et al., 1981; Rickels, Lipman, Garcia, & Fisher, 1972). Although the HSCL and SCL-90 each contain several subscales, analyses have repeatedly shown that both instruments reflect a large general distress factor (e.g., Dinning & Evans, 1977; Gotlib, 1984).

The BDI (Beck et al., 1961) is a 21-item self-report measure of depressive symptomatology. Subjects rate whether they have experienced each symptom during the past few days. The BDI is commonly used to assess mild to moderate levels of depression, and studies have generally supported its validity in this context (e.g., Bumberry, Oliver, & McClure, 1978; Coyne & Gotlib, 1983; Hammen, 1980).

The A-State (Spielberger et al., 1970) is a 20-item scale that asks subjects to rate their current affect. Researchers have used the A-State to study subjects' responses to a variety of stressful and aversive events, including surgery, shock, pain, failure, criticism, interviews, and exams (see Watson & Clark, 1984).

Correlations between the PANAS scales and the HSCL, BDI, and A-State are presented in Table 7. Looking first at the HSCL, Table 7 indicates that it is largely a measure of NA, although it also shows modest (negative) correlations with PA. In fact, the correlations between the HSCL and the PANAS NA scale are high enough to suggest that the two measures are roughly interchangeable, at least in normal populations. Insofar as this is the case, the PANAS NA scale seems to offer a shorter (10 vs. 58 items), simpler, and conceptually more straightforward measure of general psychological distress.

The BDI is also substantially correlated with the PANAS NA scale, but the coefficients are not so high as to indicate interchangeability. In addition, the BDI has significant (negative) correlations with PA, consistent with previous findings that depressive symptomatology is affectively complex (Tellegen, 1985; Watson & Clark, 1984; Watson, Clark, & Carey, in press). That is, it involves the lack of pleasurable experiences (low PA) in addition to anger, guilt, apprehension, and general psychological distress (high NA). The PANAS scales offer the advantage of providing reliable and independent measures of these two affective components. Researchers interested in studying depressed affect might therefore want to use the PANAS scales as a complement to more traditional depression measures.

The A-State is also a mixture of high NA and low PA, replicating the results of Watson and Clark (1984, Table 4) using NA and PA factor scores. An inspection of the A-State's items indicates why this is the case. Many of the items tap mood states traditionally associated with anxiety (e.g., feeling *tense, upset, worried, anxious, nervous, jittery,* and *highstrung*) or its absence (e.g., feeling *calm, relaxed,* and *content*), and such items will produce a substantial correlation with the PANAS NA scale. Other (reverse-keyed) items, however, reflect pleasant or high PA states (e.g., feeling *joyful, pleasant, self-confident,* and *rested*) that account for the A-State's significant correlation with PA. The A-State has repeatedly demonstrated its usefulness as a sensitive measure of unpleasant mood states; but, as with the BDI, the PANAS scales offer the advantage of assessing these two affective components separately.

Intraindividual analyses of nontest correlates.[1] When used with short-term time frame instructions (i.e., moment or today), the PANAS scales are sensitive to changing internal or external circumstances. We have used the PANAS scales in three large scale within-subjects investigations that illustrate their usefulness in studying qualitatively distinctive intraindividual mood fluctuations. In the first (Watson, 1988), 80 subjects completed a PANAS questionnaire each evening for 5–7 weeks, using today time instructions. At each assessment the subjects also estimated their social activity (number of hours spent with friends that day) and rated the level of stress they had experienced. A total of 3,554 measurements were collected (M = 44.4 per subject). As hypothesized, within-subject variations in perceived stress were strongly correlated with fluctuations in NA but not in PA. Also, as expected, social activity was more highly related to PA than to NA.

The other two studies were primarily concerned with diurnal variation in mood. Clark and Watson (1986) had 123 subjects fill out a PANAS form every 3 waking hours for a week using moment time instructions. Subjects also rated their current stress and noted whether they had been interacting socially within the past hour. A total of 5,476 assessments were collected (M = 44.9 per subject). Leeka (1987) replicated this design with an additional 73 subjects (a total of 3,206 measurements; M = 43.9 per subject). In both studies, perceived stress was again consistently correlated with intraindividual fluctuations in NA but not in PA. And, as before, social interaction was more strongly related to PA than to NA.

PA also showed a strong time-of-day effect in both studies. Specifically, PA scores tended to rise throughout the morning,

[1] The data reported in Watson (1988) and Clark and Watson (1986) are based on PA and NA factor scores. We have reanalyzed these data using the PANAS scales and have obtained virtually identical results.

remain steady during the rest of the day, and then decline again during the evening. However, NA did not exhibit a significant diurnal pattern in either sample.

Conclusion

We have presented information regarding the development of brief scales to measure the two primary dimensions of mood—Positive and Negative Affect. Whereas existing scales are unreliable, have poor convergent or discriminant properties, or are cumbersome in length, these 10-item scales are internally consistent and have excellent convergent and discriminant correlations with lengthier measures of the underlying mood factors. They also demonstrate appropriate stability over a 2-month time period. When used with short-term instructions (e.g., *right now* or *today*), they are sensitive to fluctuations in mood, whereas they exhibit traitlike stability when longer-term instructions are used (e.g., *past year* or *general*). The scales correlate at predicted levels with measures of related constructs and show the same pattern of relations with external variables that have been seen in other studies. For example, the PA scale (but not the NA scale) is related to social activity and shows significant diurnal variation, whereas the NA scale (but not the PA scale) is significantly related to perceived stress and shows no circadian pattern.

Thus, we offer the Positive and Negative Affect Schedule as a reliable, valid, and efficient means for measuring these two important dimensions of mood.

References

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4,* 561–571.

Beiser, M. (1974). Components and correlates of mental well-being. *Journal of Health and Social Behavior, 15,* 320–327.

Bradburn, N. M. (1969). *The structure of psychological well-being.* Chicago: Aldine.

Brenner, B. (1975). Enjoyment as a preventive of depressive affect. *Journal of Community Psychology, 3,* 346–357.

Bumberry, W., Oliver, J. M., & McClure, J. (1978). Validation of the Beck Depression Inventory in a university population using psychiatric estimate as a criterion. *Journal of Consulting and Clinical Psychology, 46,* 150–155.

Clark, L. A., & Watson, D. (1986, August). *Diurnal variation in mood: Interaction with daily events and personality.* Paper presented at the meeting of the American Psychological Association, Washington, DC.

Clark, L. A., & Watson, D. (1988). Mood and the mundane: Relations between daily life events and self-reported mood. *Journal of Personality and Social Psychology, 54,* 296–308.

Costa, P. T., Jr., & McCrae, R. R. (1980). Influence of extraversion and neuroticism on subjective well-being: Happy and unhappy people. *Journal of Personality and Social Psychology, 38,* 668–678.

Coyne, J. C., & Gotlib, I. H. (1983). The role of cognition in depression: A critical appraisal. *Psychological Bulletin, 94,* 472–505.

Derogatis, L. R., Lipman, R. S., Rickels, K., Uhlenhuth, E. H., & Covi, L. (1974). The Hopkins Symptom Checklist (HSCL): A self-report symptom inventory. *Behavioral Science, 19,* 1–15.

Derogatis, L. R., Rickels, K., & Rock, A. (1976). The SCL-90 and the MMPI: A step in the validation of a new self-report scale. *British Journal of Psychiatry, 128,* 280–289.

Diener, E., & Emmons, R. A. (1984). The independence of positive and negative affect. *Journal of Personality and Social Psychology, 47,* 1105–1117.

Diener, E., & Iran-Nejad, A. (1986). The relationship in experience between various types of affect. *Journal of Personality and Social Psychology, 50,* 1031–1038.

Diener, E., & Larsen, R. J. (1984). Temporal stability and cross-situational consistency of affective, behavioral, and cognitive responses. *Journal of Personality and Social Psychology, 47,* 871–883.

Diener, E., Larsen, R. J., Levine, S., & Emmons, R. A. (1985). Intensity and frequency: Dimensions underlying positive and negative affect. *Journal of Personality and Social Psychology, 48,* 1253–1265.

Dinning, W. D., & Evans, R. G. (1977). Discriminant and convergent validity of the SCL-90 in psychiatric inpatients. *Journal of Personality Assessment, 41,* 304–310.

Epstein, S. (1979). The stability of behavior: I. On predicting most of the people much of the time. *Journal of Personality and Social Psychology, 37,* 1097–1126.

Gotlib, I. H. (1984). Depression and general psychopathology in university students. *Journal of Abnormal Psychology, 93,* 19–30.

Hall, C. A. (1977). *Differential relationships of pleasure and distress with depression and anxiety over a past, present, and future time framework.* Unpublished doctoral dissertation, University of Minnesota, Minneapolis.

Hammen, C. L. (1980). Depression in college students: Beyond the Beck Depression Inventory. *Journal of Consulting and Clinical Psychology, 48,* 126–128.

Harding, S. D. (1982). Psychological well-being in Great Britain: An evaluation of the Bradburn Affect Balance Scale. *Personality and Individual Differences, 3,* 167–175.

Hedges, S. M., Jandorf, L., & Stone, A. A. (1985). Meaning of daily mood assessments. *Journal of Personality and Social Psychology, 48,* 428–434.

Kammann, R., Christie, D., Irwin, R., & Dixon, G. (1979). Properties of an inventory to measure happiness (and psychological health). *New Zealand Psychologist, 8,* 1–9.

Kanner, A. D., Coyne, J. C., Schaefer, C., & Lazarus, R. S. (1981). Comparison of two modes of stress measurement: Daily hassles and uplifts versus major life events. *Journal of Behavioral Medicine, 4,* 1–39.

Leeka, J. (1987). *Patterns of diurnal variation in mood in depressed and nondepressed college students.* Unpublished master's thesis, Southern Methodist University, Dallas, TX.

McAdams, D. P., & Constantian, C. A. (1983). Intimacy and affiliation motives in daily living: An experience sampling analysis. *Journal of Personality and Social Psychology, 45,* 851–861.

Moriwaki, S. Y. (1974). The Affect Balance Scale: A validity study with aged samples. *Journal of Gerontology, 29,* 73–78.

Rickels, K., Lipman, R. S., Garcia, C. R., & Fisher, E. (1972). Evaluating clinical improvement in anxious outpatients. *American Journal of Psychiatry, 128,* 119–123.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39,* 1161–1178.

Russell, J. A. (1983). Pancultural aspects of the human conceptual organization of emotions. *Journal of Personality and Social Psychology, 45,* 1281–1288.

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). *Manual for the State-Trait Anxiety Inventory.* Palo Alto, CA: Consulting Psychologists Press.

Stone, A. A. (1981). The association between perceptions of daily experiences and self- and spouse-rated mood. *Journal of Research in Personality, 15,* 510–522.

Stone, A. A. (1987). Event content in a daily survey is differentially associated with concurrent mood. *Journal of Personality and Social Psychology, 52,* 56–58.

Stone, A. A., Hedges, S. M., Neale, J. M., & Satin, M. S. (1985). Prospective and cross-sectional mood reports offer no evidence of a

"Blue Monday" phenomenon. *Journal of Personality and Social Psychology, 49,* 129–134.

Tellegen, A. (1985). Structures of mood and personality and their relevance to assessing anxiety, with an emphasis on self-report. In A. H. Tuma & J. D. Maser (Eds.), *Anxiety and the anxiety disorders* (pp. 681–706). Hillsdale, NJ: Erlbaum.

Tessler, R., & Mechanic, D. (1978). Psychological distress and perceived health status. *Journal of Health and Social Behavior, 19,* 254–262.

Warr, P. (1978). A study of psychological well-being. *British Journal of Psychology, 69,* 111–121.

Warr, P., Barter, J., & Brownbridge, G. (1983). On the independence of positive and negative affect. *Journal of Personality and Social Psychology, 44,* 644–651.

Watson, D. (in press). The vicissitudes of mood measurement: Effects of varying descriptors, time frames, and response formats on measures of Positive and Negative Affect. *Journal of Personality and Social Psychology.*

Watson, D. (1988). Intraindividual and interindividual analyses of Positive and Negative Affect: Their relation to health complaints, perceived stress, and daily activities. *Journal of Personality and Social Psychology, 54,* 1020–1030.

Watson, D., & Clark, L. A. (1984). Negative Affectivity: The disposition to experience aversive emotional states. *Psychological Bulletin, 96,* 465–490.

Watson, D., Clark, L. A., & Carey, G. (in press). Positive and Negative Affectivity and their relation to anxiety and depressive disorders. *Journal of Abnormal Psychology.*

Watson, D., Clark, L. A., & Tellegen, A. (1984). Cross-cultural convergence in the structure of mood: A Japanese replication and a comparison with U.S. findings. *Journal of Personality and Social Psychology, 47,* 127–144.

Watson, D., & Pennebaker, J. W. (in press). Health complaints, stress, and distress: Exploring the central role of negative affectivity. *Psychological Review.*

Watson, D., & Tellegen, A. (1985) Toward a consensual structure of mood. *Psychological Bulletin, 98,* 219–235.

Wills, T. A. (1986). Stress and coping in early adolescence: Relationships to substance use in urban school samples. *Health Psychology, 5,* 503–529.

Zevon, M. A., & Tellegen, A. (1982). The structure of mood change: An idiographic/nomothetic analysis. *Journal of Personality and Social Psychology, 43,* 111–122.

Appendix

The PANAS

This scale consists of a number of words that describe different feelings and emotions. Read each item and then mark the appropriate answer in the space next to that word. Indicate to what extent [INSERT APPROPRIATE TIME INSTRUCTIONS HERE]. Use the following scale to record your answers.

1	2	3	4	5
very slightly or not at all	a little	moderately	quite a bit	extremely

_____ interested	_____ irritable
_____ distressed	_____ alert
_____ excited	_____ ashamed
_____ upset	_____ inspired
_____ strong	_____ nervous
_____ guilty	_____ determined
_____ scared	_____ attentive
_____ hostile	_____ jittery
_____ enthusiastic	_____ active
_____ proud	_____ afraid

We have used PANAS with the following time instructions:

Moment	(you feel this way right now, that is, at the present moment)
Today	(you have felt this way today)
Past few days	(you have felt this way during the past few days)
Week	(you have felt this way during the past week)
Past few weeks	(you have felt this way during the past few weeks)
Year	(you have felt this way during the past year)
General	(you generally feel this way, that is, how you feel on the average)

Received May 10, 1987
Revision received September 14, 1987
Accepted November 11, 1987 ■